SAPIENZA
UNIVERSITÀ DI ROMA

An Analysis of the Visuomotor Behavior of Upper Limb Amputees to Improve Prosthetic Control

Department of Computer, Control, and Management Engineering

Dottorato di Ricerca in Engineering in Computer Science – XXXII Ciclo

Candidate

Valentina Gregori

ID number 1387986

Thesis Advisor

Prof. Barbara Caputo

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Ingegneria Informatica

October 2019

An Analysis of the Visuomotor Behavior of Upper Limb Amputees to Improve Prosthetic Control

Ph.D. thesis. Sapienza – University of Rome

This thesis has been typeset by LaTeX and the Sapthesis class.

Author's email: gregori@diag.uniroma1.it

Abstract

Upper limb amputation is a traumatic event with a dramatic impact on the everyday life of a person. The available solutions to restore the functionality of the missing hand via myoelectric prostheses have become ever more advanced in terms of hardware, but they are still inadequate in providing natural and robust control. One of the main difficulties is the variability and degradation of the electromyographic signals, which are also affected by amputation-related factors. To overcome this problem, it has been posited to combine surface electromyography with other sources of information that are less affected by the amputation.

Some recent studies have proposed to improve the control by integrating gaze, as visual attention is often predictive for future actions in humans. For instance, in manipulation tasks the eyes tend to fixate an object of interest even before the reach-to-grasp is initiated. However, the initial investigations reported in literature that combine vision with surface electromyography do so in an unnatural manner, meaning that the users need to alter their behavior to accommodate the system. The successful exploitation of gaze envisioned in this work is the opposite, namely that the prosthetic system would interpret the subject's natural behavior. This requires a detailed understanding of the visuomotor coordination of amputated people to determine when and for how long gaze may provide helpful information for an upcoming grasp. Moreover, while some studies have investigated the disruption of gaze behavior when using a prosthesis, no study has considered whether there is any disruption in visuomotor coordination due to the amputation itself.

In this work, we verify and quantify the gaze and motor behavior of 14 transradial amputees who were asked to grasp and manipulate common household objects with their missing limb. For comparison, we also include data from 30 able-bodied subjects who executed the same protocol with their right arm. The dataset contains gaze, first person video, angular velocities of the head, and electromyography and accelerometry of the forearm. To analyze the large amount of video, we developed a procedure based on recent deep learning methods to automatically detect and segment all objects of interest. This allowed us to accurately determine the pixel distances between the gaze point, the target object, and the limb in each individual frame. Our analysis shows a clear coordination between the eyes and the limb in the reach-to-grasp phase, confirming that both intact and amputated subjects precede the grasp with their eyes by more than 500 ms. Furthermore, we note that the gaze behavior of amputees was remarkably similar to that of the able-bodied control group, despite their inability to physically manipulate the objects.

Based on this knowledge, we show in a proof of concept that the combination of gaze and surface electromyography improves grasp recognition, both for intact and amputated subjects, compared to when only the latter modality is used. To make the integration natural for the user, we devised a method that allows a simultaneous combination of these modalities and weights the visual features based on their relevance. This evaluation is addressed as a proof of concept since the experiments were executed in a standard laboratory environment. We conclude the work therefore with a study to highlight the difficulty that machine learning based approaches need to overcome to become practically relevant also in daily living conditions.

Acronyms

ADL activity of daily living

AP Average Precision

CNN Convolutional Neural Network

COCO Common Objects in COntext

EMG electromyography

EOG electrooculography

FCN Fully Convolutional Network

FMG forcemyography

FPN Feature Pyramid Network

IDT Identification Dispersion Threshold

IHMM Identification Hidden Markov Model

IMST Identification Minimum Spanning Tree

IMU inertial measurement unit

IVT Identification Velocity Threshold

JSON JavaScript Object Notation

KNN K-Nearest Neighbors

KRLS Kernel Regularized Least Squares

LDA Linear Discriminant Analysis

MAV Mean Absolute Value

mDWT marginal Discrete Wavelet Transform

MeganePro Myo-Electricity, Gaze And Artificial-intelligence for Neurocognitive Examination & Prosthetics

MMG mechanomyography

MUAP motor unit action potential

PR Pattern Recognition

RBF Radial Basis Function

RLS Regularized Least Squares

RMS Root Mean Square

RNN Recurrent Neural Network

sEMG surface electromyography

SHAP Southampton Hand Assessment Protocol

STFT short-time Fourier transform

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Hands are an indispensable tool that allow humans to interact with the environment. The loss of a limb is therefore a traumatic event with a severe impact on the private, working, and social life of a patient. After the amputation the level of autonomy diminishes dramatically, for instance people struggle to perform even simple self-care activities (Niedernhuber et al. 2018). This factor has an important impact also in the working life and as a result many amputees remain unemployed or have to change their occupation (Burger 2009). For these reasons a big sense of frustration is often experienced after amputation. This may also be increased by the low acceptance of many prosthetic devices that the users do not feel as part of their own body (Niedernhuber et al. 2018).

Non invasive upper-limb prostheses that aim to substitute the missing arm can be divided in passive and active variants. Cosmetic hands, belonging to the first group, can either be static or adjustable (Maat et al. 2018). In the first case, the purpose of the prosthesis is purely aesthetic since no movement is possible, whereas in the second case it can manually be adjusted in a few configurations. Active prostheses, as the name suggests, can be actuated voluntarily by the user rather than by external factors (Castellini and Smagt 2009). This group includes body-powered and myoelectric devices (Carey et al. 2015). The first type is mechanically driven by shoulder movements that allow to open and close a hook. Myoelectric prostheses, on the other hand, are actuated by the muscles in the forearm. In this case, electrodes placed on the upper limb record changes in electric current from the skin after a neurological activation of the muscles. This recording of myoelectric signals from the skin in a non-invasive manner is known as surface electromyography (sEMG). These muscular signals are then converted into control commands for the prosthesis, which may range from simple opening and closing to more advanced commands. The three described types of prostheses are reported in Figure 1.1.

Though the myoelectric devices may seem the most sophisticated, a review by Carey et al. (2015) highlights that body-powered prostheses are perceived as more durable and practical, and require shorter training time. Moreover the cost of myoelectric prostheses is substantially higher than other options (Resnik et al. 2012; Huang et al. 2001). Furthermore, they typically use “direct control”, in which sEMG is collected from a pair of antagonist residual muscles to just open and close

Figure 1.1. Examples of cosmetic, body-powered, and myoelectric prostheses.

the hand (Ison and Artemiadis 2014). Only recent developments in commercial multi-articulated hands have allowed multiple grasp configurations from which a user can manually or semi-manually select one (Belter et al. 2013). The main problem of these devices is posed by the control, which is difficult and time consuming (Atkins et al. 1996; Castellini et al. 2014). Due to this and other problems, the mean rejection rate of these devices is high (Biddiss and Chau 2007b; Biddiss et al. 2007). A telling example of all these issues is given by the experience at the last Cybathlon 2016, which is a competition where people with physical disabilities use state of the art assistive devices to complete everyday life activities. The upper limb competition was won by a user wearing a body-powered prosthesis, which was far less advanced and much more economical than the myoelectric options worn by his competitors.

In literature several studies have attempted to interpret the intention of amputees by analyzing sEMG collected from the residual limb by means of Pattern Recognition (PR) (Herberts et al. 1973; Graupe et al. 1977; Wirta et al. 1978; Zecca et al. 2002; Roche et al. 2014; Hakonen et al. 2015, among others). However, the majority of these findings did not lead to an effective improvement in the clinical setting or real life (Farina et al. 2014; Roche et al. 2014; Resnik et al. 2018; Simon et al. 2019). The primary obstacle of these studies is related to the difficulty of decoding sEMG reliably in a realistic setting. Myoelectric signals are user-dependent and this holds in particular for amputees, as the amputation and subsequent muscular (dis)use have a considerable impact on the nature of these signals (Farina et al. 2002). Moreover, their quality changes in time due to fatigue, displacement of the electrodes, and sweat (Young et al. 2011; Castellini et al. 2014; Stango et al. 2014). To improve the control of myoelectric prostheses a number of strategies have been proposed (see Castellini et al. 2014; Farina et al. 2014; Madusanka et al. 2015; Herrera-Luna et al. 2019, and references therein). The idea behind many of these methods is to improve movement recognition by integrating or replacing sEMG with other modalities that are less influenced or deteriorated by amputation.

1.1 Visual Integration in Prostheses

The strong relationship between visual perception and manipulation makes the subject's gaze an appealing choice as supportive modality for myoelectric prostheses (Markovic et al. 2014; Hao et al. 2013). Vision plays an important role during

activities of daily living (ADLs), not only to guide the activity itself but also in the initial planning phase. Gaze is thus said to be anticipatory and can be used to understand an individual's intentions even before they manifest themselves in the motor domain (Land et al. 1999; Johansson et al. 2001, among others). Moreover, a careful observation of the eyes' activity not only informs on the intention to grasp, but can also help in identifying which object to grasp or, at the very least, its size and shape. Not surprisingly, several studies have attempted to use this information to help disabled people. For instance, in a robot assistant scenario the understanding of gaze fixations may be used to interpret the user's intentions (Admoni and Srinivasa 2016; Saran et al. 2018) or to aid tetraplegic patients who operate an exoskeleton in grasping activities (McMullen et al. 2013; Corbett et al. 2014).

In the prosthetic context, Castellini and Sandini (2006) proposed to use gaze to determine which object a user intends to grasp. The integration of gaze and vision as contextual information could be helpful especially during the initial transient phase of a movement, when the hand is preshaping its configuration to match the desired grasp. Due to its ambiguous nature, this reaching phase is the most challenging part of a movement to obtain a correct grasp classification from sEMG (Hargrove et al. 2007a). Since gaze is said to precede a manipulation, it seems a promising candidate to support the grasp recognition in a natural way. Leveraging over natural behavior of the user would make the recognition process faster and more intuitive, contributing to lower the cognitive burden. Since fatigue is one of the causes of variability in sEMG data, reducing it may also help to stabilize control.

The idea of using gaze besides sEMG to improve grasp recognition in myoelectric prosthesis has barely been explored in literature (Hao et al. 2013; Markovic et al. 2014; Markovic et al. 2015). In these studies, the subject is often equipped with customized glasses with a scene camera or eye tracking devices. Once the object has been successfully detected the user can proceed with the grasp, which is often selected based purely on the visual recognition of the object itself without considering sEMG. In these approaches, the visual behavior is imposed by the experimental requirements (i.e., to fixate an object until its recognition); a clear understanding of the natural visual planning was therefore unnecessary. A few unrelated studies have started to investigate the visuomotor coordination of prosthetic users while performing grasping or manipulation activities with their prosthesis (Bouwsema et al. 2012; Sobuh et al. 2014; Hebert et al. 2019; Aronson and Admoni 2018). The idea behind these studies is that the visual behavior can be used to evaluate the confidence that the users have with their prosthetic hands. These investigations highlighted indeed that prosthetic users spend more time looking the grasp related areas rather than planning forthcoming actions, contrary to what would be expected from intact users. A preliminary research, involving two amputated subjects with different skills in using a prosthesis, has hypothesized that the gaze behavior may "normalize" with an increasing confidence in the control response of the device (Chadwell et al. 2016). This research topic is however in its initial stages and a clear understanding of the visuomotor coordination of amputated people requires more investigation.

