Gaze2Grasp: Vision-based system for pre-grasp prosthesis control

Maxim Karrenbach

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

Master of Science

University of Washington

2021

Committee:

Amy Orsborn

Eric Rombokas

Program Authorized to Offer Degree:
Electrical and Computer Engineering

University of Washington

Abstract

Gaze2Grasp: Vision-based system
for pre-grasp prosthesis control

Maxim Karrenbach

Co-Chairs of the Supervisory Committee:
Dr. Amy Orsborn
Electrical and Computer Engineering

Dr. Eric Rombokas
Electrical and Computer Engineering

Users of prosthetic limbs face many challenges operating their devices in everyday tasks. The many aspects of doing tasks like stair descent and object grasping are largely done subconsciously in a person without any limb impairment. A person with a limb amputation or impairment bears the full weight of this cognitive load, and are often forced to deal with the limitations of the prosthesis by employing compensation strategies, like "overhanging toe" in stair descent and shoulder compensations in grasping tasks.

My research aims to ease the burden placed on prosthetic limb users by using the learning capabilities from data-driven methods of neural networks particularly in object grasping. *Gaze2Grasp* is a machine learning algorithm that uses eye-tracked gaze to predict preferable wrist rotations of a virtual prosthesis. By using an input of vision and providing outputs of wrist rotations, *Gaze2Grasp* could be implemented in any prosthetic controller with wrist degrees of freedom, and shows the value of vision in object grasping. The algorithm may also be useful for other applications like remote-operated robotic grasping systems.

ACKNOWLEDGMENTS

I would like to express my gratitude towards my academic advisers Amy Orsborn and Eric Rombokas. To Amy, I am grateful for your interest in my work and your willingness to provide support to my research. To Eric, I am grateful for your everlasting patience and for providing me with an environment of growth and curiosity during my time.

I extend my thanks towards my colleagues, past and present, in my research lab for your collaboration and expertise. Furthermore, I would like to thank Astrini Sie for taking a chance on me and bringing me into the world of research. I would also like to thank David Boe for your constant guidance during my research projects. The mentorship that you both provided me has given me an appreciation for the work that I do and a lifelong interest in research.

To the friends that I made over the years, I want to thank you for the good conversations, the shared struggles, and the strong bonds. Though I met many great people, I remember fondly each and every one of you and I hope for more good times together.

To Rose and Mako, I want to thank you for the fun adventures, for making sure that I take care of myself during times of intense work, for pushing me to improve myself, and for always celebrating all successes, no matter how small.

Finally, I would like to thank my family for their unconditional encouragement for pursuing what I love. Thank you for always believing in my choices and supporting my path throughout life.

DEDICATION

To those who strive to
Engineer to help others
May our work succeed

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 Motivation

1.1.1 Need for improvement in prosthetic devices

Mobility and natural human movement are functions that are not often thought about until a loss. This can be in the form of an accident, an impairment, or a congenital condition. It was found in 2005 that 1.6 million persons in the US were living with the loss of a limb. While more recent numbers are not available, it is estimated that the person with amputation (PWA) population will increase to 3.6 million persons by 2050. [1]. It is estimated that there are around 185,000 persons undergo amputations every year. For most of these PWA, a prosthesis will be an option in terms of rehabilitation. Yet, a survey done on 808 veterans with prosthetic limbs found that 379 users reported abandoning their prosthesis at some point during their life. This rate of abandonment can be blamed on various shortcomings that exist in almost all current versions of upper and lower limb prostheses [2]. With the increasing rates of amputations, fueled by rising rates of diseases like diabetes [3], it is important to improve the control and performance of prosthetic limbs.

1.1.2 Prediction in prosthesis control

A general prosthesis control flow begins with the task presented to the user and the user considering the best approach to the task. Due to the presence of the impairment, a cognitive load is placed upon the user to accommodate the impairment. While prostheses are meant to restore natural movement, the cognitive load is something that most prostheses do not address. However, introducing predictive systems to the prosthesis can have an impact of reducing cognitive load [4]. The prediction, in the form of feedback, cueing, actual prosthesis control, or other information, is able to reduce the amount of cognitive load that is present during the task, which allows the user a better experience overall.

1.1.3 ***Smart prostheses***

Smart prostheses incorporate some sort of algorithm or model in order to better operate closer to natural movements. In most cases, machine learning is used to predict the outcome of a task or the optimal method to reach this outcome. While many different learning methods exist, artificial neural networks (ANN) have been found to be a good choice for these types of learning problems [5, 6]. Shehata et al finds that while ANNs have long training times and large computational costs, they show great promise in upper-limb control strategies. [5] In both classification and regression problems, ANNs are able to reliably learn the relationship between the inputs and the desired outputs.

In upper-limb control, Ameri et al designed a control scheme using an ANN is able to identify reliable motor controls from electromygraphy signals (EMG) [7]. In lower limb control systems, Woodward et al shows that ANNs have been used to predict the intent of a user to switch between preprogrammed modes of use (climbing stairs, flat ground walking, etc) [6]. This implementation using ANNs are significantly more accurate than other learning methods, like linear discriminant analysis (LDA). In the realm of prosthetic development, smart prostheses using predictive control strategies with underlying machine learning algorithms are creating a path towards more intuitive and useful prosthetic devices.

1.2 Research Goal

The goal of my research is to predictive control strategies to create technology to assist PWA with everyday tasks and restore natural movement. Both people with upper-limb and lower-limb impairments benefit from predictive control strategies, and my research will include both of these groups.

For people with lower limb amputations, I hope to allow a restoration to more natural movements, especially in the accomplishment of stair descent tasks. For this goal to be recognized, I collaborated on a project called SmartStep. The SmartStep device focuses on cueing users to match their foot to a predicted foot placement. My contributions to this project came in two sections:

Aim 1.1*: Investigate the ability of subjects to adjust strides based on haptic cueing*

To figure out if subjects can be cued to adjust their stride length in a mobility task, we conducted a study on participants doing a flat-ground walking task. The subject wore sensors and a feedback device which sent vibrations to the wrist, and gait data was collected to determine when the signal should be sent for the user to adjust a target foot placement.

Findings from this sub-aims are to be published in [8].

Aim 1.2*: Collecting a large-scale dataset of stair descent gait data from intact subjects in a non-laboratory setting*

To train a model to predict future foot placements of a user when descending stairs, a large dataset is necessary. This dataset was taken on a large number of subjects wearing sensors and insoldes while descending stairs in a non-laboratory setting. This dataset is distinguished from others in the field, as, to our knowledge, this is the first large scale dataset of gait data for stair descent that is public and not collected in a laboratory.

Findings from this sub-aims are to be published in [9].

For people who use upper-limb prostheses, I wish to investigate a way for users to perform everyday tasks in a manner that is more intuitive and less of a cognitive burden. Specifically, I focus on improving the execution of object grasping tasks. In order to realize this goal, I am developing the Gaze2Grasp algorithm to predict optimal wrist angles from the user's eye-tracked gaze. This algorithm will be created and tested in the following two sub-aims:

Aim 2.1*: Collect and train algorithm on joint kinematic and eye-tracked gaze data from intact subjects performing object grasping tasks in virtual reality*

To be able to train a model, I require a dataset of gaze frames and corresponding wrist orientations. I perform data collection on intact users in virtual reality, using eye-tracking software and a hand-tracking device. The most similar datasets are done with gaze-tracking cameras (head-tracked or eye-tracked) for the scene, but not explicitly record any joint kinematics. These datasets are mainly for grasp classification and grasp synthesis, which is a separate goal from ours.

The dataset will be used to train a model, on which cross-validation and hyperparameter tuning will be done to create the highest-performing models. A test set is held apart from the training and validation set to test against some baseline models.

Aim 2.2*: Predicting optimal wrist configurations during virtual object manipulation tasks and investigating performance against non-predictive prosthesis*

I evaluate the qualitative performance of the algorithm by implementing it into a virtual prosthesis and performing grasping tasks. I compare this configuration with a wrist-locked

configuration. This user study is performed on a set of intact subjects, and the performance is evaluated.

Findings from this aim are to be published in [10].

1.3 Scope and future implementations

In Aim 1, I will collaborate to create a dataset of haptic cueing gait data for analysis on stride adjustment, as well as a dataset for training an algorithm to predict future foot placement. This research will focus mainly on the dataset collection, while collaborators will do further work on implementing the system as a whole.

This research will focus primarily on the development of the device hardware and the predictive system of the device. The final implementation of this device into real-time systems for PWA will not be addressed in this work.

In Aim 2, I will create and validate a system which calculates the optimal wrist rotation based on the eye-tracked gaze of the user. The dataset collection, the development of this algorithm and the shown improvement over wrist-locked prostheses is the main goal of the research.

This research will only focus on the development and verification of the algorithm. Implementing this algorithm into a real life prosthesis will not be considered in this work. However, this algorithm, with modifications for latency and synchronization, would be an interesting addition to a upper limb prosthesis with wrist degrees-of-freedom. The setup for this would require an eye-tracking depth camera worn by the user, and the algorithm incorporated into the controller of the prosthetic limb.

1.4 Organization

This manuscript is organized as follows:

- Chapter 1: Motivation, aims, and scope of the research.
- Chapter 2: Background, current works, and state of the art related to the research
- Chapter 3: Aim 1
- Chapter 4: Aim 2.1
- Chapter 5: Aim 2.2
- Chapter 6: Conclusion

Chapter 2

BACKGROUND AND PRIOR WORK

2.1 Prosthesis accessibility and abandonment

Despite prostheses being necessary for a person with amputation (PWA) to regain mobility in the lost limb, current options for prostheses [11, 12] that offer high functionality and a restoration of natural movement are limited. These options are further limited by cost, weight, and fit. While body-powered or passive prostheses cost much less than these, it is important to note that users have already had to go through amputations, rehabilitation, and other such costly procedures. Furthermore, prosthetic limbs require repair, replacements, and upkeep, and so the decision to use a prosthetic limb can be costly. In order for this cost to be worth it, the functionality must improve to provide more natural movement.

There is a problem that plagues all areas of prosthetic rehabilitation; the loss of a human capability leads to either compensation or abandonment. When a prosthesis introduced more problems than it solves, then it fails to provide the user with an assistive experience and rather becomes a burden to use. There is a variety of reasons for why a user would reject a prosthetic limb, including weight, comfort and functionality [13, 14, 15] Surveys on prosthetic rejection and abandonment have found that the abandonment and rejection rate can be as high as 34 and 44% in subjects with upper-limb prostheses caused by congenital arm defects [16] and trauma, respectively [17]. Even with users of ortheses, the rate of abandonment was as high as 20% [18], though upper-limb ortheses and prostheses were found to have higher levels of abandonment. In a survey of veterans with amputations, it was found in Figure 2.1 that reasons like "lack of comfort", "too much fuss", "fit/comfort", and "heavy/fatiguing" were all reasons given for prosthesis abandonment [2]. However, a survey of PWA who rejected or abandoned a prosthetic device found that 72% of these subjects may be willing to reconsider prosthesis use if the devices had a lower cost and increased functionality [19].

2.1.1 Compensatory strategies of lower-limb prosthesis users during stair descent

As many lower-limb prostheses do not have proper function to accomodate stair descent, some users tend to avoid using stairs altogether. For those with active protheses, there are a few strategies to use. The first is the "step-to" technique , where the user will lead with the

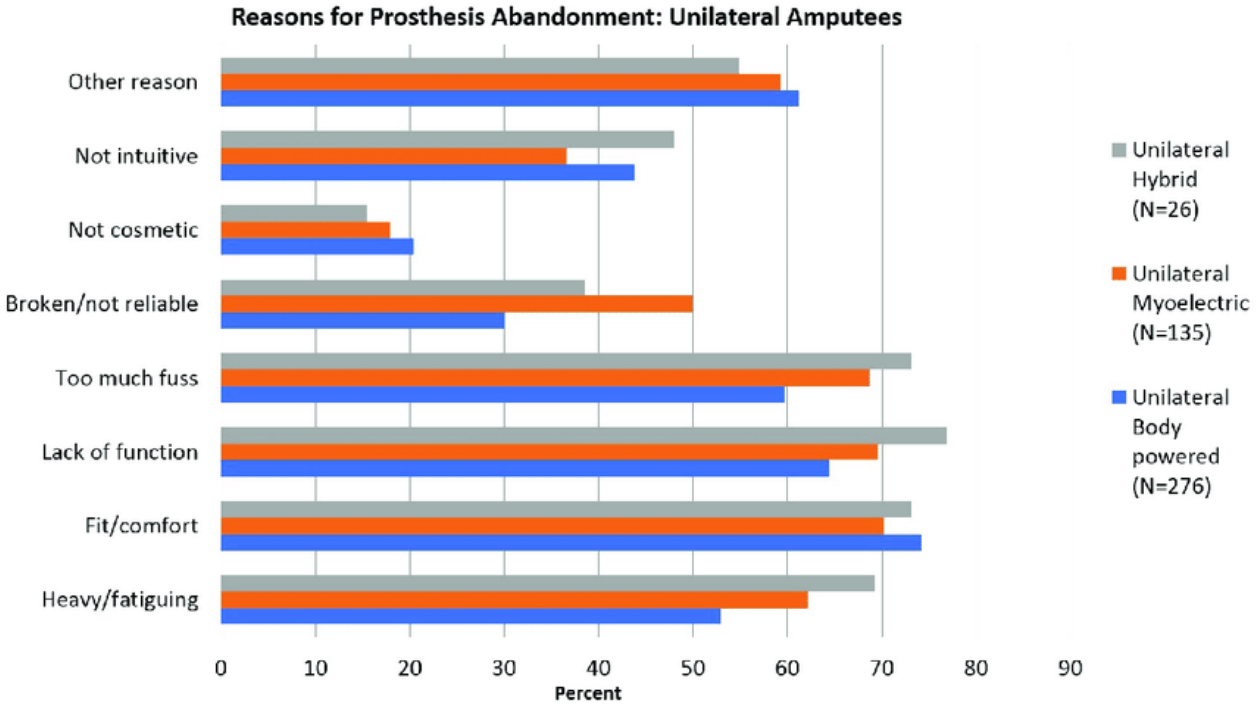

Figure 2.1: Responses for reasons of prosthesis abandonment, taken for unilateral users [2]

prosthetic limb, achieve ground contact, then move their intact foot to the prosthetic foot and repeat. [20] Other users will approach the stairs like intact users, but use "toe overhang", where the foot is placed a fixed distance to the step edge and use the step edge as a pivot for a rolling motion. [21, 22]. In all strategies, vision is used heavily to confirm accurate foot placement, which both takes the vision system away from other uses and interrupts a smooth descent.