1.2 Research Questions and Contributions

In this thesis we explore the idea to fuse sEMG and gaze in a natural manner, rather than asking the users to accommodate the system by changing their visual behavior. Successfully exploiting the user's gaze behavior requires a precise understanding of natural eye-hand coordination. The primary contribution of this thesis is therefore a quantitative and qualitative investigation on the visuomotor coordination of amputees during reaching and grasping. A specific aim of this investigation is to determine the window of opportunity in which gaze can provide useful information for intent recognition before the realization of the grasp. Contrary to prior work, in this thesis we study the so-called "movements without movement" (Raffin et al. 2012b) to understand whether the eye-hand coordination has changed as a result of the amputation, rather than due to difficulties controlling a prosthesis. In other words, the participants were required to perform the movements as naturally as possible with their missing limb. This "ideal" setting does not imply that the results are not relevant for the prosthetic setting; the disruption of gaze strategies is actually characterized by a markedly longer reaching phase, while still maintaining the majority of the fixations on the target object (Sobuh et al. 2014; Hebert et al. 2019). The window of opportunity for gaze integration in the prosthetic setting is therefore expected to be considerably longer than the one we identify here.

More specifically, the main contributions of this work are the following.

Data Acquisition: We established an acquisition protocol to collect sEMG and gaze data from 30 intact and 15 amputated subjects engaged in multiple grasp and manipulation activities. Although all the tasks were performed in a laboratory environment, these were designed to include several levels of variability from the point of view of the manipulated objects, the limb position, and the complexity level of the grasp. The intact subjects executed the experiment with their right hand, the amputees were instead instructed to attempt to execute the action as naturally as possible "as if their missing limb were still there".

Estimate Feasibility Window: We precisely studied the gaze, head, and hand coordination in the reach-to-grasp phase both for intact and amputated subjects to estimate the time elapsed from the first fixation on the target object to the grasp. This evaluation is meant to quantify the temporal window in which the gaze information are useful to guide the grasp. To perform this analysis, we developed an automatic framework that employs state-of-the-art deep learning techniques to automatically detect and segment all objects of interest from the videos collected by the eye tracking glasses worn during the acquisitions.

Visuomotor Comparison: We investigated the visuomotor coordination of amputees quantitatively and qualitatively during several functional tasks and compared the obtained results with those of intact subject involved in the same activities. The aim of this study was to understand whether the amputation has influenced the manipulation and visual strategies in the absence of external factors, like the presence of a prosthesis, that could bias the outcome.

Proof of Concept: We devised a method to integrate sEMG and gaze data at the kernel level of a classifier to discriminate eleven hand configurations. The contribution of the gaze modality is weighted based on the distance of the eyes from the objects: the closer the gaze is to a known object, the more importance the information has in the grasp evaluation. The proposed method was used in a proof of concept to evaluate whether the inclusion of the visual information improves the recognition of the grasps performed by amputated subjects.

Analysis of Grasp Recognition during ADLs: We performed several analyses at the hand of a dedicated data acquisition composed of a typical laboratory training session in the first phase and a set of activities of daily living in a home setting afterwards. The objective was to provide a best-case analysis on whether grasps can be recognized if they are part of a composite, goal-oriented manipulation action such as an ADL.

1.3 Outline

The thesis starts in Chapter 2 with a general introduction on the relevant topics on which the work is based. The chapter opens with a description of surface electromyography and its use in myoelectric prostheses, subsequently it presents the machine learning methods used for the analysis of such data. The second part of the same chapter is dedicated to gaze and eye tracking techniques. Chapter 3 provides an overview of state of the art methods to improve prosthetic control via auxiliary modalities. A specific emphasis is given to the recent studies that proposed to integrate visual information with sEMG. We also report relevant findings on the visuomotor coordination of intact and amputated subjects. The dataset used in the majority of the studies presented in this work is described in Chapter 4, which includes also the preliminary tests on the acquisition protocol and the final technical validation on the acquired data. In Chapter 5 we introduce the automatic framework that was developed to analyze the huge amount of visual data collected in the dataset; specific attention is given to the deep learning methods on which this approach was built. The results obtained using this method on the visuomotor coordination of intact and amputated subjects are presented and examined in Chapter 6, which includes a comparison between the two groups of subjects and a discussion on the potential application in the prosthetic context. Based on these results, Chapter 7 contains a proof of concept approach on integrating sEMG and gaze data. It describes the approach and analyzes in detail the improvements gained from this integration. In Chapter 8, we extend some of the previous analyses evaluating specifically the realistic activities of daily living. This chapter begins with a description of the collected dataset and follows with the performed analysis and a thoughtful discussion on the differences between experiments conducted in controlled and unconstrained environments. Finally, the conclusions and directions for future work are presented in Chapter 9.

1.4 Declaration

This thesis presents the original work of its author. It is important however to underline that some parts of the work come from shared effort. In particular, the acquisition of the dataset that is presented in Chapter 4 was performed as part of the MeganePro project and all its members contributed to different aspects of this topic. The other chapters are instead primarily the work of the author of this thesis, executed under the supervision and support of the co-authors of the corresponding publications.

The following list gives an overview of the author's publications in chronological order.

1. Valentina Gregori, Arjan Gijsberts, and Barbara Caputo. "Adaptive learning to speed-up control of prosthetic hands: A few things everybody should know". In: *IEEE International Conference on Rehabilitation Robotics (ICORR)*. 2017, pp. 1130–1135[1]

2. Francesca Giordaniello, Matteo Cognolato, Mara Graziani, Arjan Gijsberts, Valentina Gregori, Gianluca Saetta, Anne-Gabrielle Mittaz Hager, Cesare Tiengo, Franco Bassetto, Peter Brugger, Barbara Caputo, Henning Müller, and Manfredo Atzori. "Megane Pro: Myo-Electricity, Visual and Gaze Tracking Data Acquisitions to Improve Hand Prosthetics". In: *IEEE International Conference on Rehabilitation Robotics (ICORR)*. 2017, pp. 1148–1153

3. Valentina Gregori, Barbara Caputo, and Arjan Gijsberts. "The Difficulty of Recognizing Grasps from sEMG during Activities of Daily Living". In: *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*. 2018, pp. 583–588

4. Andrea Gigli, Valentina Gregori, Matteo Cognolato, Manfredo Atzori, and Arjan Gijsberts. "Visual Cues to Improve Myoelectric Control of Upper Limb Prostheses". In: *IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*. 2018, pp. 783–788

5. Valentina Gregori, Matteo Cognolato, Gianluca Saetta, Manfredo Atzori, The MeganePro Consortium, and Arjan Gijsberts. "On the Visuomotor Behavior of Amputees and Able-Bodied People During Grasping". In: *Frontiers in Bioengineering and Biotechnology* 7 (2019), p. 316

6. Matteo Cognolato, Arjan Gijsberts, Valentina Gregori, Gianluca Saetta, Katia Giacomino, Anne-Gabrielle Mittaz Hager, Andrea Gigli, Diego Faccio, Cesare Tiengo, Franco Bassetto, Barbara Caputo, Peter Brugger, Manfredo Atzori, and Henning Müller. "Gaze, Visual, Myoelectric, and Inertial Data of Grasps for Intelligent Prosthetics". In: *Scientific Data* 7 (2020), p. 43

[1] This work is not presented in this thesis.

Chapter 2

Background

The work of this thesis is multi-disciplinary by nature, touching on topics related to machine learning, the biophysics of sEMG of the forearm's muscles, and the gaze behavior in response to external stimuli. This chapter is meant to give a general overview of the main principles and concepts behind the topics treated in the rest of the work. In Section 2.1 we introduce the techniques generally adopted to collect signals from the muscles during a movement and we explain how this information is used to control a myoelectric prosthesis. In the academic setting, many pattern recognition or machine learning approaches have been proposed to control advanced multi-channel and poly-articulated prostheses. For this reason, in Section 2.2 we introduce the general concepts of machine learning with an emphasis on the methods that are used in the following chapters to analyze sEMG data. Since in this thesis we investigate the benefit of integrating sEMG with gaze, in Section 2.3 we detail the main eye tracking techniques that are used to collect visual information. Moreover, we explain how the data collected with such methods are processed to obtain the information that describes the gaze behavior of a person.

2.1 Surface Electromyography for Myoelectric Prostheses

The motor system is a complex structure that is organized in central and peripheral units, which when activated in coordination allows humans to move (Rizzolatti and Luppino 2001; Augustine 2008). Motor programming takes place in specific parts of the cerebral cortex, the bioelectric signal is then transmitted to the spinal cord, and finally reaches the skeletal muscle of the limb leading to its contraction and therefore a movement. The electrical activity observed in the skeletal muscles is composed of a train of so-called motor unit action potentials (MUAPs), which are considered the basic elements of the signal. This muscular activity can be recorded via invasive or non invasive EMG (Merletti et al. 2004), as illustrated in Figure 2.1.

Intramuscular EMG is a method that allows to detect MUAPs in a small volume by means of intramuscular needles (Adrian and Bronk 1929; Farina and Negro 2012; LeFever and De Luca 1982). This approach is highly invasive but provides localized information concerning superficial or deep muscles, depending on where the needle

(a) intramuscular EMG **(b)** surface EMG

Figure 2.1. Schematic example of the techniques used to acquire EMG signals. In (a) EMG is recorded using an invasive method, with needles that record localized information within the muscles. In (b) the global muscle activity is recorded via surface electrodes placed on the skin.

is inserted. In many situations it may be difficult or unfeasible to perform invasive recordings; for instance, invasive measurements poses large difficulties when intended to be used daily as a control modality of a prosthesis. In such applications, the collection of EMG from the surface of the skin is therefore often preferred. This non-invasive superficial technique has however its limitations, as the acquired signal is not localized and it is nearly impossible to isolate the sEMG recorded from a single muscle. More generally, the so-called "cross-talk" effect is when some components of the signal from a muscle interfere with the signals of another muscle (De Luca and Merletti 1988; Winter et al. 1994; Dimitrova et al. 2002). Some studies have proposed an atlas of electrode placement to standardize their positioning based on the specificity of the information that may be collected in each area (Basmajian and Blumenstein 1980; Criswell 2010), but these guides have not become a widely accepted standard.

Most amputees still have some volitional control over their residual muscles and EMG can therefore be measured by surface electrodes on their stump. These signals have been used to control active, myoelectric prostheses in a non invasive manner (Geethanjali 2016). In the most common myoelectric prostheses, two channels of sEMG activity are collected from a pair of agonist-antagonist muscles (Zecca et al. 2002). This allows to open and close the hand via a solution known as direct control, as already introduced in Chapter 1. More articulated prostheses can extend this by allowing the user to switch among a set of possible grasps via a special activation pattern (Belter et al. 2013; Van Der Riet et al. 2013). For instance, the Michelangelo hand[1] has 7 available grips and the Bebionic hand allows 14 grasps[2] (Van Der Riet et al. 2013). A co-contraction or another predefined movement allows to iterate through the set of available grasps; once the desired grasp has been selected, the hand can be opened or closed with a direct control (Mastinu et al. 2018). To make the grasp selection faster, the i-Limb[3] hand allows to select a grasp also via a phone application or using special adhesive stickers that indicate the proximity of

[1] `https://www.ottobockus.com/prosthetics/upper-limb-prosthetics/solution-overview/michelangelo-prosthetic-hand/`

[2] `https://www.ottobockus.com/prosthetics/upper-limb-prosthetics/solution-overview/bebionic-hand/`

[3] `https://www.ossur.com/prosthetic-solutions/products/touch-solutions/i-limb-quantum`

grasp-labeled objects via the near-field communication protocol.

The previous manner to control prostheses is not intuitive and users often need a lot of practice to become confident with this mechanism (Peerdeman et al. 2011). In other words, the increased dexterity of recent advanced prosthesis comes at the cost of increased control difficulty. One of the limiting factors is the use of only two electrodes to collect the sEMG from the stump, which impedes control methods based on PR or machine learning. Such methods promise to allow both natural and more advanced control than the traditional direct control. In the academic environment, the first PR approaches date back to the 1960s - 1970s (Herberts et al. 1973; Graupe et al. 1977; Wirta et al. 1978). Machine learning applied to rich, multi-channel sEMG allows in essence the direct recognition of multiple grasps, thus bypassing the sequential grasp selection. However, despite decades of research the majority of these approaches have not led to an effective improvement in the quality of life of the patients. In fact, only one PR-based approach has recently been released commercially (Coapt, LLC 2015). As already explained in Chapter 1, the challenge consists in recognizing the amputee's grasp intent in a reliable manner from sEMG, which depends on factors as age, muscular use, and level of amputation (Farina et al. 2002; Criswell 2010; Young et al. 2011; Stango et al. 2014).

2.2 Machine Learning

Machine learning algorithms construct a mathematical model to solve a problem without being explicitly programmed to solve that problem (Mitchell 1997; Shalev-Shwartz and Ben-David 2014; Mohri et al. 2018). Such algorithms are said to be "data-driven", since they learn from examples and, based on this knowledge, construct a model to make decisions about future unknown data. The input to the algorithm is a set of *training* data, which is composed of observations concerning the problem to solve, and the responses that correspond to these observations, usually referred to as their *labels*. In a classification problem, which is the type of problem treated in this work, the label for an example is one of a discrete set of classes. The goal is then to create a general model that is able to predict the correct output class for unseen samples. These unknown samples are called *test* data and serve to estimate the *performance* of the model, which is measured by comparing the predicted labels with the ground-truth of the test set. In the context of myoelectric prostheses, the inputs are the sEMG signals and the labels are the movements chosen from a set of possible hand configurations. Based on the relevant features of the signals, a classification model is trained to distinguish between these grasps. Then in the test phase, the model assigns to each unknown sEMG sample one of the possible classes.

More specifically, let us define with $\mathcal{X} \subseteq \mathbb{R}^d$ and $\mathcal{Y} \subseteq \mathbb{R}$ the input and the output spaces. We indicate an input datum as $\boldsymbol{x}_i \in \mathcal{X}$ and the associated label as $y_i \in \mathcal{Y}$. In particular, we focus on supervised classification problems where the output space is known and represented by a set of G possible classes $\mathcal{Y} = \{1, \ldots, G\}$. The training set is represented by a set of N input-output pairs $D = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$ drawn from an unknown probability distribution $P(\boldsymbol{x}, y)$. To guarantee that this training set is representative also for future samples, it is assumed that the training data are independent and identically distributed. Similarly, the test set is indicated

as $D' = \{\boldsymbol{x}_i, y_i\}_{i=1}^{M}$ and, in standard machine learning approaches, it is assumed that train and test data come from the same joint distribution $P(\boldsymbol{x}, y)$.

The goal of a machine learning algorithm is to find a function

$$h \in \mathcal{F} \quad s.t. \quad h : \mathcal{X} \mapsto \mathcal{Y} \tag{2.1}$$

that for a future input vector $\boldsymbol{x}$ determines the corresponding output y. A measure of quality for this hypothesis function h is given by the loss function $\mathcal{L} : \mathcal{Y} \times \mathcal{Y} \mapsto \mathbb{R}_+$ defined as $\mathcal{L}(y, h(\boldsymbol{x}))$. The optimal choice of h corresponds then to the function that minimizes the loss over the joint distribution $P(\boldsymbol{x}, y)$ of the input and output space. This quantity is referred to as the expected risk

$$\mathcal{R}_{exp}(h) = \int_{\mathcal{X}\times\mathcal{Y}} \mathcal{L}(y, h(\boldsymbol{x}))dP(\boldsymbol{x}, y) \ . \tag{2.2}$$

In practice, it is impossible to calculate this quantity since the distribution $P(\boldsymbol{x}, y)$ is unknown. Therefore the minimization of the loss is calculated over the training samples to obtain the empirical risk

$$\mathcal{R}_{emp}(h) = \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}(y_i, h(\boldsymbol{x}_i)) \ . \tag{2.3}$$

The model obtained when optimizing the empirical risk is however not necessarily able to generalize to unseen samples. For instance, the model may simply memorize all the training samples, such that it would perform optimally on the training data but it would not be able to predict anything meaningful about new samples. This is the well-known overfitting problem. To avoid this issue, we optimize instead

$$\lambda\Omega(h) + \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}(y_i, h(\boldsymbol{x}_i)) \ . \tag{2.4}$$

In this equation, $\Omega(h)$ is a regularizer that avoids overfitting by penalizing the complexity of function h; λ is instead a parameter that balances the tradeoff between the two terms of the equation.

2.2.1 Linear Classifiers

The easiest group of problems approached in machine learning can be addressed with linear classifiers. Let us suppose for simplicity that the classification problem is binary, so that $\mathcal{Y} = \{-1, +1\}$. In this case, the samples belonging to the two classes can be separated by a linear function in the space

$$\mathcal{F} = \left\{\boldsymbol{x} \mapsto \text{sign}(\boldsymbol{w} \cdot \boldsymbol{x} + b) : \boldsymbol{w} \in \mathbb{R}^d, b \in \mathbb{R}\right\} \ . \tag{2.5}$$

The hypothesis function $h(\boldsymbol{x}) = \text{sign}(\boldsymbol{w} \cdot \boldsymbol{x} + b)$ assigns a positive label to the samples falling on one side of the hyperplane $\boldsymbol{w} \cdot \boldsymbol{x} + b = 0$ and a negative label to the others.

A multiclass classification problem can be tackled by reducing it to multiple binary classification tasks following the *one versus all* or *one versus one* approach. In the first case, we solve G independent binary classification problems where in

turn each of the classes is discriminated from the $G-1$ remaining classes. For each subproblem, a function $f_g(\boldsymbol{x}) = \boldsymbol{w} \cdot \boldsymbol{x} + b$ is learned and the final multiclass hypothesis function is rewritten as

$$h(\boldsymbol{x}) = \underset{g \in \mathcal{Y}}{\operatorname{argmax}} f_g(\boldsymbol{x}) \ . \tag{2.6}$$

In the latter case, a classifier is instead learned for each pair of classes $(g, g\prime)$ with the resulting hypothesis function $h_{gg\prime}$. The multiclass hypothesis function is then derived via majority vote

$$h(\boldsymbol{x}) = \underset{g\prime \in \mathcal{Y}}{\operatorname{argmax}} \Big| \big\{ g : h_{gg\prime}(\boldsymbol{x}) = 1 \big\} \Big| \ . \tag{2.7}$$

Regularized Least Squares One possible way to choose the linear function h is to select the hyperplane $\boldsymbol{w}$ that separates the classes of the training data with the maximum margin. Let us ignore the bias b in Equation 2.5 for convenience, as it can trivially be embedded in the weight vector $\boldsymbol{w}$ by adding a unitary last element to each input sample. With this simplification, the hypothesis function becomes $h(\boldsymbol{x}) = \operatorname{sign}(\langle \boldsymbol{w}, \boldsymbol{x} \rangle)$. It can mathematically be shown that the L_2 norm of the weight vector is proportional to the inverse of the margin between both classes. So choosing this norm as the regularization term and combining it with the common squared loss function, Equation 2.4 can be rewritten as

$$J(\boldsymbol{w}, \lambda) = \frac{\lambda}{2}\|\boldsymbol{w}\|^2 + \frac{1}{2}\sum_{i=1}^{N}(y_i - h(\boldsymbol{x}_i))^2 = \frac{\lambda}{2}\|\boldsymbol{w}\|^2 + \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|^2 \ . \tag{2.8}$$

In the second equation we use matrix notation, therefore $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]^T$ is an $N \times d$ matrix of all the training samples and, similarly, $\boldsymbol{y} = [y_1, \ldots, y_N]$ is the vector of the labels of the training data. The factor $\frac{1}{2}$ was added for mathematical convenience, but it is not relevant for the calculation. By minimizing the previous equation with respect to $\boldsymbol{w}$ we can determine the hyperplane that minimizes the prediction errors while maximizing the margin between both classes as

$$\frac{\partial J(\boldsymbol{w}, \lambda)}{\partial \boldsymbol{w}} = 0 \quad \Rightarrow \quad \boldsymbol{w}^\star = (\lambda \boldsymbol{I} + \boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} \ , \tag{2.9}$$

where $\boldsymbol{I}$ is an identity matrix of dimension $d \times d$.

Linear Discriminant Analysis In the previous method we calculated a linear function that minimizes the error between the predicted and true labels of a set of training data. Linear Discriminant Analysis (LDA) instead aims to find the best data projection that maximizes the class separation while contemporaneously minimizing the samples intra-class variance (Bishop 2006). Let us consider a binary problem with N_1 samples belonging to the first class and N_2 to the second. The mean and variance for each group are respectively $(\boldsymbol{m}_1, \boldsymbol{s}_1)$ and $(\boldsymbol{m}_2, \boldsymbol{s}_2)$. The objective function to maximize is then the ratio

$$J(\boldsymbol{w}) = \frac{\boldsymbol{w}^T \boldsymbol{S}_B \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{S}_W \boldsymbol{w}} \tag{2.10}$$

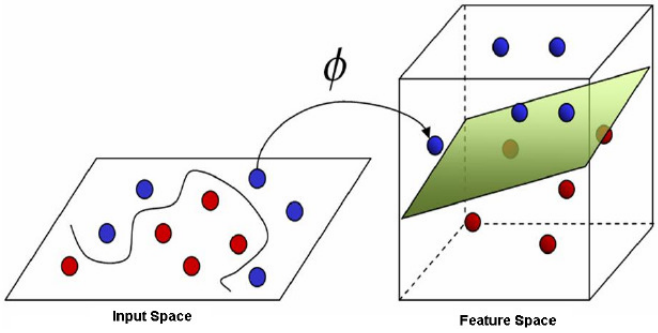


Figure 2.2. Overview of the kernel trick. In the input space the samples are non linearly separable, but by mapping them in an higher dimensional feature space using the feature mapping function Φ a linear classification problem is obtained.

of the *between-class* covariance matrix

$$\boldsymbol{S}_B = (\boldsymbol{m}_2 - \boldsymbol{m}_1)(\boldsymbol{m}_2 - \boldsymbol{m}_1)^T \tag{2.11}$$

divided by the *within-class* covariance matrix

$$\boldsymbol{S}_W = \sum_{i=1}^{N_1}(\boldsymbol{x}_i - \boldsymbol{m}_1)(\boldsymbol{x}_i - \boldsymbol{m}_1)^T + \sum_{i=1}^{N_2}(\boldsymbol{x}_i - \boldsymbol{m}_2)(\boldsymbol{x}_i - \boldsymbol{m}_2)^T \ . \tag{2.12}$$

The optimization of the objective function leads to the separation hyperplane

$$\boldsymbol{w} \propto \boldsymbol{S}_W^{-1}(\boldsymbol{m}_2 - \boldsymbol{m}_1) \ . \tag{2.13}$$

Due to its simplicity LDA has been popular in the context of prosthetic control (Englehart and Hudgins 2003).

2.2.2 Non Linear Classifiers

Linear classifiers are simple and computationally lightweight, but often not sufficiently representative for the majority of practical problems. An approach to define non linear decision boundaries while in essence still solving a linear method is given by the so-called kernel trick. Instead of working with the original data $\boldsymbol{x}$, we map them in a higher, possibly even infinite dimensional feature space using the mapping

$$\Phi : \mathcal{X} \mapsto \mathcal{H} \ , \tag{2.14}$$

where $\mathcal{H}$ is the Hilbert space. With this mapping we can apply the same linear algorithms, but in a more "powerful" high dimensional space, as shown in Figure 2.2.