2.1.2 Compensation for low degree-of-freedom upper limb prostheses in reaching tasks

A common problem in modern prosthetic hands is that degrees of freedom of wrist rotation are overlooked. Even complex prostheses strive for finger dexterity and the ability to create different grasp shapes, but often only include a manually operated wrist [11, 12]. In more accessible options, like hooks and simple graspers, wrist rotation is commonly manually operated by the intact arm. The user then has two choices of compensatory strategies: Using the intact arm, whether to manually rotate the prosthetic wrist unit or to do the task altogether, or using the shoulder to rotate the wrist as desired. The first strategy of using the intact arm leads to less use of the prosthesis altogether, while users of the second strategy

find the burden of hand orientation to be passed to the forearm, upper arm, or chest. The user performs compensatory motions in order to orient the hand in a way to make up for the lack of movement in the wrist, which when constantly repeated, can lead to limb pain and overuse syndrome in the limb [15]. This is one of the main causes for prosthesis rejection and abandonment [19]. The cognitive load of having to orient the hand based on shoulder or chest movements is then also introduced to the user.

In the survey of veterans with amputations, upper-limb prosthetic users were surveyed about certain conditions. From this, it was found that 71% report pain in other arm and 51% report conditions like arthritis or carpal tunnel in that arm. This is a result of using the intact limb as a compensatory strategy, as the user is overusing the other limb to do all tasks or rotate the prosthetic hand manually. It was also found that 72% reported back pain and 60% reported neck pain, which is a direct result of using shoulder and chest movement as a compensatory strategy, because those users are overusing the shoulder system. Overall, all compensatory strategies have drawbacks and stem from the lack of wrist function.

2.2 Data-driven prediction of reaching tasks

2.2.1 Vision and grasping tasks

Vision is an important system in many daily mobility tasks. The addition of extra visual systems is also beneficial to performing mobility tasks. He et al designed an artificial vision system with thermal imaging, and found that with the extra information provided, subjects had less errors and collisions, but did not compromise walking speed [23]. Furthermore, the lack of vision information, or the presence of particular visual impairments, leads to a decrease in mobility performance [24]. Several studies to design prosthetic devices to restore vision have been done [25, 26] because of the impact that vision has on mobility tasks. Karl et al found that the loss of vision in a reaching task led to more elevated trajectory of the hand, and required more contact adjustments [27].

Human object grasping is a complex task requiring fine motor control. Grasping can be divided into three phases: the first phase is the pre-grasping phase, which is comprised of two parallel tasks performed by the user's hand, which are hand transportation and hand pre-shaping. Hand transportation involves the hand trajectory, which is influenced by the distance of the object's grasping point to the hand [28, 29], while pre-shape encompasses the rotational changes of the hand joints to match the forecasted grasp shape [28, 30]. The second and third phase are grasping and manipulation, in which the object is contacted with a stable hold and further manipulated with respect to the environment or the hand via translation, rotation, or precision movements [28, 31, 32].

Vision has been shown to be predictive of intent and point of contact in everyday reaching

tasks. Regardless of the environment, the gaze tends to focus solely on objects involved in a task [33], ignoring any irrelevant sections of a scene [34]. During the time of gaze focus, the eyes tend to focus towards the natural grasping points of objects [35], known as affordances. These affordances are defined as the properties of objects that are directly linked to motor performances [36]. Being able to identify grasping contact points provides information on the pre-grasp approach, and so vision can be leveraged to create intelligent grasping devices.

2.2.2 Parallels of vision in robotic grasping

Data-driven analysis methods are commonly used to approach the task of robotic grasping. Most studies focus on point clouds as a dataset, as other studies that include the object models [37] or aim to reconstruct the object model from the available data [38] tend to have less success in grasps for novel objects [37, 38]. A good example of utilizing vision with incomplete point clouds is the study done by Kopicki et al, in which researchers synthesized grasps with only point clouds taken with a depth camera [39]. A limitation of grasping algorithms in general was highlighted in this study, which was the issue of multiple grasp choices with limited points of view of the object. This is a common problem in datasets of robotic grasping, which have been available for the study of various training methods [40, 41]. However, many of these datasets are collected from robotic grasping, and not on human subjects, which does not provide much insight towards the impact of human vision on grasping. Since eye-tracked gaze has been shown to predict grasp contacts, we implement gaze via eye-tracking to obtain such information for human grasp replication. However, in the realm of intelligent prosthetic devices, replacing human grasping has proven to be challenging.

Previous works have established a multitude of techniques through which the problem of robotic grasping has been approached. One set of approaches are the techniques using learning algorithms (SVM [42], Markov Chains [43], and deep learning [44]) to classify the type of grasp for a given reach. Classification is a useful approach to determine the type of grasp. Classification hierarchies have various groupings (i.e. a power grasp or a precision grasp) and subgroupings within each group may exist to further specify a better grasp choice (i.e. a palmar pinch grasp or a tripod grasp) [45]. However, classification-based systems can be limited since the chosen grasp may not fit the object exactly. The adjustments to the hand to validly grasp an object are then pushed to the grasping phrase, via techniques like force-closure grasping or task-oriented grasping [46, 47]. Thus, a big challenge in grasp synthesis comes from the broad range of objects that an algorithm has not seen.

2.3 Similar works in upper-limb prosthesis control

Recent studies have shown great promise in improvements in grasp synthesis for advanced myoelectric prosthetic hands by using eye-tracked gaze. For example, Krausz [48] et al created a prosthesis controller based on the combination of eye-tracked gaze and EMG to synthesize grasps. The study reconfirmed that gaze preceded hand motion during grasping tasks, establishing the impact that gaze can have on reaching tasks. Our study differs from this by focusing on predicting wrist rotations from gaze. While this study found that vision had a impact on predicting grasp poses, it is unknown whether gaze has the same impact on solely the wrist rotations.

Cognolato [49] et al provides an approach towards grasp classification by using vision and gaze tracking on amputees that already use prosthetic hands in order to provide an improvement in creating pre-grasp hand poses. A large dataset was collected on both intact subjects and amputees, and provides gaze data, as well as myoelectric signals, and various learning algorithms were implemented for grasp classification. A previous study by Cognolato [50] used a convolutional neural network to perform object classification with classes created from grasp types. In these two studies, subjects were instructed to grasp the objects in a certain way according to the grasp class in which the object existed in. While we also use eye-tracking gaze and a similar architecture, our experiment differs from these two studies because our study instructs subjects to grasp the objects in the way that feels most natural to them. This incorporates intent and natural grasping into our experiment, and we implement eye-tracking gaze to identify the choice of affordance of the subject and the corresponding wrist orientation. Furthermore, these recent studies tend to focus on whole-hand kinematics, which requires advanced prostheses with high degrees of freedom, which are usually inaccessible to a majority of the population [51], and so we differ from these studies by focusing on possible improvements to any prosthesis with wrist degrees of freedom.

2.4 Conclusion

Current implementations of prostheses that are accessible to a vast majority of amputees fail to afford their users with basic function, whether it is lack of sensory feedback or wrist functionality. Users must then employ compensatory strategies in order to make up for this lack of function, and end up with cognitive load and physical consequences in other systems of the body.

To assist with this problem, prostheses which use predictive control systems, as well as assistive devices that use predictive strategies, are being created. I propose to use data-driven methods in the form of neural networks to create predictive systems that allow more function to the user. In lower-limb devices, I collaborate on a project called SmartStep, which uses

neural networks to predict future foot placements and cue the user with haptic feedback. For upper-limb prostheses, I propose a system called Gaze2Grasp, which uses neural networks to predict wrist rotations from the user's vision data. Both of these systems aim to assist the user by using prediction to provide useful information to the user.

Chapter 3

DATA COLLECTION FOR USE IN PREDICTIVE LOWER-LIMB SYSTEMS

3.1 Data Collection: Gait Data of Vibrotactile Feedback-Driven Foot Placement Adjustment Task

This dataset consists of feedback-driven flat-ground walking gait data collected from 10 participants recruited in accordance to a University of Washington Institutional Review Board protocol. The study analyzing this dataset has a publication under review in [8]

3.1.1 Sensors and Equipment

The experiment requires two main sets of equipment: the hardware and the data recording system.

The main hardware in this experiment is a wrist-worn vibrotactile feedback device, shown in Figure 3.1. This device consists of two motors (Vibrating Mini Motor Disc 1201, Adafruit, New York City, NY, USA) on the dorsal and palmar side of the wrist driven by motor drivers (DRV2605, Texas Instruments Incorporated, Dallas, TX, USA). The motor cues are controlled by an Adafruit microcontroller (Adafruit Feather M0 Adalogger microcontroller 2796, Adafruit, New York City, NY, USA) connected to a multiplexer (TCA9548A I^2C, Texas Instruments Incorporated, Dallas, TX, USA). The system interfaced to a computer using a bluetooth transmitter (HC-05, Shenzhen HiLetgo Technology Co., Ltd, Shenzhen, China). All components are encased in a 3D-printed case, secured to an generic elastic band with Velcro, except the palmar motor, which is secured within the elastic band separately.

The experiment requires the collection of full body kinematic data and ground reaction force data. To collect the full body kinematic datadata, we used a Qualisys motion capture system (Qualisys AB, Göteborg, Sweden). This system consisted of reflective motion capture marks attached to specific landmarks by adhesive, as well as limb clusters which were worn via elastic straps. The ground reaction force data was captured using four force plates (Kistler Group, Winterthur, Switzerland) in the path of the participant.

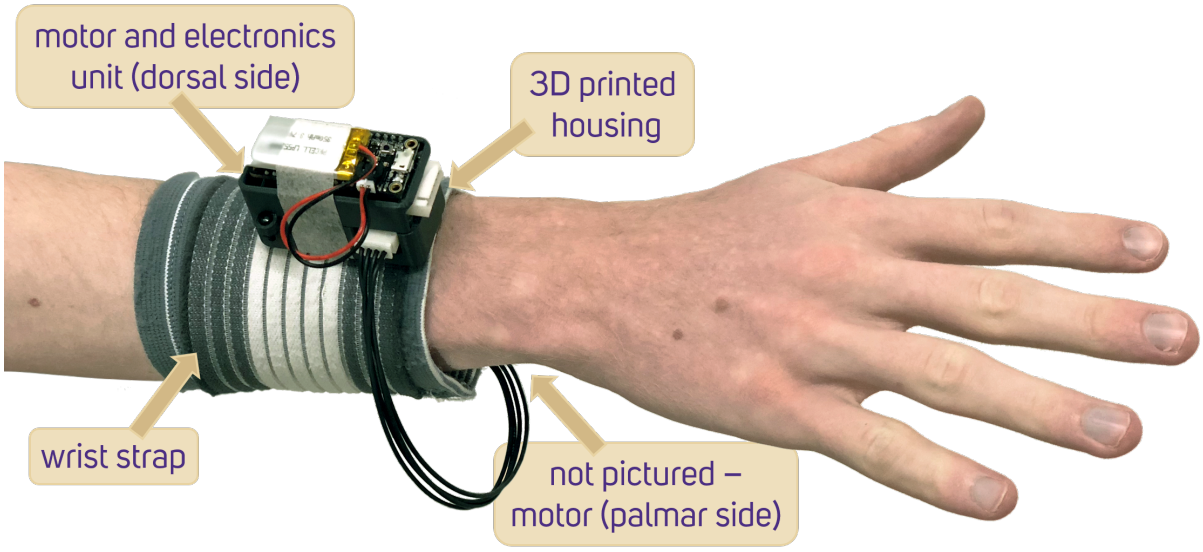

Figure 3.1: Device used for sending vibrotactile cues to the dorsal and palmar side of the wrist for stride adjustment [8].

3.1.2 Task

The participant is instructed to walk in a forward pass along a walkway, shown in Figure 3.2, during which haptic cues were sent via the wrist apparatus, shown in Figure 3.1. The participant first did a calibration for the motion capture system, and then established a starting position, which was approximately 5 strides from the first force place. The participant is then instructed to adjust the stride length of the next right foot based on the haptic cues received from the apparatus, with a longer step if the cue is felt on the palmar side and a shorter step if the cue is felt on the dorsal side. The participants are instructed to avoid stopping during the forward pass, but are not given any instruction regarding walking speed, stride length adjustment, or subsequent adjustments. One trial consists of a forward pass, in which one haptic cue is sent, and a return to the starting position. Each data collection session recorded 84 trials, and participants were allowed to take a break between trials if needed.

Each trial had either one haptic cue, which could come at one of six different times and could be either a longer step cue or a shorter step cue, or lacked a haptic cue. This haptic cue would happen after a specific gait event is detected, according to the stimulation time, and is delivered at a random location along the walkway.

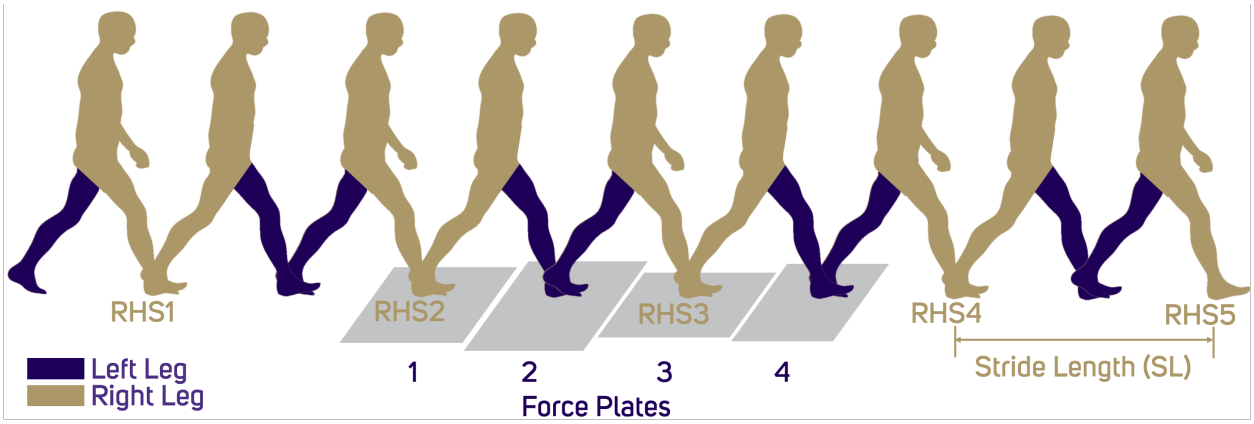

Figure 3.2: Visualization of the task. Subjects were instructed to walk in a straight pathway over force places while awaiting haptic cues. [8].

3.1.3 Data Processing

The gait data collected from the motion capture system required manual cleaning to track each of the anatomical landmarks and clusters throughout each trial. In order for the data to be usable, each anatomical landmark and cluster had to be recognized and labelled during the majority of the trial. For each trial, markers were hand-labelled and tracked to ensure that there were no dropped labels or unrecognized markers.

For analysis, the foot strikes also had to be identified in order to center the target foot. For each trial, at least five heel strikes were recorded, so manual identification of the first right heel strike had to be done. the first heel strike was defined as the right heel strike happening prior to contact with the first force place. We were then able to identify at most four heel strikes prior to the target foot strike and at most three heel strikes after the target foot strike.