Since $\Phi(\boldsymbol{x})$ is a mapping into an high dimensional space, its computation may not be efficient or even impossible in case of an infinite dimensionality. However, for many algorithms it is not necessary to *explicitly* map the samples into $\mathcal{H}$; instead, it is sufficient to be able to calculate the inner product

$$K(\boldsymbol{x}, \boldsymbol{x}') = \langle \Phi(\boldsymbol{x}), \Phi(\boldsymbol{x}') \rangle \quad \forall \ \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X} \tag{2.15}$$

between any two samples, where $K : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ is the so-called kernel function. The computational advantage is then given by the specific set of kernels that can calculate this inner product directly from the original samples in the input space $\mathcal{X}$, without explicitly mapping them into $\mathcal{H}$. Some popular examples of such kernels are:

- Polynomial kernel:

$$K(\boldsymbol{x}, \boldsymbol{x}') = (\boldsymbol{x} \cdot \boldsymbol{x}' + c)^d \quad \text{with} \quad c > 0 \ , \tag{2.16}$$

  where d is the degree of the polynomial. Note that the linear kernel is a special case of the polynomial kernel with $c = 0$ and $d = 1$.

- Gaussian or Radial Basis Function (RBF) kernel:

$$K(\boldsymbol{x}, \boldsymbol{x}') = \exp\Big(-\gamma \Big\| \boldsymbol{x}' - \boldsymbol{x} \Big\|^2 \Big) \quad \text{with} \quad \gamma > 0 \ . \tag{2.17}$$

- Exponential χ^2 kernel:

$$K(\boldsymbol{x}, \boldsymbol{x}') = \exp\Big(-\gamma \sum_{i=1}^{d} \frac{2(x_i - x'_i)^2}{(x_i + x'_i)} \Big) \quad \text{with} \quad \gamma > 0 \ . \tag{2.18}$$

Since the kernel has been defined as an inner product, it can also be interpreted as a measure of similarity between two samples $(\boldsymbol{x}, \boldsymbol{x'})$ in the feature space. However, a kernel function corresponds to an inner product in some feature space $\mathcal{H}$ only under certain conditions. These are specified by *Mercer's condition* (Mercer 1909), which requires the kernel to be symmetric positive semi-definite. If this condition is satisfied, then it can also be demonstrated that kernels can be combined to create other kernels. More specifically, (1) the sum of two kernels is a kernel, (2) the product of kernels is a kernel, and (3) the product of kernel with a constant factor is a kernel (Shawe-Taylor and Cristianini 2004).

Kernel Regularized Least Squares Using the kernel approach the Regularized Least Squares (RLS) method can be extended to the non linear case using a family of hypothesis functions $h(\boldsymbol{x}) = \text{sign}(\langle \boldsymbol{w}, \Phi(\boldsymbol{x}) \rangle)$. Equation 2.9 can be rewritten as

$$\boldsymbol{w}^\star = (\lambda \boldsymbol{I} + \boldsymbol{\Phi}\boldsymbol{\Phi}^T)^{-1}\boldsymbol{\Phi}^T\boldsymbol{y} = \boldsymbol{\Phi}^T\boldsymbol{\alpha} = \sum_{i=1}^{N} \alpha_i \Phi(\boldsymbol{x}_i) \ , \tag{2.19}$$

where $\boldsymbol{\Phi} = [\Phi(\boldsymbol{x}_1), \ldots, \Phi(\boldsymbol{x}_N)]^T$ and $\boldsymbol{\alpha} = (\boldsymbol{K} + \lambda \boldsymbol{I})^{-1}\boldsymbol{y}$ with $\boldsymbol{K} = \boldsymbol{\Phi}\boldsymbol{\Phi}^T$. The optimal hypothesis function is then obtained as

$$h^\star(\boldsymbol{x}) = \text{sign}(\langle \boldsymbol{\Phi}^T\boldsymbol{\alpha}, \Phi(\boldsymbol{x}) \rangle) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i \langle \Phi(\boldsymbol{x}_i), \Phi(\boldsymbol{x}) \rangle \right) = \text{sign}(\langle \boldsymbol{\alpha}, \boldsymbol{k} \rangle) \ , \tag{2.20}$$

where $\boldsymbol{k} = [k(\boldsymbol{x}, \boldsymbol{x}_1), \ldots, k(\boldsymbol{x}, \boldsymbol{x}_N)]^T$ is the vector of the kernel similarity between a new sample $\boldsymbol{x}$ and the training set.

2.2.3 Classification Scheme and Features

In the previous section, we indicated the sEMG data with the vector $\boldsymbol{x}$, however this is not entirely accurate. Raw sEMG measures trains of MUAPs, so rather than looking at the electrical current at one point of time one should consider a window of historic data to get meaningful information on the level of muscle activation. This slicing of the entire sEMG signal in overlapping windows is shown in Figure 2.3. The final grasp decision is then made for each window individually. This processing scheme was proposed by Englehart and Hudgins (2003) and later followed by the majority of the studies in the field. Furthermore, rather than using the entire sequence of raw sEMG as input, a more compact representation of this sequence is used, which should encode the relevant information of the muscle activity. This process is known as feature extraction and it is performed for each window of data (Hudgins et al. 1993; Zecca et al. 2002; Englehart and Hudgins 2003).

A wide variety of features has been proposed in literature (Zardoshti-Kermani et al. 1995; Englehart et al. 1999; Micera et al. 2010). The *time domain* features, based on the evaluation of sEMG amplitude, are the easiest to extract since the signal does not require additional preprocessing steps. Moreover it has been shown that there is a quasi linear relationship between the muscle force and the Root Mean Square (RMS) amplitude of the signal (De Luca 1997a). With the aim to capture this relation, a simple data aggregation like RMS or Mean Absolute Value (MAV) is often used (Phinyomark et al. 2012). The *frequency domain* features instead rely on the spectral characteristics of the signal, which seem to be related to the velocity of muscle fibers (Farina et al. 2004). Other more advanced features, such as the short-time Fourier transform (STFT) and marginal Discrete Wavelet Transform (mDWT), capture information in both the time and the frequency domains, aiming to preserve as much information as possible from the original signal. An extensive list of the main features is provided by Zecca et al. (2002) and Micera et al. (2010).

The ideal window length of historic data is a tradeoff between the minimum number of samples that allows to obtain a correct classification of the movement and the lowest possible delay. It is clear that a large window length may capture more statistical properties of the performed grasp. On the other hand, to wait for the acquisition of such information increases the delay between data collection and movement execution. The idea of overlapping windows of historic data is meant to reduce as much as possible the time delay from the signal reception to the grasp prediction, rather than waiting the entire window length before processing another segment. In literature various window lengths have been analyzed, the ideal size was found to be between 150 ms to 250 ms (Smith et al. 2010), but also shorter windows provide a good classification accuracy (Englehart and Hudgins 2003).

2.3 Eye Tracking

While the first part of this chapter focused on the hand, this second part is dedicated to the eyes. An overview of the eye's anatomy is presented in Figure 2.4. The visible part of the eye is composed of the sclera (i.e., the white part), the iris (i.e., the colored part), and the pupil, which is located at the center of the iris and regulates the amount of light in the eyes by changing its diameter (Snell and Lemp

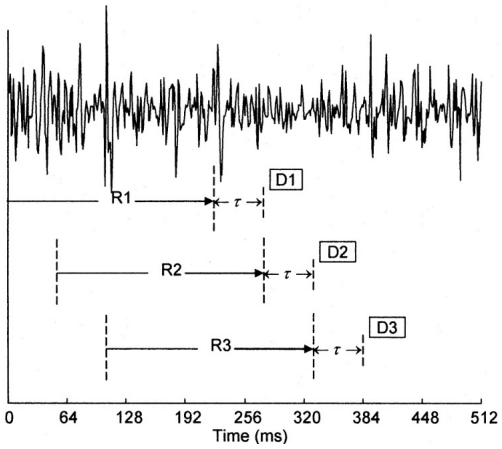


Figure 2.3. Overlapping windows of historic sEMG data. Figure credit: Englehart and Hudgins (2003).

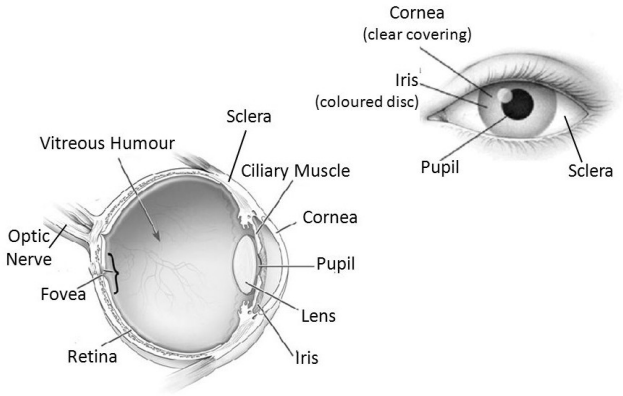


Figure 2.4. Anatomy of the eye.

2013; Duchowski 2017). The iris is protected by the cornea, a transparent external membrane. Internally, behind the iris, a biconvex lens converges the light rays on the retina, an inner photosensitive layer. A particular region of the retina, known as the fovea, is the part of the eye that has the sharpest vision and that is responsible for most of color perception.

Eye- and pupil-related information is generally collected for clinical and scientific purposes. Over the last decades, considerable efforts have been spent on the development of new eye tracking devices that allow to estimate where a person is looking. Among the traditional eye tracking techniques, Morimoto and Mimica (2005) distinguish between intrusive and non-intrusive methods. In the first case, the recording device is placed in direct contact with the user, such as for instance contact lenses, electrodes, or head-mounted devices. The method is non-intrusive when remote cameras are used to capture images of the eyes. Some camera based devices can be considered as semi-intrusive if they are head mounted. Intrusive eye tracking techniques are generally more accurate than the non-intrusive techniques. One of the most precise and invasive tracking is obtained with special contact

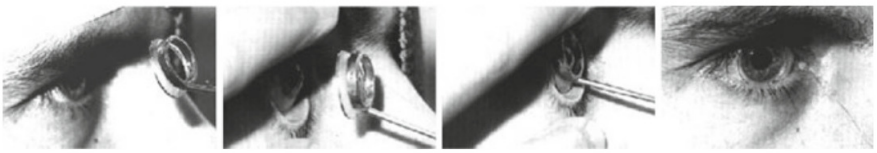


Figure 2.5. Procedure to place the contact lenses for eye tracking. Figure credit: Duchowski 2017.

Figure 2.6. Overview of the EOG setup. Figure credit: Duchowski 2017.

lenses that remain tightly attached over the cornea and the sclera so that the lenses move with the eyes (Yarbus 2013; Duchowski 2017). Such contact lenses, shown in Figure 2.5, estimate the movements of the eyes using embedded mirrors (Matin and Pearce 1964) or coils to measure their position via magnetic fields (Robinson 1963). Electrooculography (EOG), introduced by Mowrer et al. (1935), is another invasive method that measures the changes of the corneoretinal potential with skin electrodes placed around the eyes (Duchowski 2017). As shown in Figure 2.6, pairs of electrodes are placed above and below, and to the left and right side of the eyes. When the eyes move, also the cornea-retina dipole moves and the electrode pairs record a difference in potential, which indicates the eye's position.