3.1.4 Further Work

The next steps in the project. consisted of more processing steps, like normalization, statistical analysis, and a full analysis of the stride length, gait data, walking speed on adjustments, and more. This work was done by other contributing authors, and is described in [8]. The main relevant conclusions were that the user could be cued to adjust the stride of the target foot. Stimulation timings are shown in Figure 3.3. For shorter strides, the user needed the cue sent before the toe-off of the left foot (ST2) in the previous step, while for longer strides,

the user needed the cue sent before the midswing of the left foot (ST3) in the previous step.

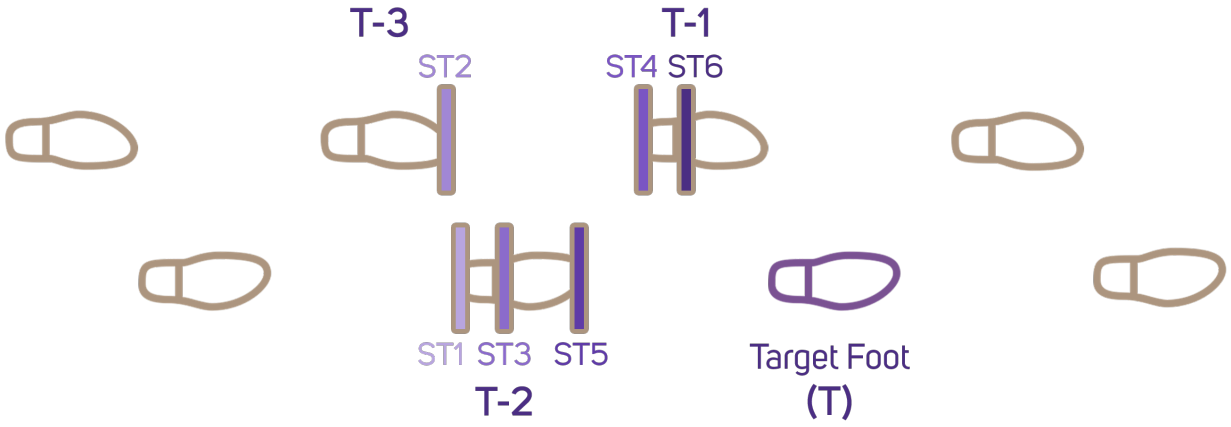

Figure 3.3: Stimulation times sent during the trials. Findings show that shorter stride adjustments needed the cue sent before ST2 and longer stride adjustments needed the cue sent before ST3 [8].

3.2 Data Collection: Gait Data of Stair Descent Task

This dataset consists of stair ambulation gait data collected from 101 participants recruited in accordance to a University of Washington Institutional Review Board protocol. This dataset, which is collected specifically for stair ambulation, has a publication under review in [9]

3.2.1 Sensors and Equipment

This experiment requires the capturing of full body kinematics and forces of the foot. To capture the full body kinematics, we used an XSENS motion capture system (MTw Awinda, Xsens Technologies B.V., Enschede, Netherlands). This system consisted of a lycra shirt, headband, and inertial measurement units attached on the head, torso, and limbs via elastic straps and placements on the shirt and headband. To collect the normal forces under each foot, we used Moticon insoles (Moticon SCIENCE ver. 1 and Moticon SCIENCE ver. 2, Moticon ReGo AG, Munich, Germany). The Moticon insoles used were European shoe size

44/45, or US Men's 9, so participants who had equal or larger shoe sizes would wear them inside a non-arched athletic shoe. Participants with smaller shoe sizes wore rubber overshoes with the insole between their shoe and the rubber overshoe. When using the Moticon ver. 1 insoles, an Arduino board (Arduino MEGA 2560, Arduino, Somerville, MA, USA) and an wifi shield (Adafruit WINC 1500 Wifi Shield, Adafruit Industries, New York, NY, USA) were used to record and sync the data. However, the Moticon ver. 2 had capabilities of on-board data storage, and these components became unused.

3.2.2 Task

The experiment consists of the subject ascending and descending a staircase (rise of 15.5cm and run of 33cm) with 13 steps. Before any data is collected, the subject does a calibration for the motion capture system. One trial consists of the subject descending and ascending the staircase, with a one second pause when they reach the bottom. No instructions were given with regards to the leading foot, the pace, or the turn direction between descents and ascents. Each data collection session ranged between 20 and 50 trials, and most participants took breaks every 10 or 15 trials. The sensors, as well as data visualizations of the ascents and steps, are shown in Figure 3.4.

3.2.3 Data Recording

The data that was recorded consisted of full body kinematics recorded in the XSENS program, in a specific format (.mvn), and the foot force data recorded on-board on the Moticon insole. The full body kinematic data consists of the time series of joint angles and 3D positions of each of the XSENS sensors. These values were sampled at a rate of 60 Hz for the duration of the sessions between breaks. The Moticon data consisted of the time series of pressures of each sensor in the insoles, as well as the foots's total acceleration, force, and center of pressure. These values were sampled at a rate of 50 Hz for the duration of the sessions between breaks. Each subject had several sets of Moticon and XSENS files depending on how many breaks were taken, and each set of Moticon and XSENS files had several trials of data.

3.2.4 Data Preprocessing

The full body kinematic data is reprocessed in HD to provide for cleaner data, and then is exported to a comma-separated-value file using a Matlab plugin for the XSENS program. The Moticon data is downloaded from the insole's on-board storage into a plain text format (.txt).

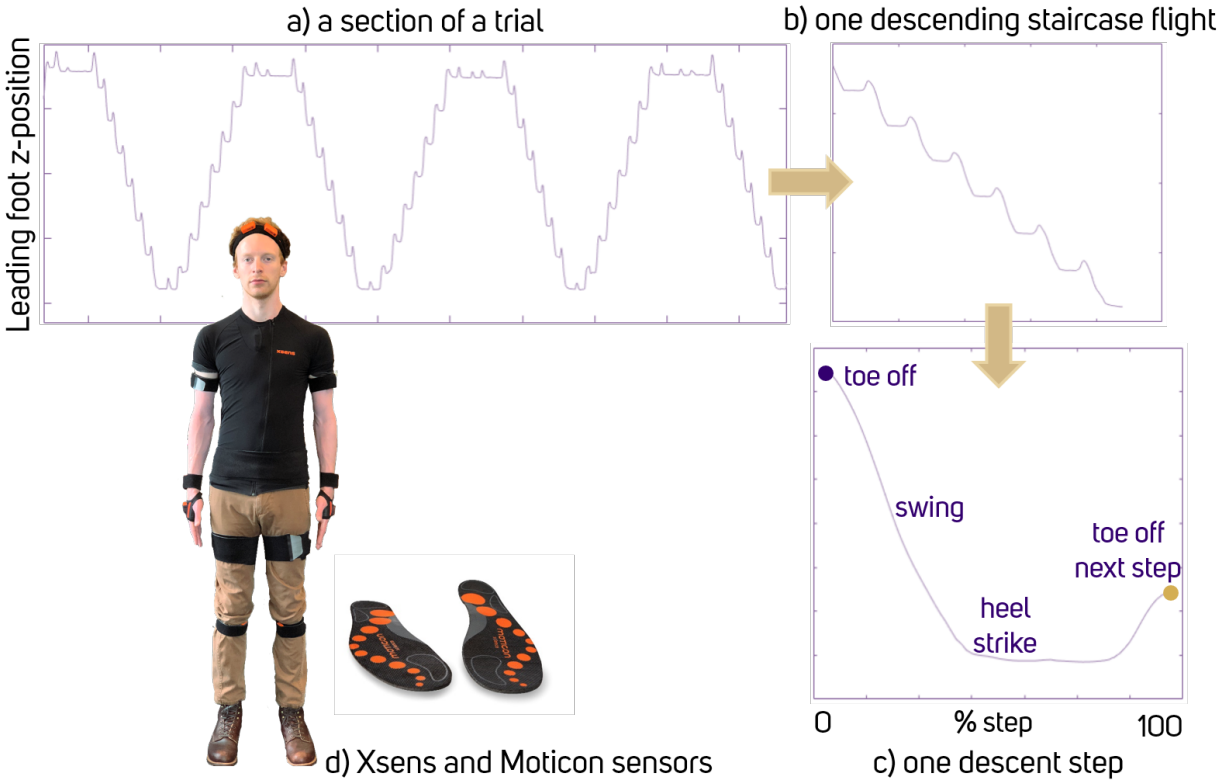

Figure 3.4: Segmentation of the data shown for a trial section, staircase steps, and individual step. Also shown are the sensors and insoles worn by the participant [9].

3.2.5 Further Work

The next steps in the project. consisted of more preprocessing steps, like resampling, syncing and filtering, segmentation both to descents and to steps, and analysis of the data. This work was done by other contributing authors, and is described in [9].

3.3 Conclusion

For predictive controllers to be implemented into prosthetic systems, datasets are crucial to train the system. In order for the predictive controller to mimic human movement as closely as possible, these datasets should be based on human subjects where possible to gain insight into how these tasks are approached by human systems. Datasets such as the ones

described in the above sections are necessary to analyze human cognitive response and to develop systems which can perform in human-in-the-loop systems.

Chapter 4

IMPACT OF EYE-TRACKED GAZE ON HAND PRE-GRASP POSE: DATASET AND MODEL DEVELOPMENT

4.1 Dataset Collection

In this experiment, we collect a dataset of vision and hand kinematic data on subjects reaching in a virtual environment to grasp objects. Subjects in this study consisted of 6 intact (non-amputee) right-handed subjects (5 males, 1 female). Subjects were compensated for their contributions and all subjects gave informed consent to the study. All experimental procedures were approved by the Institutional Review Board of the University of Washington. The study analyzing this dataset has a publication under review in [10]

4.1.1 Objects

We used the Yale-CMU-Berkeley Object and Model Set [52] for the objects in the experiment. Only a subset of these objects was used, in order to avoid using multiple objects with similar object geometries. Each object's geometry was remodeled in Unity. This subset, of which some are shown in Figure 4.1, contains 30 objects of varying object geometry and features.

4.1.2 Sensors and Equipment

This experiment requires the capturing of gaze data as well as kinematic data of the hand. To capture the gaze data, as well as run the experiment in Virtual Reality (VR), a HTC Vive Pro Eye (Vive Pro Eye, HTC Corporation), while the kinematic hand data was captured via a Leap Motion Controller (Leap Motion Controller, Ultraleap).We used the Tobii SDK (TobiiXR SDK, Tobii) eye tracking software.

4.1.3 Task

A VR program in Unity was created in order to capture the hand kinematics and gaze data. The experiment consists of the subject sitting at a desk while objects spawn in front of the

Figure 4.1: Examples of various objects used in this experiment. A range of object geometries are used, with repeated geometries having different scaling factors.

subject. A trial consist of an object spawning in a randomized position and rotation. A red indicator light informs the subject that the object is spawned, and the subject is instructed to reach for the object. When contact is made, the object disappears and the indicator light becomes green to inform the user to move to a pre-defined neutral position. At this location, the indicator light becomes blue and the subject is given one second to rest until the next object is spawned. Each data collection session ranged between 300 and 400 trials due to the randomization of objects' locations and orientations as well as corrupted trials. The set of objects used in this task is a subset of the objects described above.

4.1.4 Data Recording

During the reaching phase of each trial, each frame of gaze data and kinematic hand data is being recorded according to the fixed frame rate in Unity. The gaze data, shown in Figure 4.2, consists of a 100x100 depth cloud centered around the gaze point. The center of the gaze was also tracked over time, essentially creating a "heat map" of the subjects' gaze. The hand kinematic data consists of joint positions and rotations for all joints in the hand and wrist.

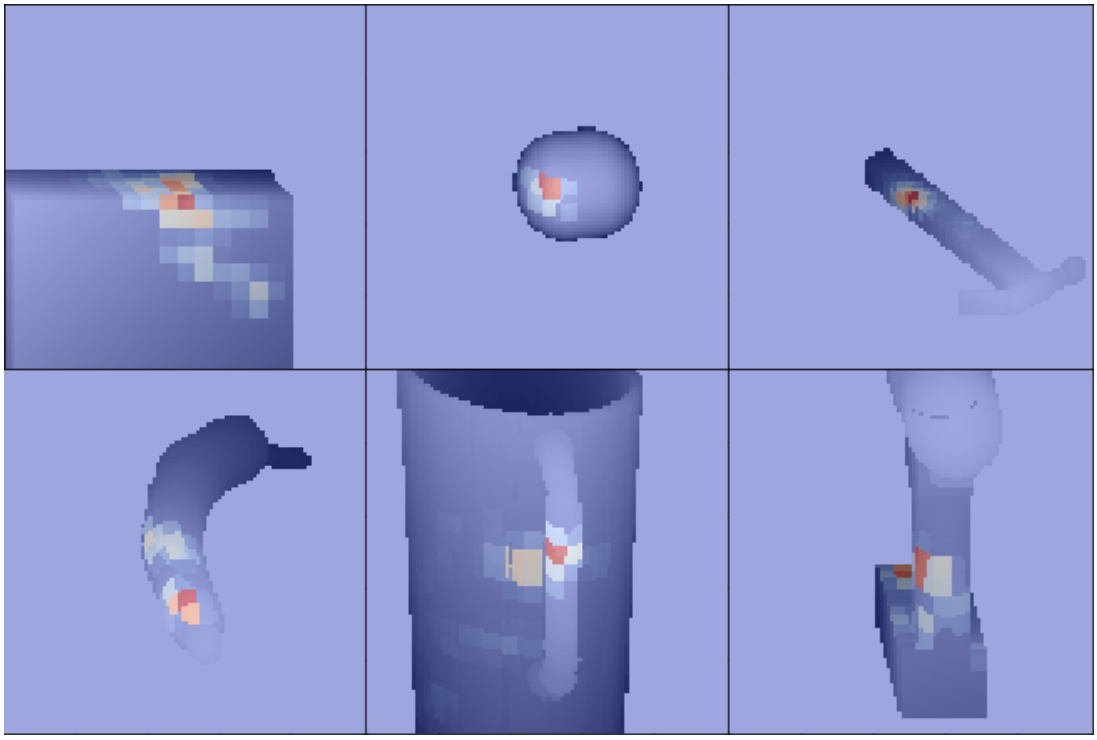

Figure 4.2: Gaze visualizations on several objects. The depth data provides for general object geometry, while the heat data provides information on locations where the gaze lingered.

4.1.5 Data Processing

A trial's data file is processed in a Python program, where five frames are selected as the optimal frames for the trial. The median of the gaze centers are calculated and the frame which has a center that is closest to this median is the one most likely to contain the correct center of gaze. The frame is then split into a depth image of the object, as well as a heat map of gaze lingering, which are both normalized between 0 and 1.

The hand kinematics are selected as the last fives frames before the trial ends, as this should be the most accurate to the desired grasping pose. The joint data is transformed using sine and cosine operations to remove discontinuities between 0 and 359 degrees.