In contrast to the previous methods, the so-called camera based eye tracking techniques do not require direct contact with the eyes or the skin (Morimoto and Mimica 2005). Videooculography methods make use of one or more cameras to determine the movement of the eyes by analyzing certain features captured in the recordings (Duchowski 2017). The general features used for tracking are in this case the pupil and the limbus, which is the boundary between the iris and the sclera. However, the pupil is often difficult to detect due to the low contrast with the iris, which is why methods have been explored over the years to accentuate this difference. One of the first methods consisted in illuminating the eyes with visible light to produce the so-called corneal reflection, which is a glint on the cornea of the eyes (see Thomas and Stasiak 1964; Young and Sheena 1975 and references therein). More recent techniques use instead infrared or near-infrared light to enhance the contrast between pupil and iris. The advantage with respect to visible light is that the user is not distracted by the beam. Also in this case the light generates a corneal reflection in the iris, but with this method also the pupil can be "segmented". There are two infrared eye tracking techniques that are commonly used: bright-pupil and

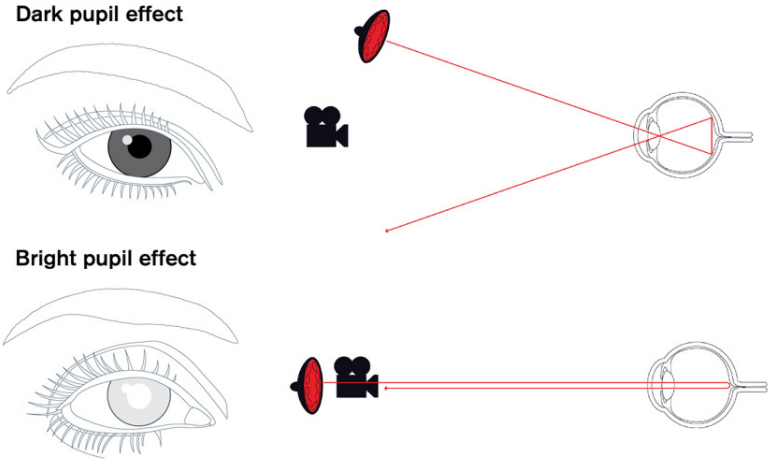


Figure 2.7. Infrared camera based eye tracking techniques with dark (top) and bright (bottom) pupils. In both cases the glint of the corneal reflection is observed in the top part of the eye. Figure credit: `www.tobiipro.com`.

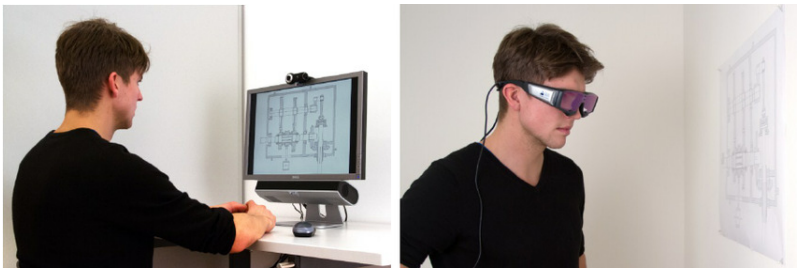


Figure 2.8. Remote (left) and head mounted (rigth) infrared camera based eye tracking systems. In the first case the device is placed under the display monitor, in the second case it is embedded in the glasses worn by the user. Figure credit: Lohmeyer et al. 2013.

dark-pupil (Morimoto et al. 2000; Morimoto and Mimica 2005). In the first case the illumination source is (almost) coaxial with respect to the optical axis of the camera and thanks to the reflection the pupil appears bright—this is the same phenomenon that makes eyes appear red in pictures. When the illumination source is offset from the optical path, the pupil becomes darker than the iris. A representation of both methods is shown in Figure 2.7. These camera based methods exist both in invasive and non-invasive variants, depending on whether the device is head mounted or remote as shown in Figure 2.8. However, while in the first case the user can freely move around, in the second case the space in which the experiment can be performed is strictly bounded around the monitor under which the eye tracker device is placed.

2.3.1 Gaze Coordinates

The abovementioned methods obtain the rotation of the eyes using different techniques, namely the estimation of the movement of lenses, the electric potential of

the skin, or optical characteristics of the pupils. Modern eye tracking devices are typically based on infrared corneal reflection, as this technique is more practical and less invasive. In this case, the center of the pupil is estimated via the pupil's reflection and combined with the corneal glint to obtain the gaze direction (Poole and Ball 2006). Subsequently, combining the gaze direction $\boldsymbol{d}$ and pupil center $\boldsymbol{p}$ of each eye it is possible to approximate the gaze point $\boldsymbol{g}$ in 3-dimensional world coordinates. These vectors are shown in Figure 2.9, where we decided to represent a head mounted system since it is the device that is used in the remainder of this work. To calculate the 3-dimensional coordinates, the left and right gaze directions are expressed in their parametric form as lines passing through the pupil centers:

$$\begin{aligned} \boldsymbol{v}_l &= \boldsymbol{p}_l + t_l \boldsymbol{d}_l \ , \\ \boldsymbol{v}_r &= \boldsymbol{p}_r + t_r \boldsymbol{d}_r \ . \end{aligned} \tag{2.21}$$

Contrary to the representation in Figure 2.9, these two lines generally do not strictly intersect when working in finite resolution and due to measurement errors. This situation is shown in Figure 2.10. Therefore instead of the intersection point, we calculate the points where the distance between two skew lines is shortest. This distance lies along a line that is perpendicular to both $\boldsymbol{v}_l$ and $\boldsymbol{v}_r$; the unit vector of this normal line is given by

$$\boldsymbol{n} = \frac{\boldsymbol{d}_l \times \boldsymbol{d}_r}{|\boldsymbol{d}_l \times \boldsymbol{d}_r|} \ . \tag{2.22}$$

Let us define the vector $\Delta\boldsymbol{p} = (\boldsymbol{p}_l - \boldsymbol{p}_r)$ that goes from the left to the right pupil center. The minimum distance between the lines is then the projection of this difference on the unit vector of Equation 2.22 perpendicular to the lines

$$|\Delta\boldsymbol{p} \cdot \boldsymbol{n}| \ . \tag{2.23}$$

The intersection of $\boldsymbol{v}_l$ with the plane formed by the translations of $\boldsymbol{v}_r$ along $\boldsymbol{n}$, which is identified by its normal vector $\boldsymbol{n}_r = \boldsymbol{d}_r \times \boldsymbol{n}$, gives the point on $\boldsymbol{v}_l$ nearest to $\boldsymbol{v}_r$. Performing the same calculation for $\boldsymbol{v}_r$, the nearest points lying respectively on the left and right lines are

$$\begin{aligned} \boldsymbol{g}_l &= \boldsymbol{p}_l + \frac{(\boldsymbol{p}_r - \boldsymbol{p}_l) \cdot \boldsymbol{n}_r}{\boldsymbol{d}_l \cdot \boldsymbol{n}_r} \boldsymbol{d}_l \\ \boldsymbol{g}_r &= \boldsymbol{p}_r + \frac{(\boldsymbol{p}_l - \boldsymbol{p}_r) \cdot \boldsymbol{n}_l}{\boldsymbol{d}_r \cdot \boldsymbol{n}_l} \boldsymbol{d}_r \ . \end{aligned} \tag{2.24}$$

Finally, the 3-dimensional gaze $\boldsymbol{g} = (g_x, g_y, g_z) = \frac{1}{2}\boldsymbol{g}_l + \frac{1}{2}\boldsymbol{g}_r$ is right in the middle between the left and right points calculated in Equation 2.24.

Many eye trackers also record a first person video using a scene camera. In this case it is relevant to project the 3-dimensional gaze point in the frame coordinate system to match the recorded video with the subject's gaze, as shown in Figure 2.11. If the horizontal and vertical field of view of the camera $\boldsymbol{\alpha} = (\alpha_x, \alpha_y)$ and the video resolution (w, h) are known, then the coordinates of the 2-dimensional gaze point, $\boldsymbol{g'}$, are

$$\begin{aligned} g'_x &= \frac{w g_x}{2 g_z \tan \dfrac{\alpha_x}{2}} \\ g'_y &= \frac{h g_y}{2 g_z \tan \dfrac{\alpha_y}{2}} \ . \end{aligned} \tag{2.25}$$

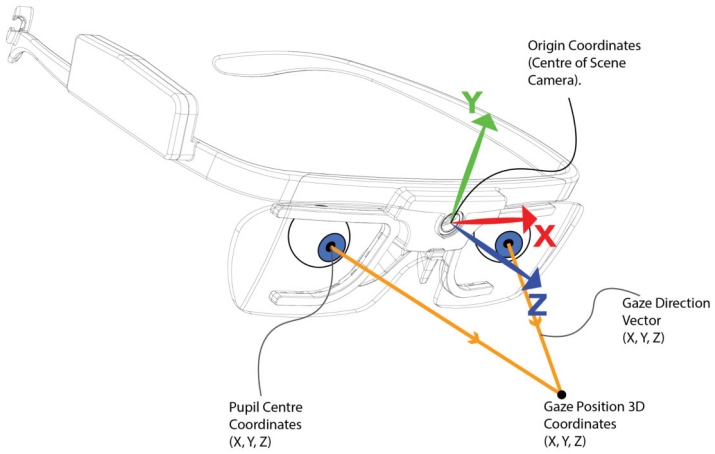


Figure 2.9. Overview of the gaze quantities calculated by an head mounted eye tracking system. Figure credit Tobii AB 2017.

Although this equation captures the essence of the mapping in terms of eccentricity and distance, in practice the calculation is not as straightforward. Additional factors play a role, such as camera distortion, possible rotation or translation of the camera with respect its expected position, and the calibration procedure of the device.

2.3.2 Fixations and Saccades

Eye tracking has gained popularity in many fields, as it allows to easily study human gaze behavior in a wide range of situations. The primary categorization of eye movements is in terms of fixations and saccades (Duchowski 2017; Holland and Komogortsev 2013). Fixations occur when the eyes are maintained in a specific position, namely the fovea remains centered on an object of interest. A natural fixation generally lasts about 500 ms (Johansson et al. 2001; Hessels et al. 2017). Even though the eyes are commonly assumed to be steady during a fixation, they may in fact perform small movements, such as tremor, microsaccades, or drift that serve to maintain visibility (Pritchard 1961; Martinez-Conde et al. 2004). Movements smaller than 5° are generally ignored in the fixation analysis and considered simply as random noise around the fixation area (Carpenter 1988). Saccades on the other hand are rapid, ballistic eye movements that reposition the fovea from one location to another. Small saccades with a size between 2.5° to 20° typically deal with near objects, whereas larger ones indicate a search of further objects (Land et al. 1999). Saccade events happen fast and generally have a duration ranging from just 10 ms to 100 ms (Baloh et al. 1975; Duchowski 2017).

Several mathematical models have been proposed to classify fixations and saccades. Komogortsev et al. (2010) presented two groups of eye movement classification algorithms: velocity and position based methods. The first group takes advantage of the fact that during fixations the eye velocity is lower than during saccades (Salvucci and Goldberg 2000). The Identification Velocity Threshold (IVT) method, belonging to this group, computes the point-to-point angular velocity for each couple of consecutive gaze coordinates and compares it with a threshold. If the velocity is

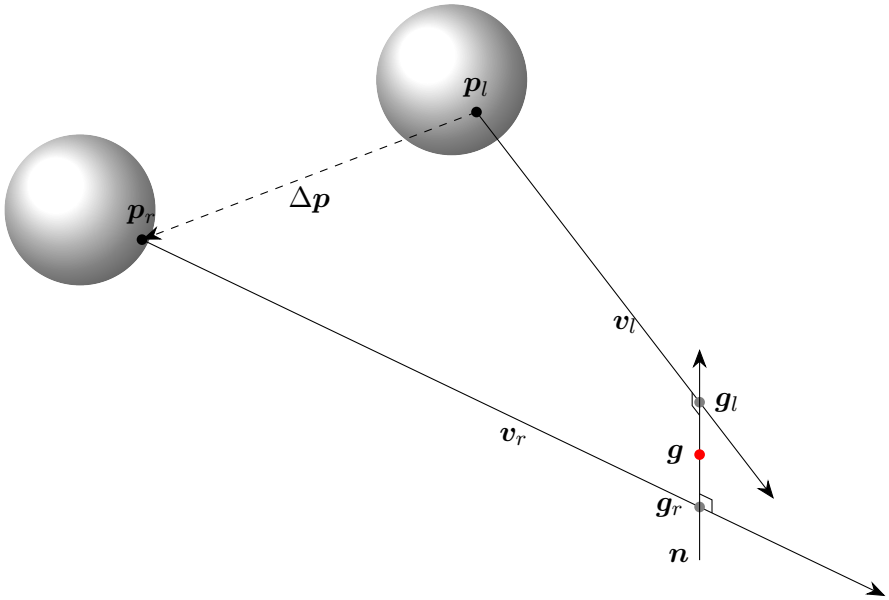


Figure 2.10. A schematic overview of non intersecting lines passing through the pupil centers and oriented in the left and right gaze directions. The points g_l and g_r represent the points on each of the two lines where their distance is minimal. They also lie on the vector that is normal to both v_l and v_r identified by the unit vector n.

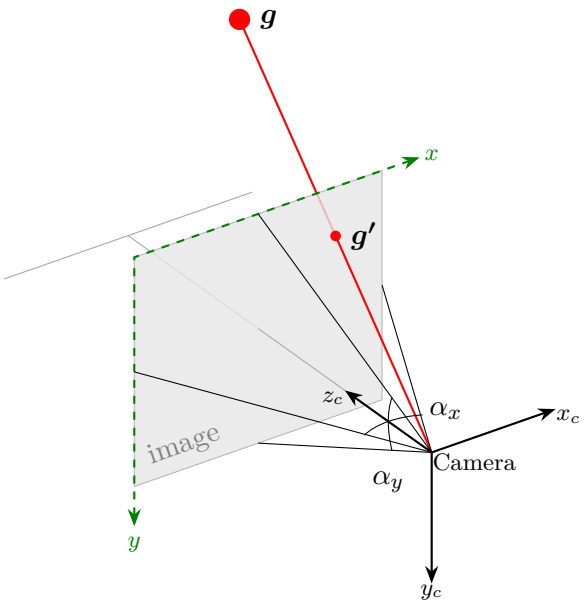


Figure 2.11. Projection of the 3-dimensional gaze point on the camera frame. The angles α_x and α_y represent the horizontal and vertical field of view of the camera.

lower than this threshold it is classified as a fixation, otherwise as a saccade. A more complicated approach is Identification Hidden Markov Model (IHMM) (Salvucci and R. Anderson 1998), which is a probabilistic method that models fixations and saccades as two states, each represented by a velocity distribution. The transition probability from one state to another is the likelihood that a new gaze sample is of the same type as the previous sample or that it changes instead from fixation to saccade or vice-versa. Position based algorithms differ in that they analyze the spatial dispersion of gaze points. These algorithms are based on the assumptions that (1) points belonging to the same fixation will be close to each other and (2) an increasing distance among successive data indicates a saccade (Salvucci and Goldberg 2000). For instance, Identification Dispersion Threshold (IDT) considers sliding windows over the gaze position signal (Nyström and Holmqvist 2010). If the spatial dispersion calculated in each window is lower than a predefined threshold, then the samples are classified as fixations, otherwise as the window moves the new incoming samples are recognized as saccades. The Identification Minimum Spanning Tree (IMST) algorithm (Goldberg and Schryver 1995) builds instead a tree to connect 2-dimensional gaze coordinates with segments. The idea behind this approach is that short line segments connects fixation points and, instead, longer segments represent a saccade.

2.3.3 Gaze Velocity

As mentioned, velocity based algorithms take into account the angular velocity of the gaze (Salvucci and Goldberg 2000). Given two consecutive 3-dimensional gaze vectors $\boldsymbol{g}_{i-1}$ and $\boldsymbol{g}_i$, the angular difference between them can easily be calculated by means of their dot product (Duchowski 2017)

$$\alpha_i = \arccos\left(\frac{\boldsymbol{g}_i \cdot \boldsymbol{g}_{i-1}}{\|\boldsymbol{g}_i\|\|\boldsymbol{g}_{i-1}\|}\right), \quad \forall i \in \{2, ..., N\} \ . \tag{2.26}$$

An approximation of the instantaneous gaze velocity at time t_i then follows as

$$v_i = \frac{\alpha_i}{t_i - t_{i-1}}, \quad \forall i \in \{2, ..., N\} \ . \tag{2.27}$$

Although the Tobii glasses used in the remainder of this thesis provide a unit gaze direction vector for both eyes (see Figure 2.9), we instead use the gaze point in world coordinates to estimate the common angle of the eyes. These world coordinates had fewer missing data and were slightly cleaner in practice due to onboard processing. They are however relative to the position of the scene camera rather than the eyes, as highlighted by the reference system reported in Figure 2.9. Since this camera is located on top of the frame of the glasses, this may lead to some inaccuracy at small gaze distances. The gaze point can therefore be mapped in a coordinate system that is centered between the left and right pupils

$$\widehat{\boldsymbol{g}}_i = \boldsymbol{g}_i - \bar{\boldsymbol{p}}_i, \quad \forall i \in \{1, ..., N\} \ , \tag{2.28}$$

where $\bar{\boldsymbol{p}}_i$ is the average of the left and right pupil locations relative to the scene camera.

Chapter 3

Related Work

In the previous chapters we highlighted that the limitations of myoelectric prostheses are primary related to the difficulty of interpreting and using sEMG as a control modality. The signals depend strongly on the anatomical characteristics of the user, the positioning and the contact of the electrodes on the skin, changes in arm positions, and fatigue (Farina et al. 2002). Moreover, the training procedure to learn to use the prosthesis with satisfactory performance is often perceived as painful and long (Peerdeman et al. 2011). Various studies have proposed to overcome the controllability problems by integrating or replacing sEMG with modalities that are more stable or less affected by amputation-related factors. In Section 3.1, we present an overview of studies that proposed to improve prosthetic control via multimodal integration techniques. Since this thesis investigates the use of the gaze as support modality, we first summarize the main findings in the field of visuomotor coordination in Section 3.2. This is followed by an overview of the studies that integrated gaze or visual information for prosthetic control. Finally, we report the main findings in literature regarding visuomotor coordination of upper limb amputees while manipulating objects with a prosthesis.

3.1 Multimodal Control of Prostheses

To improve the controllability of myoelectric prostheses, both muscular and "contextual" modalities have been considered in literature as alternative or additional information to sEMG to understand the intended grasp (Herrera-Luna et al. 2019; Madusanka et al. 2015). In the following we report how both approaches have been used in previous studies.

3.1.1 Muscular Information

Medical ultrasound imaging allows to observe the interior of the arm and therefore the contractions of the hand and wrist muscles in the forearm. An example of an ultrasound device is shown in Figure 3.1. This technique was initially applied to recognize the flexion of the fingers and the adduction of the thumb (Castellini and Passig 2011), movements of flexion/extension of the elbow and knee, rotation of

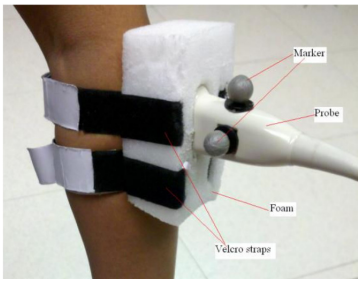


Figure 3.1. Ultrasound device used to observe muscle movements. Figure credit: Zhou and Zheng (2012).

wrist and ankle (Zhou and Zheng 2012), and wrist angles (Guo et al. 2013). In these cases, commercially available ultrasound equipment was used to record a stream of images that were then analyzed with common vision processing methods, which provided results comparable to similar studies using sEMG. Moreover, while the sEMG electrodes should be placed on a muscle belly to minimize the noise, the positioning of the ultrasound device on the arm does not seem to alter the output performance, as reported by McIntosh et al. (2017). While previous studies focused on position control, Sierra González and Castellini (2013) investigated the prediction of the finger forces and demonstrated the existence of a linear relationship between ultrasound image features acquired from the forearm and fingertip forces. Despite the obtained results are surprisingly good in term of movement recognition, in terms of wearability the equipment required to record ultrasound is considerable more cumbersome than sEMG electrodes (Castellini et al. 2014).

While performing a movement or a grasp with the upper limb, the muscles contract inside the body. These contractions propagate externally in terms of change of the forearm's volume and this can be recorded via forcemyography (FMG) with force-sensitive resistors. Among the first studies that introduced this idea in the prosthetic context were Curcie et al. (2001) and Phillips and Craelius (2005), who discriminated specific finger flexion and finger taps from FMG acquired from the residual limb of amputated users. This technique was also employed to predict the grip force during grasps (Wininger et al. 2008) and to classify eight hand and wrist movements (Radmand et al. 2016) obtaining a level of performance that was comparable with sEMG-based methods. Instead of using only FMG for movement recognition, recent studies have also started to combine this modality with standard sEMG. Nissler et al. (2017) presented a preliminary work in which tactile-myography and sEMG were independently acquired from six intact subjects and one amputee while performing wrist and hand movements. The comparison between the two modalities showed that FMG generally outperforms sEMG in terms of number of successful concluded tasks and completion time. This result was confirmed by a similar experiment in a follow-up (Jaquier et al. 2017). In this case, a combination of tactile-myography and sEMG was shown to improve movement recognition.

In addition to changing the arm's volume, a muscle contraction also elicits mechanical vibrations in the range of 5 Hz to 100 Hz that can be captured. This measure is called mechanomyography (MMG) and it is recorded either by means of microphones (Goldenberg et al. 1991; Courteville et al. 1998) or by accelerometers (Silva

et al. 2003b). Advantages of MMG in the prosthetic context are that, contrary to sEMG, it is not affected by factors like skin impedance, moisture level, and interference (Islam et al. 2013). This modality has been used to move a prosthetic wrist (Silva et al. 2003a) and to recognize four hand postures (Zeng et al. 2009). Recently, Wilson and Vaidyanathan (2017) proposed to fuse inertial measurement and MMG in a custom hardware design to control a commercial prosthetic hand. This setup allowed to distinguish several hand movements both in online and offline situations. The integration of MMG and sEMG was instead successfully executed by Xiloyannis et al. (2015) to predict finger movements.

3.1.2 Contextual Information

So far we have described modalities that complement or substitute sEMG by gathering different information from the muscles. There are other approaches that instead consider external or contextual factors, which have the advantage of not depending on the condition of the residual limb. For instance, accelerometry can be used to map the arm's orientation and movements in the inertial reference frame. The acquisition of this information is cost-effective, because accelerometers are cheap and some research-oriented electrodes already come with a three-axial accelerometer embedded. Several studies have shown that the combination of sEMG and accelerometry outperforms the use of the solely former modality for hand gestures recognition (Chen et al. 2007; Fougner et al. 2011; Gijsberts et al. 2014; Georgi et al. 2015). Bennett and Goldfarb (2017) proposed instead a sequential approach in which the inertial measurement is used to rotate the wrist of a myoelectric prosthesis, whereas sEMG is subsequently employed to control the hand. Similarly, Fougner et al. (2011) presented a "two-stage classifier" to first detect the limb position by means of the accelerometer and then to decide the hand movement by means of sEMG.