The raw wrist rotation data was processed into 4 classes. These classes were chosen empirically from the data that was collected, which showed two peaks in the wrist pronation/supination and the wrist flexion/extension, visible in Figure 4.3. The four classes represent all possible combinations of the peaks, visualized in Figure 4.4.

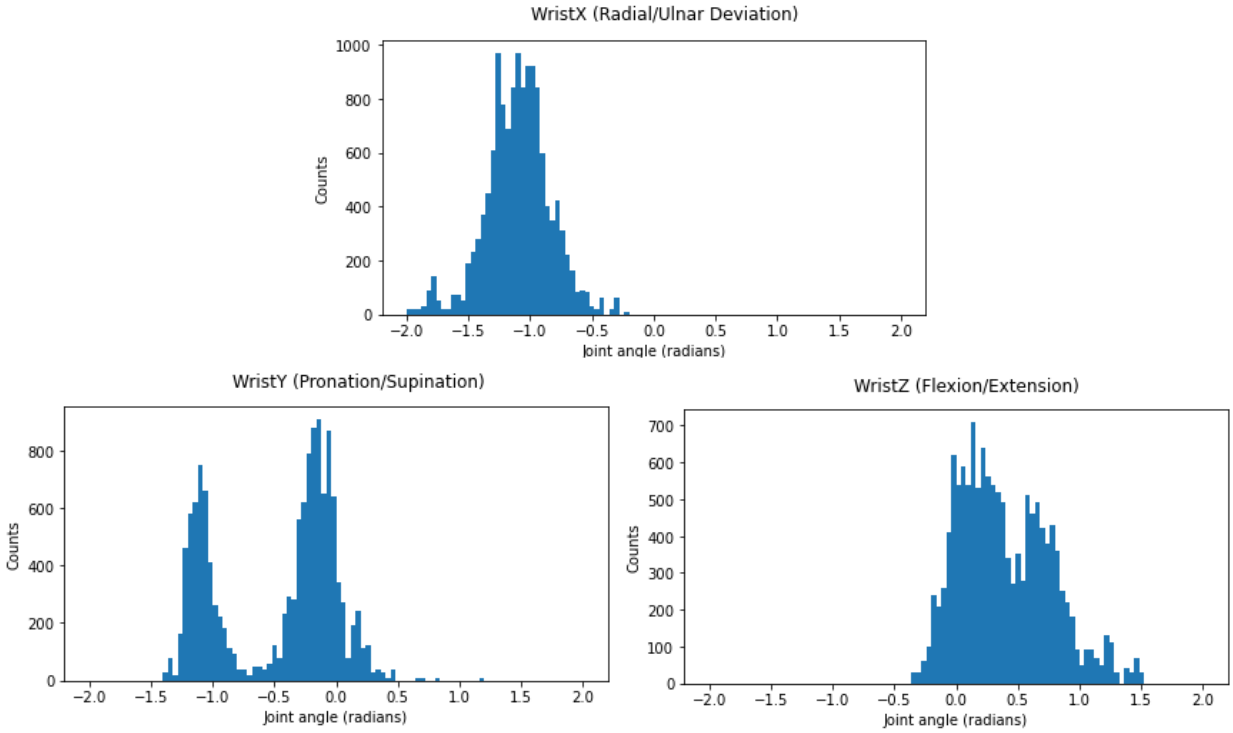

Figure 4.3: Distributions of the wrist angles in the data. In the x-direction (radial/ulnar deviation), there is a single peak, and in the y-direction (pronation/supination) and the z-direction (flexion/extension) there are two peaks.

The data is also split into various subsets of data for training, validation, and testing. The test set consists of the complete set of data from one subject which is not included in the training or validation set. The remaining data is split between training and validation by randomly choosing 10% of the data to be in the validation set and the rest in the training set.

4.2 Model Development

We aim to develop a model that can take frames of gaze data as input and output a prediction for the user's wrist rotations.

4.2.1 Feature Selection

The input and output features will be selected from the dataset described in the previous section. These features are chosen in a way that fits our goal the most. The first consideration in the problem is the choice to view this problem as a regression or as a classification problem. While regression seems the most natural choice to a problem regarding human kinematics, we find that most grasps in the dataset have specific values in the distributions that they tend towards.

The input features chosen for the task are the entirety of the collected gaze data. After processing this data, this consists of two channels of a 101x101 square matrix of gaze data. The two channels correspond to the depth image of the object and the time-tracked heat map of the gaze over the object. Both channels are normalized between 0 and 1 to help with the model learning.

The output features chosen for the task are the two wrist rotations with the most deviation in the choice. For the adduction and abduction of the wrist, we see in Figure 4.3 that there is not much variance in the angles of this joint, which makes sense for most common grasping tasks. However, there is great variance and distinct peaks in the pronation/supination and the flexion/extension. This supports our choice of posing this as a classification problem and our reasoning for using these two wrist rotations. While data for all joints of the hand are recorded, we only use the joint data for the wrist because this aligns the most with our object of predicting wrist rotations.

4.2.2 Baseline Selection

To select a baseline algorithm for this task, we look at the goal of predicting the class of optimal wrist rotations given an input frame of data. Knowing this, we find several methods to choose from.

The most simple approach consists of choosing the most common class in our data. This ensures that the most represented class will always be the chosen. Given the class imbalance in the dataset, this will ensure that we have a fairly high accuracy, but fails to properly represent any of the other classes, which may have been a more optimal choice. This will serve as the first baseline to compare our algorithm to, and will be referred to as the most-frequent class (MFC) baseline.

To improve upon this simple solution, we may propose a second baseline algorithm which uses regression to predict the joint angles, then post processing this into the classes. As our data seems to have multiple distributions in the two rotations, a reasonable choice would be to use a Gaussian models. Typically this approach works the best when the distributions are Gaussian, which Figure 4.3 shows exist in our dataset. The predicted wrist rotations are

Figure 4.4: Visualization of how each class looks to the user. Each class has a flexion/extension component and a pronation/supination component.

then calculated into the classes in the same way that the target data is processed into classes for comparison. This will serve as the second baseline to compare our algorithm to, and will be referred to as the Gaussian Mixture Model (GMM) baseline

4.2.3 Final Algorithm Selection

Ultimately, the baselines will not perform very well, as there is information in the input data that the baselines will fail to account for. To show a reasonable performance in algorithm, we propose using a Convolutional Neural Network (CNN). A large body of research in image

processing has found that the CNN architecture is particularly good at most image processing tasks [53, 54, 55]. The defining aspects of the CNN is their alternating of convolutional layers and pooling layers. The convolutional layers downsample the image and learn features [46], while the pooling layers reduce the dimensions of the feature maps. An example of this can be seen in Figure 4.5, which shows the outputs of the filters in the final model on a given image.

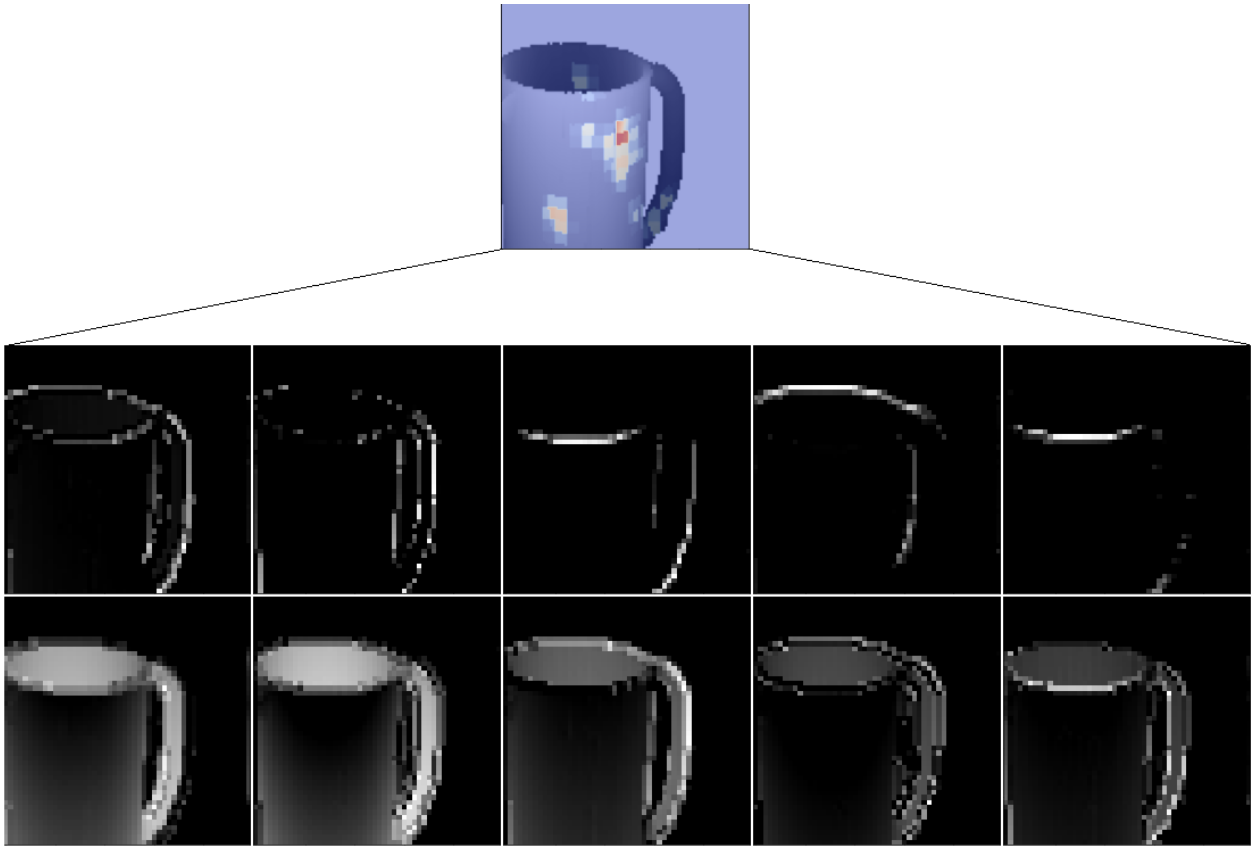

Figure 4.5: Examples of a features learned from a single gaze image in the first convolutional layer in the architecture. Filters learn various features, like edges or points of contact. A subset of filters are shown here

The final architecture, shown in Figure 4.6, consists of 2 layers of convolution and pooling. The first 2D Convolutional layer has 32 filters with filter size of 5 and a kernel size of 3. This is followed by a Max Pool 2D layer with kernel size of 3. The second 2D Convolutional layer has 64 filters with a filter size of 3 and a kernel size of 3, and is followed by a Max Pool 2D layer with kernel size of 3. The output of this layer is flattened, and input into a fully connected layer with a size of 100. This outputs to a final Dense layer for the output

with 4 classes and a softmax activation function. Dropout layers were implemented after the Flatten and the Dense layers in order to counteract overfitting.

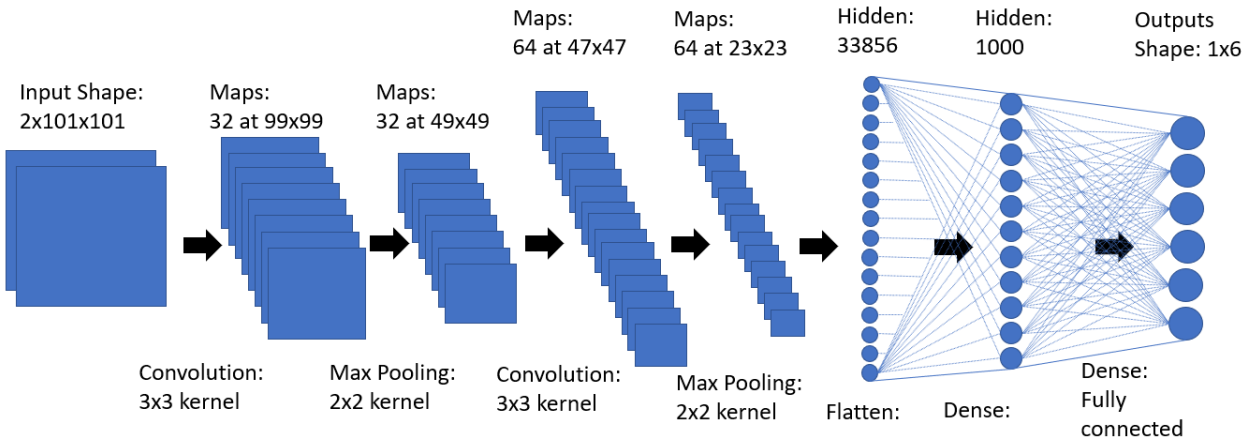

Figure 4.6: Architecture of the Convolutional Neural Network trained for this task. The layers of convolution and pooling are followed by two fully connected layers to a softmax output layer.

4.2.4 Hyperparameter Tuning

Hyperparameter turning was performed using all data except one subject with a split of 90/10 for training/validation. The outcomes of various testings are shown in Table 4.1. Several configurations of shallow networks were tested, as well as several deeper configurations. The shallow configurations generally performed worse than the deep networks and the conventional CNN architecture performed best. We found that the architectures performed best when the number of convolutional and max pooling layers was 2, and the size of these layers performed best when the convolutional layers have an increasing number of filters. The best choice for number of filters were 16 and 32 for the first and second layer, respectively. The kernel size and pooling size were parameters that I found to have less of an effect on the performance, so they are constant between layers. All models were trained using the "Adam" optimizer, and early stopping was used to determine when convergence was reached before over-fitting behavior was shown.

The final hyperparameter values are described in Table 4.2

Table 4.1: Performance metrics of various hyperparameter tunings for classification of the wrist class.

Number of Convolution Layers	*1*	*2*	*3*	*4*
Accuracy	69.6%	75.2%	73.1%	72.1%
F1 Score	48.40	54.91	52.06	49.22
Number of filters (First and Second Layer	*8,16*	*16,32*	*32,64*	
Accuracy	70.92&	72.11%	74.41%	
F1 Score	49.56	48.61	53.04	

Table 4.2: Hyperparameters tested and chosen for the final architecture.

Concatenated	
Optimizer	Adam
Learning Rate	0.001
Dropout Rate	20%
Hidden Layer Activation Function	ReLu
Batch Size	64
Epoch	10

4.2.5 Test results

We implemented the CNN algorithm to predict the class of wrist rotations from gaze data of an unseen subject, which was held out from the dataset. The results show an accuracy of 72.9% and a F1 score of 49.79. This accuracy and F1 score is slightly lower than, but comparable to, the highest accuracy and F1 scores found during validation.

Table 4.3: Performance metrics of the final implementation of the model and the baseline models on the test dataset.

Model	*MFC Baseline*	*GMM Baseline*	*CNN Architecture*
Accuracy	57.7%	59.6%	72.9%
F1 Score	18.30	22.95	49.79

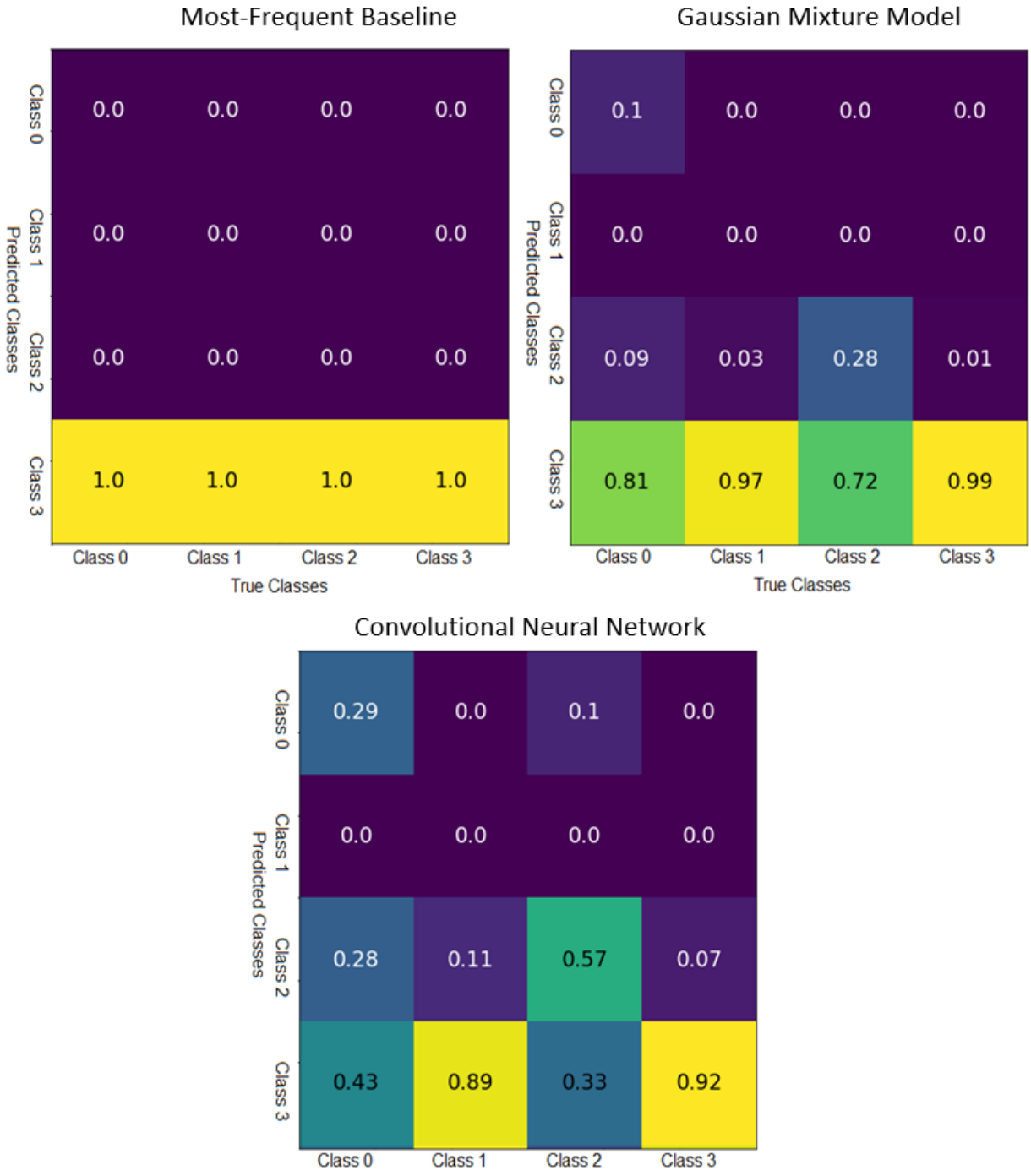

Figure 4.7: Confusion matrices of the models on the test set of data. These confusion matrices shows which the percentage of each class predicted correctly, and which class is confused the most with other classes.

The confusion matrices in Figure 4.7 show how the classes are mistakenly chosen over one another for each model, as well as the how each model predicted each class in the set.

4.3 Discussion

4.3.1 Performance metric, class imbalance, and confusion

The performance metrics chosen to evaluate the model were the accuracy and the Macro F1 score. Accuracy can be a good indicator of the general performance of the model, but cannot capture small successes and failures between classes. The Macro F1 score was chosen to counteract this, and show the success of the model in predicting each class specifically. As we see in Tables 4.1 and 4.3, an change in accuracy does not necessarily imply a corresponding change in Macro F1-score. The MFC baseline model can achieve a high accuracy since the dataset is imbalanced, but this does not mean it is a good model, as it has not learned how to predict any of the other classes. The Macro F1 score is a good representation of this, as it takes into consideration the success of the model to predict all classes in a representative manner.

The dataset has a class imbalance with the most frequent class being 58% of the data and the least frequent class comprising less than 1 percent. The most frequent class is Class 3, while the least frequent class is Class 1, which are supination/extension and pronation/extension, shown in Figure 4.4. For this reason, we use the Macro-F1 score when comparing the different algorithms to find the best performing architecture. In a practical setting, the least frequent class may be one that is used for very few object orientations, so despite being optimal in these scenarios, an alternate class may be only slightly less optimal. For this reason, we did not find it necessary to eliminate this class, as the user study may show that both classes are similar in efficiency. This may result in a lower model performance when measured by accuracy or macro-F1 score.

We find that the classes that are most frequently confused are Class 2 and Class 3. Referencing Figure 4.4, we can see that these classes represent the wrist supination, but differing flexion/extension. This suggests that when the wrist is supinated, the flexion or extension of the wrist does not have a significant impact on grasp success. We also find that most of Class 1 scenarios were incorrectly predicted to be Class 3. This is because Class 1 is the least frequent class, so the model most likely did not learn when to predict these and instead predicted the most frequent class. However, the model predicted Class 1 a non-zero amount of times, which indicates that there are some specific scenarios for Class 1, and the model is able to start to learn these nuances.

4.3.2 Performance on unseen subjects

Our results show that the algorithm is capable of performing comparably on unseen subjects, as shown by our test data, which was made of an unseen subject and on which the algorithm performed comparably to the validation data. This suggests that the algorithm is able to learn despite variations between subjects, and is valuable as a general algorithm, not a subject-specific algorithm. A general pre-trained algorithm may be useful for devices that are widespread in distribution, unlike other algorithms that require training and learning to be done by the prosthesis user. While the algorithm is shown to be general, it will still benefit from more training data with more subjects to improve performance. It is also important to test on a wider variety of subjects, as our sample of subjects shared similar traits, like age range and level of impairment. It is entirely possible that individuals with other characteristics may experience different performance from the algorithm.

4.4 Limitations

This study found that a model can be developed for predicting the wrist rotations of subjects. However, the algorithm was trained on data collected from intact subjects, and the amount of training data is not large, especially for an architecture which prefers a considerable amount of data, like the CNN. The size of the dataset used may lead the algorithm to issues of overfitting, and the small amount of intra-subject variability in the dataset may have led the algorithm to learn things that are not generalizable to the whole population. The goal of this algorithm is to be able to predict the most natural motions for reaching tasks to remove the cognitive load from a prosthesis user. However, the subjects in this dataset are not PWA, and there might be a difference in the way that PWA approach reaching tasks in terms of gaze.

4.5 Future Work

To address the issues mentioned above, a more comprehensive and representative dataset could be collected in order to train the model to be more generalizable and have better performance. Future implementations of this algorithms could also test different architectures, like implementing autoencoders or siamese CNN architectures. The dataset before any processing includes a time-series of gaze data and hand kinematic data. A future study intending to explore the reaching task as a sequence could use this data in order to better predict the final hand kinematic data by using an RNN or an LSTM architecture. The improved algorithm could be robust towards saccades or gaze misdirections by considering the entire sequence of gaze frames.

The dataset collected and used in this study is comprised of the entirety of the hand joint kinematics, not just the wrist. While we approach the problem of wrist rotation, a future exploration of the other hand kinematics could be investigated on this dataset in order to perform grasp classification, or even more ideally, mode-free full-hand kinematics prediction.

4.6 Conclusion

This aim was successful in developing a classification algorithm that can predict the optimal orientation of the wrist in a reaching task. The effectiveness of this algorithm is not entirely captured by the performance metrics, and so a user study will be performed to see the effect of this algorithm on compensatory movements and reaching task success. This type of predictive algorithm is the center of a predictive prosthesis controller, and gives insight into natural human movement and ways to restore this to users of upper limb prostheses.

Chapter 5

ANALYSIS OF PREDICTIVE WRIST CONTROL ON VIRTUAL PROSTHESIS

5.1 User Study

To test the performance of the algorithm, we implement it in a virtual environment and measure its performance in object manipulation tasks. Subjects in this study consisted of 5 right-handed subjects (4 males, 1 female) and 1 left-handed subject (1 male).

5.1.1 Sensors and Equipment

This experiment requires the capturing of gaze data as well as kinematic data of the forearm and upper arm. To capture the gaze data, as well as run the experiment in Virtual Reality (VR), a HTC Vive Pro Eye (Vive Pro Eye, HTC Corporation), while Vive Trackers (Vive Tracker, HTC Corporation) were used to capture arm kinematic data. We used the Tobii SDK (TobiiXR SDK, Tobii) eye tracking software and a Leap Motion Controller (Leap Motion Controller, Ultraleap) was used for a portion of the study.

5.1.2 Task

In a virtual environment, subjects perform object grasping and manipulation tasks. Before any tasks, the subject performs a neutral pose calibration. Three configurations of the subjects' virtual hand were used: a full virtual reconstruction of the fully tracked hand (natural condition), a virtual prosthesis with a static-wrist power grasp (wrist-locked condition), and a virtual prosthesis with the wrist predicted by our algorithm (wrist-predicted condition). In each configuration, a trial starts with an object spawning in a pseudo-random location and orientation. the subject then places their hand in the pre-determined neutral position, during which a yellow indicator light informs the subject that they are getting in the neutral position. Once neutral position has been reached, the indicator light changes to red which informs the subject to attempt to reach for the object. The subject then picks up the object, whether by the physics of the real hand or by the open/close functionality of the virtual prostheses, and moves it to a pre-determined location. The subject then opens their hand to

drop the object in the box at the final location and the indicator light turns blue to indicate that a successful trial has been done. The subject then has a second to rest before the next object is spawned and a new trial begins.

In the static power grasp configuration, the rotation of the virtual prosthesis is controlled solely by the movement and rotation of the forearm and upper arm trackers. The second virtual prosthesis configuration includes the model which predicts the pronation/supination and flexion/extension of the wrist, based on the user's gaze on the object.

Figure 5.1: Demonstration of the predictive model. The object's orientation allows for multiple grasps, and the model predicts the wrist's rotations based on the user's gaze on the object, indicated by the white "x" marker.

For the prosthesis configurations, the virtual prosthesis movement was controlled by Vive Trackers on the forearm and upper arm. A demonstration of the model in use is shown in Figure 5.1 In each configuration, the objects are placed in the same configurations and revealed to the subject in the same order. The set of objects used in this task is a subset of objects previously used in the previous task, as well as new objects that were not used in the previous task.

5.1.3 Data Recording

For the user study, we record various sets of data. The first set of data is the gaze frames during each trial. This set is not used for data analysis, but rather for visualizations when looking through the comparisons of the trials. The second set of data recorded is the 3D orientations and locations of the forearm and the upper arm. These are recorded from the Vive Trackers placed on the subject, and are used for the main data analysis later on. The final set of data that is recorded are the times taken for different parts of the trial.

5.1.4 Analysis and Performance Metrics

The three performance metrics that we look at are task times, range of motion for elbow and shoulder joints, and a prosthesis performance metric (PHAM). There are two subcomponents of total task time: preparation time and manipulation time. Preparation time is the time from the neutral position to the time when the object is grasped, while manipulation time is the time taken to move the object from the spawn location to the destination location. The task time is the total time of the trial, or the sum of these two times.

Compensatory movements for a upper-limb prosthesis are generally shown in the range of motion of the elbow joint, shoulder (shoulder) joint, and the chest/torso. In this study, we disregarded the range of motion of the chest from analysis. The decision was made due to the nature of the task that we had subjects perform. When picking up objects on a table, most chest movement from the resting position will occur due to the placement of the object on the table, regardless of the prosthesis configuration.

A global coordinate system was defined for the forearm, upper arm, and chest, and calibration was used to determine the initial positions of the elbow joint and the shoulder joint. This calibration recorded the neutral pose, where all values of angles are calculated relatively. While the recording software was able to capture absolute angles of trackers, it gives no information about the relative angles of these trackers. These angles must be calculated with respect to the neutral pose. The neutral pose had the subject sit with the shoulders relatively relaxed and had the elbows bent close to 90 degrees.

The elbow joint angles were calculated from the movement of the forearm tracker relative to the upperarm tracker and the shoulder joint angles were calculated from the movement of the upperarm tracker relative to the chest tracker. The flexion and extension angle was the only joint angle calculated for the elbow joint, while joint movement in the sagittal plane (flexion and extension) and coronal plane (abduction and adduction) was calculated for the shoulder joint. Humeral rotation was not calculated, due to the lack of movement in this plane during reaching tasks. The range of motion of each joint is calculated by finding the difference between the maximum and the minimum joint angles.

For overall performance of the virtual prostheses, we use a modified version of a metric developed by Yerrabelli et al [56] called the Prosthetic Hand Assessment Measure (PHAM). This measure is a weighted sum of the 2D translational displacement of the hand, 3D deviation of the chest, and 3D deviation of the shoulder. We substitute the 3D deviation of the chest with the 1D deviation of the elbow, as this may prove to have more information than the chest does in our selected reaching tasks and the 2D translation displacement of the hand with the 2D translation displacement of the forearm. We also take out factors associated with aspects of the study that were not present in our study, like the normalization factor and the completion rate, which is replaced by completion time. Overall, the lower the value

of the PHAM metric, the better that the configuration works as a useful prosthetic hand.

$$P = \frac{\left\|\overrightarrow{\lambda}\right\| + \left\|\overrightarrow{\delta_E}\right\| + \left\|\overrightarrow{\delta_S}\right\|}{\eta}$$

where the deviation and displacements are calculated, respectively, as:

$$\overrightarrow{\delta_X} = \sum_{n=1}^{N} \left| \begin{bmatrix} \varphi_X \\ \theta_X \\ \psi_X \end{bmatrix}_n - \begin{bmatrix} \varphi_X \\ \theta_X \\ \psi_X \end{bmatrix}_{n-1} \right|$$

$$\overrightarrow{\lambda} = \sum_{n=1}^{N} \left| \begin{bmatrix} \vec{x} \\ \vec{y} \end{bmatrix}_n - \begin{bmatrix} \vec{x} \\ \vec{y} \end{bmatrix}_{n-1} \right|$$

For the analysis of all the performance metrics, we calculated the medians and averages over all subjects and all trials within each subject's session.