Also vision has been used to gather contextual information from the surroundings to facilitate a grasp. By simply embedding a camera in the prosthesis, as shown in Figure 3.2, images can be collected while the hand approaches the object for the grasp. Modern techniques of image processing can then be used to analyze this data and provide information on the object. For instance, several studies trained Convolutional Neural Networks (CNNs) or other deep neural networks to associate an RGB or RGB-D image to a predefined grasp, rather than the standard object label (DeGol et al. 2016; Taverne et al. 2019). In these cases the only information that determines the grasp is the correct recognition of the object. Lenz et al. (2015) proposed a grasp selection process based on two-step cascade deep learning algorithms. At the first stage, a network extracted the grasping points for the object, later another network used this information to estimate the best grasp. This approach was improved by Ghazaei et al. (2017), who in their method selected the grasp type according to an abstract representation of the objects, by associating the same grasp label to multiple objects with the same visual appearance . As shown in Figure 3.3, a wrist movement triggered a signal to take a picture of the object and after the grasp was proposed by the system the prosthesis was controlled proportionally via sEMG. Similarly Fajardo et al. (2018) proposed a multi step approach in which multiple grasps were associated to an object via visual recognition

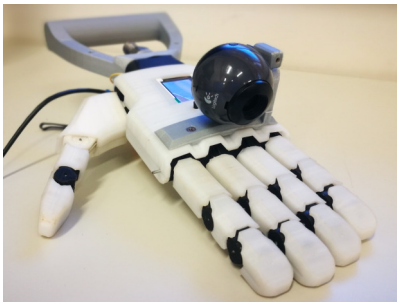


Figure 3.2. Example of a camera integrated in a hand prosthesis. Figure credit: Fajardo et al. (2018)

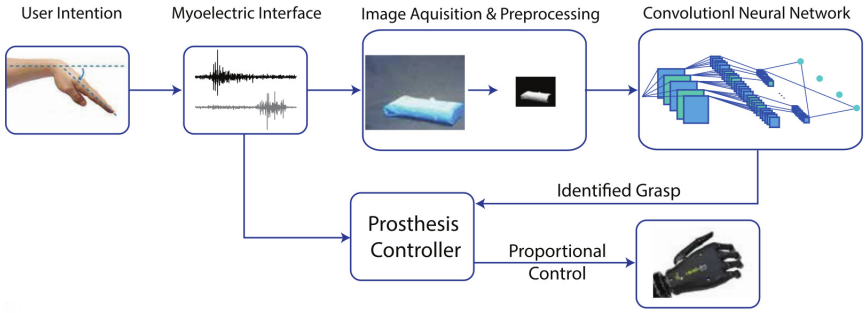


Figure 3.3. Overview of the experimental structure proposed by Ghazaei et al. (2017) to use sEMG and visual information from a camera embedded in the prosthesis. Figure credit: Ghazaei et al. (2017).

methods. Among the proposed grasps, the user could semi-manually select the right one via muscular contractions decoded by the Myo armband. In all these cases, the hand configuration is determined exclusively by the grasp-label(s) assigned to each object. However, vision can also be used to estimate higher level information, such as position, size, and shape of the objects (Klisić et al. 2009). For instance, Došen and Popović (2011) used vision-based techniques to estimate the object's size and orientation, and an ultrasound distance sensor to establish the distance from the target. The grasp is then selected by taking into account all these characteristics. Gardner et al. (2014) combined instead MMG, to recognize the intention to open and close the hand, with a camera for grasp selection (see Figure 3.4). Similarly, Došen et al. (2010) used sEMG to open, close, and orientate the hand during grasping and manipulation, while a system based on visual information selects the grasp type and the aperture size by estimating object properties. Along the same lines, Hays et al. (2019) integrated both visual and tactile data to improve the performance of a prosthetic hand based on sEMG recognition. When the tactile sensors detect slip, the prosthesis adjust the grip force. Visual feedback is instead used at the beginning to set the grasp type and wrist rotation, if the grasp is mistaken the user can make use of classical pattern recognition system based on sEMG. The main drawback of these studies is that if an object can be grasped with more than one hand configuration (DeGol et al. 2016), then just using visual information is not sufficient to resolve this ambiguity.

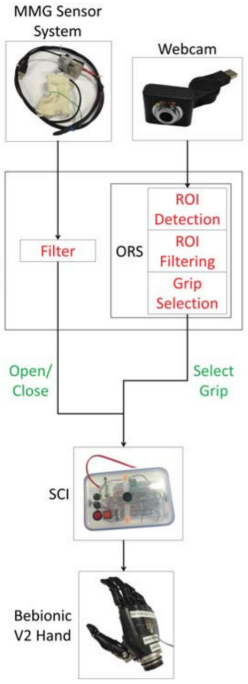


Figure 3.4. Overview of the experimental structure proposed by Gardner et al. (2014) to use MMG and visual information from a camera embedded in the prosthesis. Figure credit: Gardner et al. (2014).

3.2 Visuomotor Coordination

Eye tracking is the practice of measuring eye movements to understand the human visual behavior and where the focus of the people falls during an activity (Poole and Ball 2006). Early studies on gaze behavior typically involved constrained settings, for instance by fixating the chin to avoid head movements or by limiting the field of view to a monitor. Generally, the goal of these studies was to understand visual saliency, or in other words the features that capture visual attention while observing a static scene (see Tatler et al. 2011, and references therein). This approach however has several limitations, as highlighted in the reviews by Tatler (2014) and Tatler et al. (2011). For instance, the subject's behavior may be biased by time and space constraints: the visual stimulus is shown often for a short period of time in a field of view that is confined to a screen. Moreover, with such an experimental setup one misses out on the natural movements done while interacting with the environment as well as all the dynamism of the real world.

Unconstrained experiments became possible with the introduction of wearable eye-tracking devices that allowed the user to move freely in the environment (Land 2006). This advancement also made it possible to analyze the natural coordination of eye, head, and hand while executing more complex tasks. A typical activity studied in literature consists of a block copying task, in which the subjects are asked to pick blocks from a "resource" area and place them in a "workspace" area as to replicate a configuration demonstrated in a "model" area. This task was analyzed also when different instructions were given to the participants to capture multiple coordination

strategies. The first movement of the head generally follows the first saccade toward the target object by about 10 ms to 50 ms (Pelz et al. 2001; Smeets et al. 1996). However, depending on the exact instruction the head may also anticipate saccadic activity by about 200 ms (Pelz et al. 2001). The reported times elapsed from the first saccade to the beginning of the reaching phase are 300 ms (Pelz et al. 2001) and 170 ms to 240 ms (Smeets et al. 1996). The visuomotor coordination was also studied in similar displacement activities, where the participants were required to grasp an object and move it in a target location. When interacting with household items (i.e., a bottle, a cup, and a can) the average lag between the beginning of the first fixation on the target object and the onset of the hand motion was 253 ms (Belardinelli et al. 2016).

Within the scope of this thesis, the time that passes from the first fixation until the actual grasp of the object is of primary interest. Johansson et al. (2001), who studied the pick and place task of a bar with and without obstacles along the path, reported a delay of about 1 s. In the same task the eyes also preceded the arrival of the bar at the destination by about 800 ms. Similar eye-hand delays ranging between 0.53 s to 1 s were found while displacing a cup and a box of pasta (Lavoie et al. 2018), and while drinking from or passing a cup, a bottle, and a can (Belardinelli et al. 2016). Lastly, it is important to underline that the amount of anticipation of gaze decreases when the user observes a task done by someone else (Flanagan and Johansson 2003; Sciutti et al. 2013). Although all studies confirm the anticipatory nature of gaze, they do not always agree on the exact timing of the motor execution after the first visual fixation, for instance when the hand reaches the object. These discrepancies can probably be explained by differences in experimental setting (Smeets et al. 1996; Pelz et al. 2001), variability due to a small number of subjects, or difficulty in accurately analyzing a large number of trials.

Similar goal-oriented gaze strategies were also reported during ADLs. In a tea-making task Land and Hayhoe (2001) observed that the activation of the arm follows the first saccade on the target object on average 0.56 s later. During walking (Patla and Vickers 2003) or driving (Land and Lee 1994), on the other hand, the eyes were found to fixate the forthcoming location about 0.5 s to 1 s in advance, while in some sports the eyes precede the final position of the ball by about 100 ms (Land and McLeod 2000; Hayhoe et al. 2012).

3.3 Integration of Vision in Prosthetics

The previous studies provided solid evidence that the grasp of an object is often preceded by a visual fixation on the same object. In an early work, Castellini and Sandini (2006) hypothesized that this proactivity of gaze could be used to improve the control of a prosthetic device or in a teleoperation scenario. To the best of our knowledge, since then only a few studies have actually explored this potential integration to improve intent recognition for upper limb prostheses. Hao et al. (2013) proposed a system based on EOG, which, as explained in Section 2.3, tracks the eyes via electrodes places on the face. The users scanned the object they intend to grasp with their gaze and, based on the estimated size, one of four possible grasps was selected. This mechanism essentially allowed the user to preshape the prosthesis,

which was then activated with a wrist movement to perform the desired action. Although the subjects were able to easily select the grasp, the overall method is impractical due to the electrodes that must be placed on the face and the two phase control strategy. A conceptually similar approach was subsequently proposed by Markovic et al. (2014), who substituted the electrodes with a more practical pair of augmented reality glasses, which were also used to provide the user with artificial proprioceptive feedback. As before, the user was required to fixate the target object, which was visually segmented by the glasses when recognized. The augmented reality interface provided a visualization of the grasp that was selected on the basis of the visual information. The user could then decide to initiate the grasp via sEMG control. An overview of the entire process is reported in Figure 3.5. A similar approach was proposed by the same group in a follow-up work (Markovic et al. 2015) in which the augmented reality was substituted with custom-designed glasses with an embedded camera. The object closest to the center of the camera image was segmented to estimate its shape, size, and orientation. This information was then used to predict the grasp, which could again be triggered by the user via the myoelectric interface. Even if not directly used for the grasp selection, also Clemente et al. (2016) proposed a simple form of augmented reality feedback by visually providing the user information on the grip closure and force during a displacement action.

Even if all studies report positive results, we note several disadvantages in the proposed approaches. The methods impose an unnatural form of control, since the user is required to fixate or scan an object until the employed visual system has managed to recognize it. This means that the time elapsed from the first fixation to the grasp is long and this time increases when recognition errors occurs. Moreover, following this approach the control modalities, such as vision and sEMG, are used sequentially rather than simultaneously, so a mistake in one of these steps also affects subsequent stages.

3.4 Application of Gaze Tracking in Robotics

The study of the visual behavior has also relevant applications in the field of assistive robotics to improve the indirect control of a robotic arm. For instance, the gaze position has been used to operate an exoskeleton in virtual (Novak and Riener 2013) and real (Bergamasco et al. 2011) environments to grasp some objects. In both cases, the participants used their gaze to “indicate” the object they intend to grasp; this information was then employed to assist the user in the grasping movement. A similar strategy was also adopted by McMullen et al. 2013, who developed a system for tetraplegic patients implanted with electrocorticographic electrodes. Also in this case the object of interest was identified via eye tracking and computer vision, and a prosthetic limb was activated via brain-control using intracranial electroencephalography. In a stroke rehabilitation context, a similar approach based on the fusion of visual information and electroencephalography was used to control an exoskeleton (Frisoli et al. 2012). For tetraplegic patients it has also been proposed to combine object localization via gaze with sEMG collected from the shoulder. The latter input modality was then used to control the trajectories of

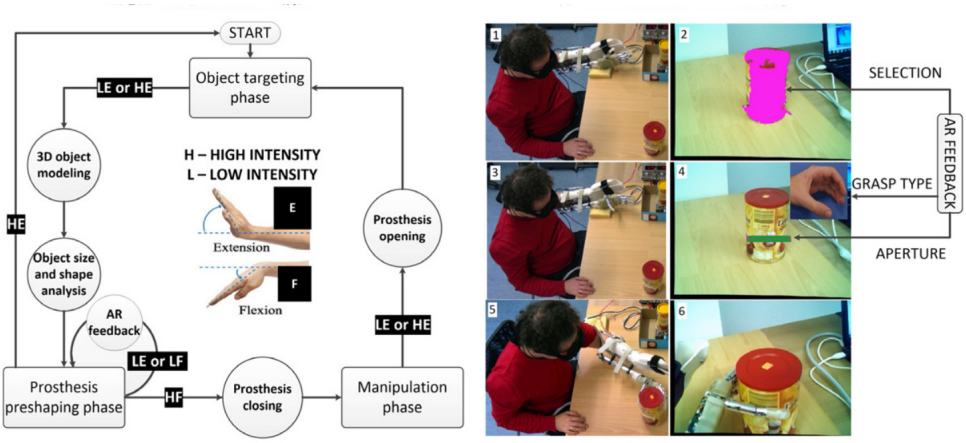


Figure 3.5. Overview of the experimental structure proposed by Markovic et al. (2014) to use augmented reality, visual information, and sEMG to improve the prosthetic control. The control loop (left) is a state machine with the transition triggered by the sEMG. During the control cycle (right) the augmented reality glasses segment the object and propose the grasp to take it. Figure credit: Markovic et al. (2014).

a robotic arm (Corbett et al. 2012; Corbett et al. 2014).