5.2 Results

The machine learning model was able to predict valid choices for wrist rotation given frames of gaze data. In this section, we report the performance of the algorithm as a model, and compare the impact that the model has when implemented in a virtual prosthesis.

5.2.1 Task Times

The task times, divided into preparation, manipulation, and total times, are shown in Figure 5.2

Across all subjects and trials, the real hand configuration has a median preparation time of 2.44s, a median manipulation time of 1.58s and a median total task time of 4.02s. For the same subjects and trials, the static power grasp configuration has median times of 3.81s, 1.75s, and 5.56s for preparation, manipulation, and total task time, respectively. The predictive power grasp configuration has 3.09s, 1.64s, and 4.73s for preparation, manipulation, and total task time.

The real hand configuration serves as a baseline for natural, intact behavior in an object reaching task. In line with this, we find that the real hand configuration has the smallest median times for preparation (2.44s) times, whereas the wrist-predicted configuration has the smallest median times for manipulation (1.64s).

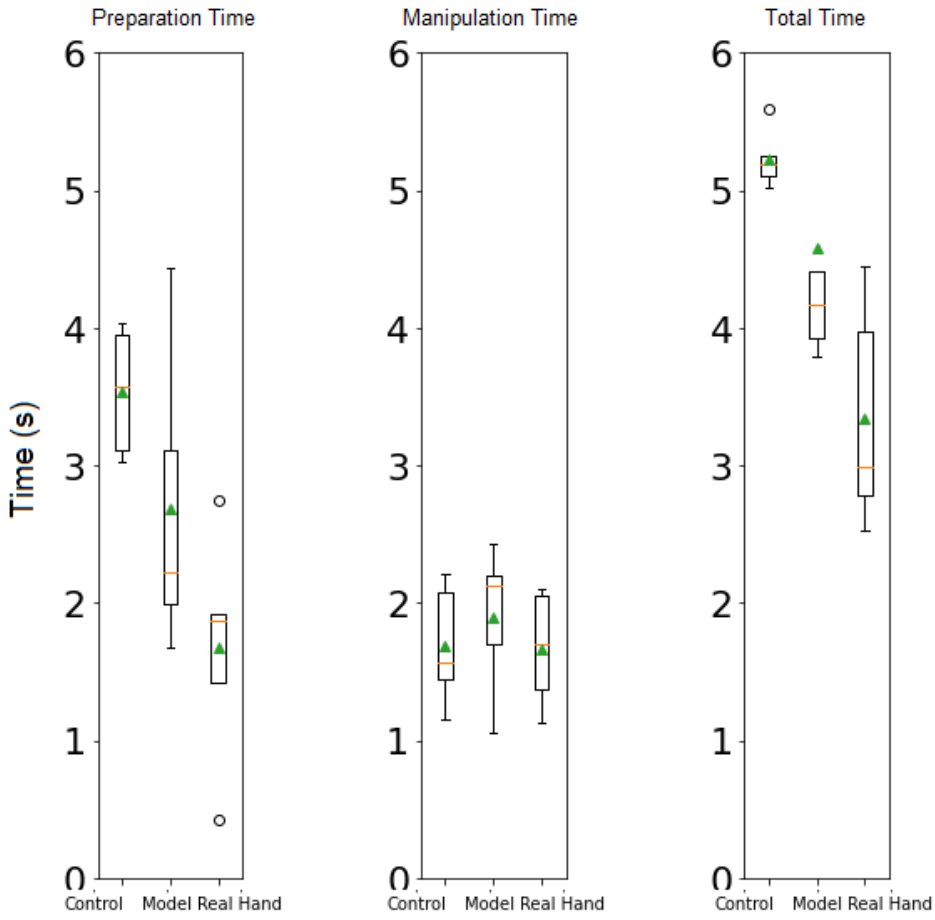

Figure 5.2: Box and whisker plot of the times for each configuration. There is little change between configurations for manipulation time, while a reduction in preparation time is shown.

We test the hypothesis that the wrist-predicted configuration shows a statistical improvement from the wrist-locked configuration in both preparation time and manipulation time. For all significance testing in this project, we use a 95% confidence interval, which corresponds to an α value of 0.05. We run a one-tailed paired t-test to compare the two experimental configurations, and the p-values for both preparation and manipulation times are shown in Table 5.1. We see that in preparation time, $p < \alpha$, which indicates a statistically significant improvement. In manipulation time, we see that $p > \alpha$, which indicates that there is no statistically significant improvement in manipulation time.

Table 5.1: P-values for each time in all configurations

	P-value
Preparation Time	0.0214
Manipulation Time	0.25944

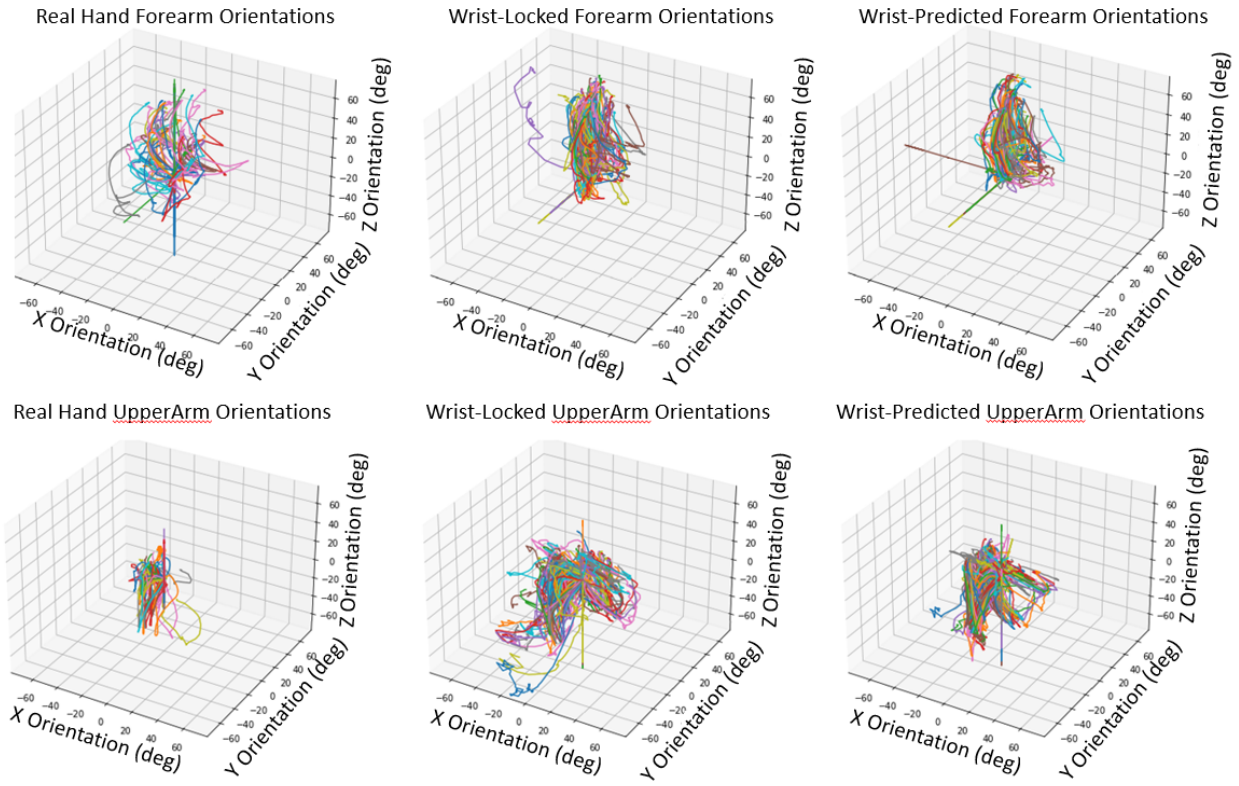

Figure 5.3: Overlaid 3D trajectories of all trials for a random subject. These trajectories are visualizations of the paths that each arm segment makes, and differences in paths can be seen in the wrist-locked and the wrist-predicted configurations.

5.2.2 Compensatory Movements: Elbow

The forearm orientations are shown in the top row of Figure 5.3. The range of motion of the elbow joint is calculated by finding the difference between the maximum and the minimum joint angles in the sagittal plane. The peak abduction and adduction is also calculated for the reaching phase of each trial, and these are displayed with the range of motion in Table

Table 5.2: The peak joint angles and the range of motion of the elbow

Configuration	*Peak Flexion*	*Peak Extension*	*Range of Motion*
Wrist-locked	32.56	1.00	30.26
Wrist-predicted	25.95	1.00	23.73

The range of motion (ROM) in the elbow joint varies largely in both virtual prosthesis configurations. This may be because reaching from the neutral position encourages extension of the elbow for the hand to reach the object. We find that the range of motion in the wrist-locked prosthesis configuration is larger than the range of motion in the wrist-predicted configuration. The wrist-locked configuration had both larger peak flexion and peak extension, meaning that the ROM increase did not come from solely more flexion or more extension.

5.2.3 Compensatory Movements: Shoulder

The shoulder orientations are shown in the bottom row of Figure 5.3. The range of motion of the shoulder joint is calculated by finding the difference between the maximum and the minimum joint angles in the sagittal and coronal planes. The peak abduction and adduction in these planes is also calculated for the reaching phase of each trial, and these are displayed with the range of motion in Table 5.3.

Table 5.3: The peak joint angles and the range of motion of the shoulder

Configuration	*Peak Flexion*	*Peak Extension*	*Sagittal Range of Motion*
Wrist-locked	44.39	18.86	14.06
Wrist-predicted	46.01	29.02	12.02
Configuration	*Peak Abduction*	*Peak Adduction*	*Coronal Range of Motion*
Wrist-locked	46.63	1.00	37.23
Wrist-predicted	32.72	0.94	24.81

We find that the range of motion in the sagittal plane for the wrist-locked configuration is larger than the motion in the same plane for the wrist-predicted configuration. For both planes of motion peak ROM values, a Kolmogorov-Smirnov test was performed between the wrist-locked configuration and the wrist-predicted configuration. The abduction and

adduction ROM of the wrist-predicted configuration has a smaller distribution than the wrist-locked configuration, and with a p-value of 0.0344, this constitutes a statistical significance. The flexion and extension of the shoulder do not seem to have a statistically significant change (p-value = 0.43).This is most likely because the objects are in front of the user, and moving from neutral pose to the object is going to be similar each time in the coronal plane. A visualization of the shoulder movements in the abduction and adduction plane is shown in Figure 5.4

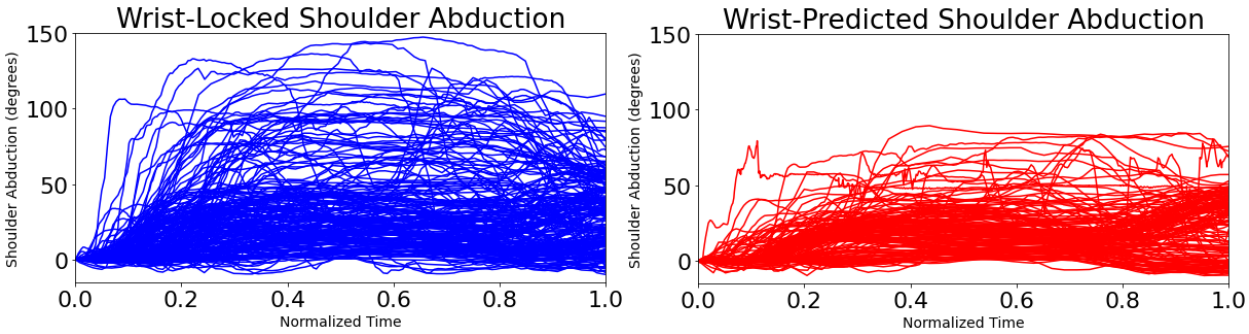

Figure 5.4: Overlaid orientations over normalized time for the shoulder abduction and adduction plane of movement. A distribution of larger ranges of motion for the wrist-locked scheme is seen, as compared to a smaller distribution in the wrist-predicted scheme.

5.2.4 Prosthesis Performance Metric

The PHAM performance metric described earlier is calculated and shown for the wrist-locked and the wrist-predicted configurations in Table 5.4

Table 5.4: Average PHAM values for each configuration

Configuration	*Average PHAM*
Wrist-locked	8.131
Wrist-predicted	4.541

5.3 Discussion

This study shows the validity of using gaze-based systems for wrist prediction to improve performance of basic prostheses. We find that wrist pronation/supination and flexion/extension are able to be predicted from gaze-centered vision data. When this algorithm is implemented into a prosthesis controller, the performance is increased relative to a static power grasp prosthesis. These results are in line with the idea that gaze plays a role in hand preshaping in object manipulation tasks, as well as the idea that wrist degrees of freedom in upper limb prostheses contribute to increased usability and lower prosthesis rejection rates. Vision captures object geometry and information about the object like relative scale, but centering the vision on gaze may provide additional information about intent in a reaching task. Though intent was not explicitly tested in our experiment, eye-tracked gaze provides granular information regarding grasping intent, which becomes valuable in non-laboratory situations where multiple objects and affordances are almost always present.

The purpose of this study was not to compare against the state-of-the-art accuracy of grasp prediction, but rather to demonstrate the value of incorporating gaze-centered vision into grasping controllers. The results also focus on emphasizing the value of degrees of freedom in the wrist when using prostheses. Robotic grasping studies tend to focus solely on success of the robotic grasper to grasp the object. However, in order to assess the true value of predictive models in prosthetic controllers, they need to be evaluated in a full human-in-the-loop system. To our knowledge, there are no studies which compare prosthesis controllers with all of the same performance metrics.

A fundamental groundwork for robotic grasping using vision and feedback-based learning exists in previous studies. In works focusing on the approach of the hand to the object, Rombokas et al showed that orthogonality of the hand to the object faces provided enough information to calculate successfully the angle of approach [57]. While focusing solely on one aspect of robotic grasping, this study shows that angle of approach can be analytically calculated through depth cloud vision. Furthermore, Herzog et al previously developed an algorithm to synthesize grasps based on object geometries and specific features of objects, which could extend to unknown objects [58]. Object features, recorded as grasp-templates, are learned and applied to later, unknown objects to calculate a suitable grasp. In studies using a depth camera, like Saxena et al, success rates dropped on the chosen first choice grasps when the depth camera only had one point of view of the object [41]. Future studies mentioned by Saxena et al propose to avoid looking at object models at all and instead focus on the identify only viable points of contact on an object. [41] Their approach consists of an algorithm that takes multiple pictures of the object and triangulates the most likely points of contact and uses force-closure gripping to pick up objects. Furthermore, this study proves that avoiding a 3D reconstruction of the object's model can prevent issues that come from reflective or transparent object reconstruction.

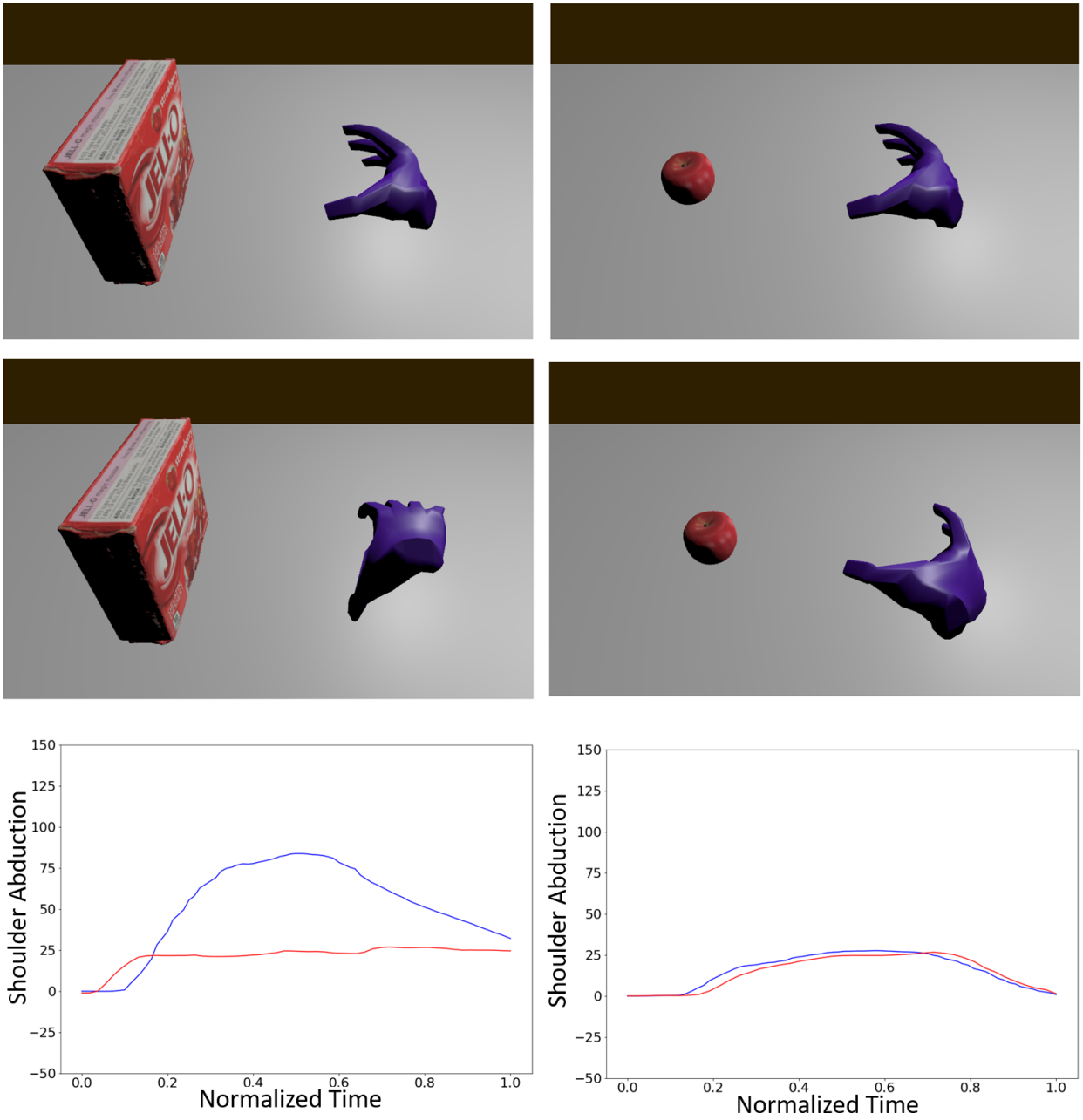

Figure 5.5: A comparison between the response when the predictive model made a large difference on the left and the response when the predictive model did not make a difference, on the right. The y-trajectory, a simple measure of shoulder trajectory in this task, is shown respectively.

We show that there are statistically significant improvements in the performance metrics when using our wrist-predictive system against a wrist-locked configuration. These improvements are more pronounced in certain trials over others. The trials where there is no improvement are trials in which the static power grasp is sufficient to grasp the object. For example, in Figure 5.5, we show a trial with significant improvement and a trial with no improvement. In the case of no improvement, the wrist-locked configuration would have the same shape and follow the same trajectory as the wrist-predictive configuration. In the case of significant improvement, it is clear that the wrist-prediction allows for an easier grasp, while the wrist-locked configuration must perform compensatory or strenuous extra movements to achieve a valid grasp on the object. The choice of using a power grasp for comparison was driven by our knowledge that a power grasp with the whole hand is one of the most common grasp choices for everyday object manipulation. If a different grasp was chosen for comparison, it is likely that a smaller amount of trials would show little improvement. However, in the cases of no improvement, it is important to note that there is no decrease in performance, and the wrist-predictive model is suited to handle these cases as well. We can in Figure 5.6 a histogram of the ranges of motion to further see that in general, the model reduced compensation significantly. These distributions reveal that the model was helpful for most cases, although there were still some object configurations which did not have a difference between the configurations.

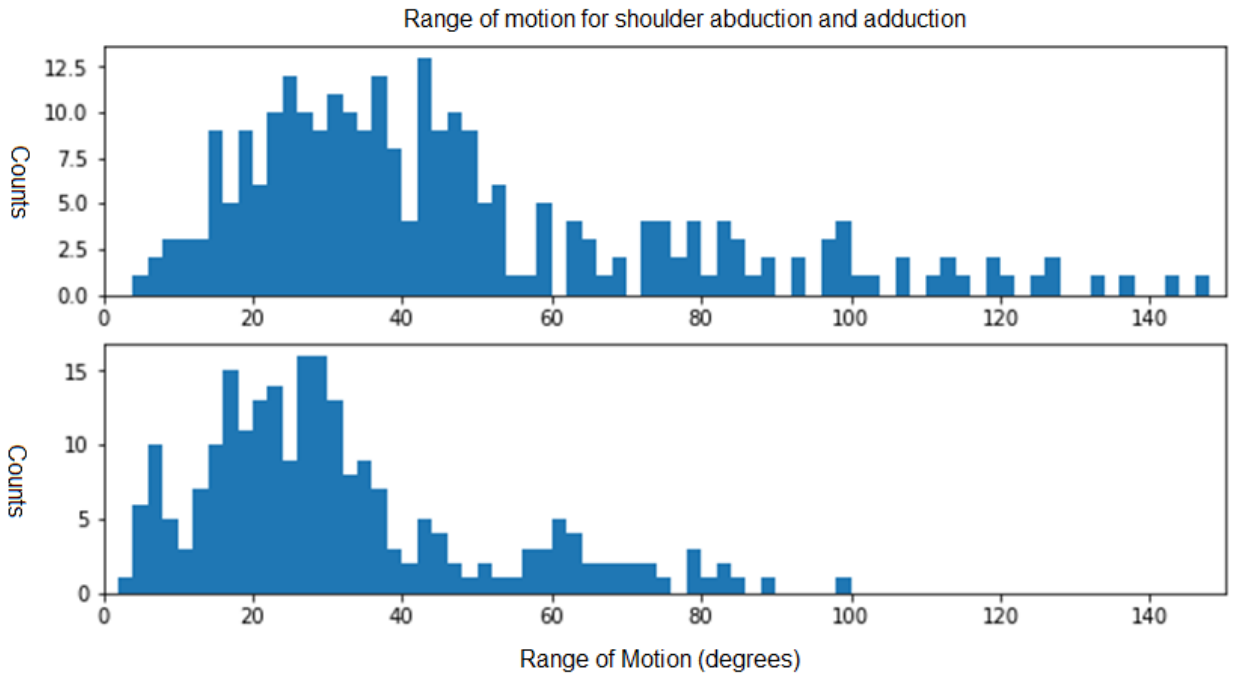

Figure 5.6: Counts for all trials over the wrist-locked configuration and the wrist-predicted configuration.

A motivation of using a predictive model was to reduce the cognitive load on the user of the prosthetic hand. We found that when using wrist-locked configuration, there was a statistically significant increase in shoulder compensatory movements, when compared to the wrist-predictive model. While this effect is found in other studies [59], the cause is undocumented. We believe that this effect is due to the increase in cognitive load when having to perform compensatory shoulder movements. These compensations are unnatural to most subjects, and so by also implementing a different degree of freedom (elbow extension/flexion), the subject may focus better on task execution. The idea of a larger cognitive requirement in the wrist-locked configuration is also supported by the larger task times for this configuration. A larger task time indicates a higher task difficulty or complexity [60], and since trials were identical across both configurations, the most likely reason for this increase in complexity was the use of extra shoulder movements. The presence of the extra complexity may have caused some subjects, intentionally or not, to use more movement in the elbow to reduce complexity to a comfortable level.

The results show that there is a statistically significant difference for shoulder range of motion in the abduction and adduction plane between each configuration. The majority of the difference comes from the change in shoulder abduction (median peak angle is 50 degrees) rather than adduction (median peak angle is 3 degrees), which can be explained by the neutral pose during calibration. These angle value are calculated with respect to the neutral pose, where the value of the shoulder sagittal plane angle is 0 in the neutral pose. Subjects at neutral pose have low degrees of adduction, hence further adduction is unlikely. However, the flexion and extension of the shoulder are not statistically different between each configuration. This is most likely because the objects are in front of the user, and moving from neutral pose to the object is going to be similar each time in the coronal plane.

5.4 Limitations

A limitation of this study is the amount and type of data in the training dataset. The model architecture of the convolutional neural network is known to require a considerable amount of data to train. This is to avoid overfitting because deep CNNs have large amounts of parameters which can overtrain quickly on a small amount of data. While the model was shown to perform well enough in the user study, the amount of error in the model performance would not be adequate for use in a prosthetic controller used by an actual PWA. Furthermore, the data is imbalanced between the different poses, with one pose comprising of over 51% of the dataset and the smallest class comprising under 2% of the dataset. While this may accurately reflect the choices made by intact humans, the imbalance in the data can be a problem for evaluating the model. The smallest class has the smallest Macro F1 score and is underrepresented in the training, so the model will not perform as well when it comes to the smallest class. Though we do not attempt object classification or provide labels, it

is inevitable that our limited training dataset does not cover all possible object geometries. Furthermore, both the data collection and the user study use a virtual environment. Virtual reality artifacts like eye strain, extra weight on the head, and latency all have an impact on the subject, which may extend to the results of the study.

Furthermore, the data from the collection stage was sparse, as it was a frame of gaze centered around an object. Most frames had a margin of zero values, though the width of this margin on each side varied due to the gaze location on the object. This sparsity of the data can make it difficult for many other choices of architectures to learn [61], but this problem can be mitigated using slow pooling layers [62].This configuration consists of the use of many small (2x2) pooling layers instead of fewer large (3x3,5x5) pooling layers. This was tested in this design, but was not found to have significant improvement.

5.5 Future Work

In this study, we used a classification-based system for predicting the wrist rotations. An interesting comparison would be to perform a user-study with a regression based system. The presence of rotations in between the designated classes indicates that there are a small number of cases where regression could predict a more accurate angle.

We also used a relatively simple convolutional neural network due to the small size of the training dataset. This was done to prevent overfitting, which can happen to architectures with a large amount of parameters but a low amount of data. We not only expect to have much better performance of the model with more data, but also the opportunity to try different variations of the CNN architecture.

In this work, the data collected was centered around the gaze of the user. To further prove the impact of this gaze-centering, a better approach would be to collect additional data from a static point, similarly to robotic grasping experiments. With both of these datasets, a model could be trained for each and a more concrete comparison between the vision data and the gaze-centered vision models could be performed.

In the user study portion of the study, we used a Virtual Reality environment, which has drawbacks described above. Furthermore, the virtual nature of the objects and the prostheses cannot compare to the experience of using a physical system with real objects. Another implementation of our user study with a real prosthesis with wrist degrees of freedom would be useful to gauge the impact of gaze-centered vision in the wrist-predictive model. This implementation would allow for any subject, intact or PWA, to be able to use the system as long as their eyes are able to be tracked. A more comprehensive study in the physical environment would be able to further solidify our findings.

5.6 Conclusion

A model to predict wrist rotations from eye-tracked gaze was trained from a dataset collected from 6 subjects and evaluated with a user study. In the user study, the wrist-predicting model was implemented into a virtual prosthesis and compared against a wrist-locked prosthesis. The comparison was evaluated with task times, elbow and shoulder movements, and a prosthetic hand performance metric. We find a statistically significant improvement in task time and shoulder compensation when using the wrist-predictive prosthesis. Gaze-centered vision holds important information about the user's intent when reaching towards and object, and the prediction of wrist rotations has a big impact on the performance of low-complexity prosthetic hands. With further exploration and data, the performance and usefulness of the algorithm would become more evident.

Chapter 6

CONCLUSION

The goal of my research is to create predictive systems for assistive technology in people with amputations (PWA). I worked in data collection for lower limb stair descent tasks, and the majority of my research focuses on building and testing a system that can predict wrist rotations to help PWA with upper limb amputations with daily reaching tasks.

In the first project (Chapter 3), I explore various data collection methods for lower limb amputations for a project called *SmartStep*. During this project, I first assisted with the data collection using a different kind of motion capture, optical motion capture, and ground force plates in order to collect gait data of subjects responding to tactile feedback when walking. This data was used to explore the required time it takes for users to modify their gait after receiving haptic cues. The second portion of the project consisted of data collection of 101 intact subjects descending stairs. Motion capture systems and foot force sensors were used during this collection. This large dataset was used in order to predict the future foot placement of a user walking down a set of stairs. These projects interested me in the ability of algorithms that can predict human motion during daily tasks.

My next project (Chapter 4), was an investigation in how models can be used in order to assist reaching tasks in users of upper limb prostheses. In this chapter, I performed data collection on subjects in Virtual Reality reaching for objects, and recorded eye-tracked gaze data and hand kinematics. This data was used in order to create a model that could predict wrist rotations by using frames of gaze data. The model consists of a convolutional neural network and outputs classes of wrist rotations. The primary finding from this project was that a model of acceptable accuracy is able to learn, from gaze data, which class of wrist rotations mimic the most natural human choices.

The model was validated by a user study (Chapter 5), in which users were asked to pick and place objects using different configurations of a virtual prosthesis (real hand movement, a wrist-locked grasper hand, and a grasper hand with the wrist rotation predicted by the model). The performance of the prostheses were evaluated by looking at the completion times, the ranges of motion of elbow and shoulder, and a prosthesis performance metric. The main finding from this user study was that the model-predicted configuration was able to reduce total task times, and reduce the amount of shoulder movement, or compensation, that the user had to do, compared with the wrist-locked configuration. This indicates that

the model has value when implemented into a prosthesis with simple open-close capabilities.

The results from Chapter 3 show the value in using prediction for a lower-limb assistive device called *SmartStep*. This device uses a model that predicts the future foot placement and cues to the user to adjust their step accordingly. This device may be used in a real-world implementation with more consideration in the future. *SmartStep* may be useful for not only those with lower limb prostheses, but also with limb impairments, elderly people, and other users who have trouble with stair descent.

The results from Chapter 4 and Chapter 5 show that there is information in a subjects' eye-tracked gaze, which can be used to predict the most natural wrist rotations of the user through a model, called *Gaze2Grasp*. This model is useful for users with upper limb prostheses to more easily grasp and manipulate objects, as well as perform daily tasks. *Gaze2Grasp* would be able to be easily implemented into any upper limb prosthesis with wrist degrees of freedom with relative ease and minimal modifications. In the future, *Gaze2Grasp* could be modified for other purposes, like predicting full-hand kinematics for a grasping task, or for predicting wrist rotations in other tasks, like terrain navigation. In general, tasks which use vision heavily and require hand kinematic output could benefit from the use of an algorithm like *Gaze2Grasp*.

Overall, the prediction-based systems done in my research show that predictive, data-driven methods are robust and generalizable. The use and implementation of these methods in prostheses and devices lead to lower decision times, a lessened cognitive load, and decreased physical compensation. The prediction of smaller, more achievable aspects of tasks, like foot placements and wrist rotations, can help the user in impactful ways.

BIBLIOGRAPHY

[1] K. Ziegler-Graham, E. J. MacKenzie, P. L. Ephraim, T. G. Travison, and R. Brookmeyer, "Estimating the prevalence of limb loss in the united states: 2005 to 2050," *Archives of Physical Medicine and Rehabilitation*, vol. 89, pp. 422–429, 3 2008.

[2] L. Resnik, S. Ekerholm, M. Borgia, and M. A. Clark, "A national study of veterans with major upper limb amputation: Survey methods, participants, and summary findings," *PLoS ONE*, vol. 14, 3 2019.

[3] J. P. Boyle, A. A. Honeycutt, K. V. Narayan, T. J. Hoerger, L. S. Geiss, H. Chen, and T. J. Thompson, "Projection of diabetes burden through 2050," *Diabetes Care*, vol. 24, pp. 1936–1940, 11 2001.

[4] M. Zahabi, M. M. White, W. Zhang, A. T. Winslow, F. Zhang, H. Huang, and D. B. Kaber, "Application of cognitive task performance modeling for assessing usability of transradial prostheses," *IEEE Transactions on Human-Machine Systems*, vol. 49, pp. 381–387, 8 2019.

[5] A. W. Shehata, H. E. Williams, J. S. Hebert, and P. M. Pillarski, "Machine learning for the control of prosthetic arms: Using electromyographic signals for improved performance," *IEEE Signal Processing Magazine*, vol. 38, pp. 46–53, 6 2021.

[6] R. B. Woodward, J. A. Spanias, and L. J. Hargrove, "User intent prediction with a scaled conjugate gradient trained artificial neural network for lower limb amputees using a powered prosthesis," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2016-October, pp. 6405–6408, 10 2016.

[7] A. Ameri, M. A. Akhaee, E. Scheme, and K. Englehart, "Regression convolutional neural network for improved simultaneous emg control," *Journal of Neural Engineering*, vol. 16, p. 036015, 4 2019.

[8] A. Sie, C. Fisher, M. Karrenbach, E. Case, C. Caraballo, B. C. Muir, and E. Rombokas, "Timing of Haptic Cues for Stride Adjustments in Mobility Task," *Prep.*, 2020.

[9] A. Sie, M. Karrenbach, C. Fisher, S. Fisher, N. Wiek, C. Caraballo, E. Case, D. Boe, B. C. Muir, and E. Rombokas, "Descending 13 Real World Steps: A Dataset and Analysis of Stair Descent," *to be Submitt. to Gait Posture*, 2020.

[10] M. Karrenbach, D. Boe, A. Sie, R. Bennett, and E. Rombokas, "Improving automatic control of upper-limb prostheses using gaze-centered tracking," *Prep.*, 2020.

[11] Ottobock, "Upper Limb Prosthetics."

[12] UtahArm, "Taska Hand."

[13] E. Biddiss, D. Beaton, and T. Chau, "Consumer design priorities for upper limb prosthetics," *http://dx.doi.org/10.1080/17483100701714733*, vol. 2, pp. 346–357, 2009.

[14] S. M. Engdahl, B. P. Christie, B. Kelly, A. Davis, C. A. Chestek, and D. H. Gates, "Surveying the interest of individuals with upper limb loss in novel prosthetic control techniques," *Journal of NeuroEngineering and Rehabilitation*, vol. 12, 6 2015.

[15] K. Østlie, R. J. Franklin, O. H. Skjeldal, A. Skrondal, and P. Magnus, "Musculoskeletal pain and overuse syndromes in adult acquired major upper-limb amputees," *Archives of Physical Medicine and Rehabilitation*, vol. 92, 2011.

[16] K. Postema, V. van der Donk, J. van Limbeek, R. A. Rijken, and M. J. Poelma, "Prosthesis rejection in children with a unilateral congenital arm defect:," *http://dx.doi.org/10.1177/026921559901300308*, vol. 13, pp. 243–249, 7 2016.

[17] S. Salminger, H. Stino, L. H. Pichler, C. Gstoettner, A. Sturma, J. A. Mayer, M. Szivak, and O. C. Aszmann, "Current rates of prosthetic usage in upper-limb amputees–have innovations had an impact on device acceptance?," *Disability and Rehabilitation*, 2020.

[18] A. T. Sugawara, V. D. Ramos, F. M. Alfieri, and L. R. Battistella, "Abandonment of assistive products: assessing abandonment levels and factors that impact on it," *Disability and Rehabilitation: Assistive Technology*, vol. 13, pp. 716–723, 10 2018.

[19] E. Biddiss and T. Chau, "Upper limb prosthesis use and abandonment: A survey of the last 25 years," *Prosthetics and Orthotics International*, vol. 31, pp. 236–257, 9 2007.

[20] I. C. Narang, B. P. Mathur, P. Singh, and V. S. Jape, "Functional capabilities of lower limb amputees:," *http://dx.doi.org/10.3109/03093648409145345*, vol. 8, pp. 43–51, 6 2016.

[21] A. Telonio, S. Blanchet, C. N. Maganaris, V. Baltzopoulos, and B. J. McFadyen, "The detailed measurement of foot clearance by young adults during stair descent," *Journal of Biomechanics*, vol. 46, pp. 1400–1402, 4 2013.

[22] M. J. Highsmith, J. T. Kahle, A. L. Lewandowski, S. H. Kim, and L. J. Mengelkoch, "A method for training step-over-step stair descent gait with stance yielding prosthetic knees: A technical note," *Journal of Prosthetics and Orthotics*, vol. 24, pp. 10–15, 1 2012.

[23] Y. He, S. Y. Sun, A. Roy, A. Caspi, and S. R. Montezuma, "Improved mobility performance with an artificial vision therapy system using a thermal sensor," *Journal of Neural Engineering*, vol. 17, p. 045011, 7 2020.

[24] S. E. Hassan, J. E. Lovie-Kitchin, and R. L. Woods, "Vision and mobility performance of subjects with age-related macular degeneration," *Optometry and Vision Science*, vol. 79, pp. 697–707, 11 2002.

[25] S. C. Chen, G. J. Suaning, J. W. Morley, and N. H. Lovell, "Simulating prosthetic vision: Ii. measuring functional capacity," *Vision Research*, vol. 49, pp. 2329–2343, 9 2009.

[26] J. S. Hayes, V. T. Yin, D. Piyathaisere, J. D. Weiland, M. S. Humayun, and G. Dagnelie, "Visually guided performance of simple tasks using simulated prosthetic vision," *Artificial Organs*, vol. 27, pp. 1016–1028, 11 2003.

[27] J. M. Karl, L. R. Schneider, and I. Q. Whishaw, "Nonvisual learning of intrinsic object properties in a reaching task dissociates grasp from reach," *Experimental Brain Research 2013 225:4*, vol. 225, pp. 465–477, 1 2013.

[28] M. A. Arbib, T. Iberall, and D. Lyons, "Coordinated control programs for movements of the hand," *Hand Function and the Neocortex*, pp. 111–129, 1985.

[29] M. Jeannerod, "The timing of natural prehension movements," *Journal of Motor Behavior*, vol. 16, pp. 235–254, 1984.

[30] R. Tomovic, G. Bekey, and W. Karplus, "A strategy for grasp synthesis with multifingered robot hands," Institute of Electrical and Electronics Engineers, 1987.

[31] S. B. Kang and K. Ikeuchi, "Grasp recognition using the contact web," *IEEE International Conference on Intelligent Robots and Systems*, vol. 1, pp. 194–201, 1992.

[32] H. Liu, T. Iberall, and G. A. Bekey, "Multi-dimensional quality of task requirements for dextrous robot hand control," pp. 452–457, 1989.

[33] M. Hayhoe, "Vision using routines: A functional account of vision," *Visual Cognition*, vol. 7, pp. 43–64, 2000.

[34] M. F. Land, "Eye movements and the control of actions in everyday life," *Progress in Retinal and Eye Research*, vol. 25, pp. 296–324, 5 2006.

[35] C. Böhme and D. Heinke, "Modeling visual affordances: The selective attention for action model (saam)," *Proceedings of the International Symposium on AI Inspired Biology - A Symposium at the AISB 2010 Convention*, pp. 37–42, 2010.

[36] J. J. Gibson, "The ecological approach to visual perception," *The Ecological Approach To Visual Perception*, 5 2013.

[37] D. Morrison, P. Corke, and J. Leitner, "Closing the loop for robotic grasping: A real-time, generative grasp synthesis approach," 4 2018.

[38] G. M. Bone, A. Lambert, and M. Edwards, "Automated modeling and robotic grasping of unknown three-dimensional objects," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 292–298, 2008.

[39] M. Kopicki, R. Detry, M. Adjigble, R. Stolkin, A. Leonardis, and J. L. Wyatt, "One-shot learning and generation of dexterous grasps for novel objects:," *https://doi.org/10.1177/0278364915594244*, vol. 35, pp. 959–976, 9 2015.

[40] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection:," *https://doi.org/10.1177/0278364917710318*, vol. 37, pp. 421–436, 6 2017.

[41] A. Saxena, J. Driemeyer, and A. Y. Ng, "Robotic grasping of novel objects using vision:," *http://dx.doi.org/10.1177/0278364907087172*, vol. 27, pp. 157–173, 2 2008.

[42] R. Pelossof, A. Miller, P. Allen, and T. Jebara, "An svm learning approach to robotic grasping," *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2004, pp. 3512–3518, 2004.

[43] K. Hsiao, L. P. Kaelbling, and T. Lozano-Pérez, "Grasping pomdps," *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 4685–4692, 2007.

[44] D. Morrison, P. Corke, and J. Leitner, "Learning robust, real-time, reactive robotic grasping:," *https://doi.org/10.1177/0278364919859066*, vol. 39, pp. 183–201, 6 2019.

[45] T. Feix, J. Romero, H. B. Schmiedmayer, A. M. Dollar, and D. Kragic, "The grasp taxonomy of human grasp types," *IEEE Transactions on Human-Machine Systems*, vol. 46, pp. 66–77, 2 2016.

[46] Y. Liu, L. Jiang, D. Yang, Y. Liu, J. Zhao, and H. Liu, "Analysis on the joint independence of hand and wrist," *IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM*, vol. 2016-September, pp. 31–37, 9 2016.

[47] R. Haschke, J. J. Steil, I. Steuwer, and H. Ritter, "Task-oriented quality measures for dextrous grasping," *Proceedings of IEEE International Symposium on Computational Intelligence in Robotics and Automation, CIRA*, pp. 689–694, 2005.

[48] N. E. Krausz, D. Lamotte, I. Batzianoulis, L. J. Hargrove, S. Micera, and A. Billard, "Intent prediction based on biomechanical coordination of emg and vision-filtered gaze for end-point control of an arm prosthesis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, pp. 1471–1480, 6 2020.

[49] M. Cognolato, A. Gijsberts, V. Gregori, G. Saetta, K. Giacomino, A.-G. M. Hager, A. Gigli, D. Faccio, C. Tiengo, F. Bassetto, B. Caputo, P. Brugger, M. Atzori, and H. Müller, "Gaze, visual, myoelectric, and inertial data of grasps for intelligent prosthetics," *Scientific Data 2020 7:1*, vol. 7, pp. 1–15, 2 2020.

[50] M. Cognolato, M. Graziani, F. Giordaniello, G. Saetta, F. Bassetto, P. Brugger, B. Caputo, H. Müller, and M. Atzori, "Semi-automatic training of an object recognition system in scene camera data using gaze tracking and accelerometers," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10528 LNCS, pp. 175–184, 2017.

[51] C. P. F. Pasquina, A. J. Carvalho, and T. P. Sheehan, "Ethics in rehabilitation: Access to prosthetics and quality care following amputation," *AMA Journal of Ethics*, vol. 17, pp. 535–546, 6 2015.

[52] YCB, "Benchmarks-Object and Model Set."

[53] N. Sharma, V. Jain, and A. Mishra, "An analysis of convolutional neural networks for image classification," *Procedia Computer Science*, vol. 132, pp. 377–384, 2018.

[54] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," *Proceedings of 2017 International Conference on Engineering and Technology, ICET 2017*, vol. 2018-January, pp. 1–6, 3 2018.

[55] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," 11 2015.

[56] R. Yerrabelli, L. Osborn, J. H. Medicine, C. Hunt, C. Clancy, R. Kaliki, and N. Thakor, "Pham: Prosthetic hand assessment measure stem cell biomechanics modeling view project prosthetic hand assessment measure view project pham: Proshetic hand assessment measure," 2017.

[57] E. Rombokas, P. Brook, J. R. Smith, and Y. Matsuoka, "Biologically inspired grasp planning using only orthogonal approach angles," *Proceedings of the IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics*, pp. 1656–1661, 2012.

[58] A. Herzog, P. Pastor, M. Kalakrishnan, L. Righetti, J. Bohg, T. Asfour, and S. Schaal, "Learning of grasp selection based on shape-templates," *Autonomous Robots*, vol. 36, pp. 51–65, 1 2014.

[59] S. L. Carey, M. J. Highsmith, M. E. Maitland, and R. V. Dubey, "Compensatory movements of transradial prosthesis users during common tasks," *Clinical Biomechanics*, vol. 23, pp. 1128–1135, 11 2008.

[60] S. Schaefer and C. Hengge, "Testing the concurrent validity of a naturalistic upper extremity reaching task," *Experimental brain research*, vol. 234, p. 229, 1 2016.

[61] S. Liu, D. C. Mocanu, A. R. R. Matavalam, Y. Pei, and M. Pechenizkiy, "Sparse evolutionary deep learning with over one million artificial neurons on commodity hardware," *Neural Computing and Applications 2020 33:7*, vol. 33, pp. 2589–2604, 7 2020.

[62] B. Graham, "Spatially-sparse convolutional neural networks," 9 2014.