Aside from controlling robots, gaze and visual information has also been used to collaborate with them in the context of human robot interaction. Aronson et al. (2018) noted that during robot teleoperation and shared manipulation tasks, the gaze of the users follows similar patterns that can be exploit to improve the human robot collaboration. Similarly, Palinko et al. (2016) compared an eye tracking and head tracking approach via quantitative and subjective evaluations in a human robot collaborative task, concluding that in the former case the human robot interaction results smoother. In a teleoperation scenario, Latif et al. (2009) demonstrated that the gaze can be used to operate a robot, allowing to free the hand of the people from the control-joystick. Moreover, the eye patterns are also predictive of teleoperation failure and can be used as a signal to prevent errors (Aronson and Admoni 2018).

3.5 Visuomotor Coordination with Prostheses

A few studies have started to investigate the gaze behavior of amputated subjects while performing activities with their prosthesis. These studies aimed to understand whether the presence of this device affects the visuomotor coordination of the users, in terms of a deviation from the reference behavior described in studies with unimpaired subjects (see Section 3.2). In a small case study, Sobuh et al. (2014) compared the behavior of upper limb amputees using their own prostheses and intact subjects using a prosthesis simulator while pouring water into a glass. Both groups did not use gaze to proactively plan subsequent actions in a task, instead they appeared more concentrated on the ongoing manipulation. Other studies have also found that, during manipulation activities, amputees tend to switch their gaze between the object and the prosthetic hand more often than intact subjects would do using their own hand (Bouwsema et al. 2012; Hebert et al. 2019). This behavior

is probably meant to monitor the proper functioning of the prosthesis to compensate for uncertainty in the grasp security due to the lack of tactile and proprioceptive feedback. The disruption of gaze strategies is also characterized by a markedly longer reaching phase, while still maintaining the majority of the fixations on the target object (Sobuh et al. 2014; Hebert et al. 2019). Similar finding were also reported when able-bodied subjects are engaged in manipulation tasks using a prosthetic simulator (Blank et al. 2010; Sobuh et al. 2014; Parr et al. 2018; Parr et al. 2019). Almost all of these studies investigated this disruption in eye-hand coordination precisely for this reason, namely to measure the subject's proficiency in controlling the prosthesis. Indeed, Chadwell et al. (2016) noted that one participant who used a prosthesis daily showed more natural gaze behavior than another less experienced participant, while Sobuh et al. (2014) observed a shorter fixation on the hand area with increasing practice.

Chapter 4

Building and Validating a Dataset for Visual Integration

Some of the studies presented in the previous chapter integrated visual information for prosthetic control. However, contrary to the objectives of the present work they required the subjects to adapt their gaze behavior to accommodate the system. Moreover previous approaches were only tested on few subjects. The Myo-Electricity, Gaze And Artificial-intelligence for Neurocognitive Examination & Prosthetics (MeganePro) project aimed to overcome these limitations and to provide the community with a public dataset of myoelectric and gaze data. A total of 15 upper-limb amputees and 30 intact subjects were engaged in several static and functional activities to evaluate both the reach-to-grasp and the manipulation actions on multiple household items. The specific tasks, objects, and grasp types were chosen after preliminary evaluations to capture in the acquisition protocol as much variability as possible. These tasks were followed by other behavioral and clinical tests unrelated to the scope of this thesis.

Although the improvement of prosthetic control is the primary motivation for the acquisition of this dataset, we strongly believe that it can also be employed in other fields, as for instance neuroscience and rehabilitation. These data allow to study the visuomotor strategy of intact and amputated subjects in a wide range of situations (see Chapter 6). This is not only an interesting psychometric study, indeed the clear understanding of coordination parameters represents a prerequisite for the prosthetic applications. Moreover, since the control group of intact subjects was chosen to match in terms of age and gender with the recruited amputees, also comparative studies among the two groups would be possible. In the end, since the dataset is publicly released, scientists can use it to test new algorithms in the field of offline prosthetic control without the need to acquire data.

We will open the chapter by introducing the devices involved in data collection, since these have been used in all the acquisitions we made. In Section 4.2, we present a preliminary version of the dataset and the analyses done to establish the final acquisition protocol. In Section 4.3, we give a detailed description of the final dataset, including the involved subjects, the protocol and the data processing routine. We then provide in Section 4.4 several technical validations to ensure the soundness of

the data.

4.1 Acquisition Setup

We designed an acquisition setup suited to the collection of gaze and visual data, and sEMG from the forearm. It is composed of electrodes to acquire the electromyographic signals, eye tracking glasses to record the gaze activity and head movements, and a laptop to store the recorded data. The devices are shown in Figure 4.1. For the experiments we took care to position them so as to not interfere with natural movements during the tasks.

The Delsys Trigno Wireless sEMG System (Delsys Inc., USA)[1] is composed of sEMG electrodes (see Figure 4.1a) used to collect the muscle's activity from the forearm of the participants. The sEMG signal is sampled at 1926 Hz with a baseline noise of less than 750 nV RMS and an inter-sensor latency lower than 500 µs. Each electrode has an embedded three-axial accelerometer that collects data with a sampling rate of 148 Hz. The electrodes communicate via wireless with a base station that is connected to the laptop.

The Tobii Pro Glasses 2 (Tobii AB, Sweden)[2] (see Figure 4.1b) were used to gather gaze and visual information. These lightweight (45 g) and wearable glasses are equipped with four eye cameras to track the eyes, a frontal camera to record the scene, an inertial measurement unit (IMU) for the head movements, and a microphone. The eyes are tracked using corneal reflection and dark pupil methods (see Section 2.3) with automatic parallax and slippage compensation. Gaze and gaze-related information are sampled at 100 Hz, while the video is recorded with a (1920 px × 1080 px) resolution at 25 frames per second. The theoretical accuracy and precision for eye tracking data is 0.5° and 0.3° RMS, the horizontal and vertical field of view of the camera is approximately 82° and 52°. The glasses are connected via a cable to a portable recording unit that locally stores the data and transmits them via wireless to the laptop. The recording unit also contains battery that allows a maximum recording time of approximately 120 min. The system can quickly and easily be calibrated with a standard single point calibration procedure. As supplementary feature, the glasses are equipped with corrective lenses, which can be applied in case of visual deficit, and several nose pads, which regulate the position of the device to ensure good tracking and comfort.

4.2 Preliminary Data

To define the final protocol we acquired a preliminary dataset in which similar grasping activities were performed in different conditions from a small group of subjects. The aim was to evaluate whether the recorded sEMG is influenced by the constraints posed for a task (Fougner et al. 2011; Geng et al. 2012; Peng et al. 2013; Khushaba et al. 2014) to eventually include these factors in the final dataset. After the description of the acquisition protocol (Gigli et al. 2018) we briefly explain the processing routine of the raw data and introduce the features and classification

[1] `http://www.delsys.com/`
[2] `http://www.tobiipro.com/`

(a) Trigno electrodes **(b)** Tobii glasses

Figure 4.1. Acquisition setup. In (a) the Delsys Trigno electrodes used to acquire the activity of the forearm are shown. In (b) the Tobii Pro Glasses 2 to collect visual data are presented.

algorithms used to analyze sEMG and accelerometer. Later we will present the analyses performed on these data and the obtained results.

4.2.1 Dataset

Five intact subjects (4M, 1F) were engaged in the acquisition. For our exercises we selected ten grasps from hand taxonomy literature (Cutkosky 1989; Sebelius et al. 2005; Crawford et al. 2005; Feix et al. 2009; Feix et al. 2015) on the base of their relevance in ADLs (Bullock et al. 2013). We also combined with each of them three objects that can be naturally manipulated using the respective grasp. The objects were chosen with the aim to ensure a many-to-many relationship, such that an object does not uniquely identify a grasp and vice-versa. Aside from multiple objects, the acquisition protocol was extended in two other manners to test variability of myoelectric signals. First, the subjects were engaged both in static and dynamic tasks. In the former case they were only asked to reach and grasp the objects, without further manipulation, and to keep the desired hand configuration for few seconds. In the dynamic part the same grasps were used to perform functional tasks on a subset of previous objects. In this case the subjects were asked to complete a manipulation action while keeping the required grasp during the whole activity. This introduces variability in the dynamic context of the hand by producing the crosstalk given by wrist and limb movements. The third form of variability is represented by the arm's orientation during the tasks. To study the limb position effect, all the static grasps were executed both seated and standing. The functional tasks were instead performed only in one manner, depending on the most common orientation in ADLs. The tasks, grasps, and objects are reported in Table 4.1.

Before the beginning of the exercise, twelve sEMG electrodes were placed in two arrays around the right forearm. The first array, composed of eight electrodes, was placed at the height of the radio-humeral joint. The second array was located approximately 45 mm below. As widely done in the field of pattern-recognition applied to myoelectric control, the sEMG electrodes were placed around the proximal part of the forearm, where most of the muscles of the hand lie, following an untargeted approach (Hakonen et al. 2015; Hargrove et al. 2007b), without mark specific muscles'

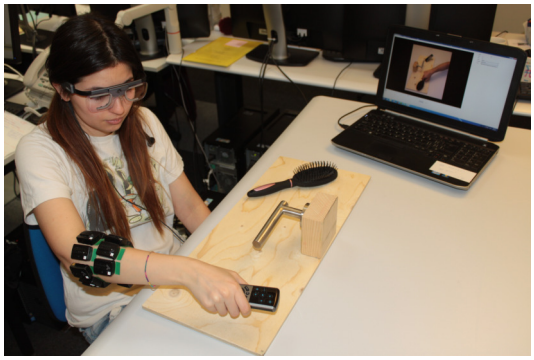


Figure 4.2. Overview of the acquisition protocol and setup during the execution of a task.

belly. A latex-free elastic band was wrapped around the electrodes to guarantee full adherence to the skin. The subject also wore the Tobii glasses and followed the calibration procedure.

During the exercises the subject was in front of a table on which at least five objects were placed to ensure visual clutter and to provoke a visual search for a given item. Before each grasp the subject was trained with a short video to clarify how the objects should be approached and to encourage familiarity with the exercise. By means of an automatic text-to-speech, the computer indicated both the beginning and the end of each grasp. Each movement-object combination was repeated four times and each grasp repetition lasted approximately 5 s followed by 4 s of rest. An overview of the acquisition protocol and setup is shown in Figure 4.2.

4.2.2 Data Processing and Classification

The sEMG and gaze data were timestamped in a shared reference time. These timestamps were used to synchronize all modalities and to upsample them to the sampling rate of sEMG. The sEMG data were filtered to remove the powerline interference and the label of each grasp was corrected taking into account the actual muscular activation by following the approach described by Kuzborskij et al. (2012). Furthermore these data were standardized to have zero mean and unitary standard deviation, based on statistics calculated exclusively on training data. For the classification we adopted the window-based approach introduced by Englehart and Hudgins (2003) and described in Section 2.2.3. We segmented the data using a sliding window of 200 ms and an increment of 10 ms (i.e., 20 samples).

After a feature extraction step we used previous segments to train and test a classifier. For the sEMG data we used the mDWT as feature representation and a Kernel Regularized Least Squares (KRLS) with exponential χ^2 kernel as classifier. From the accelerometer data the MAV was extracted and used as feature in combination with a KRLS classifier with RBF kernel. These features, classifier, and kernel types were previously introduced in Section 2.2.2 and Section 2.2.3 and were chosen for the good results presented on similar data by Gijsberts et al. (2014).

Table 4.1. List of grasp types, and static and dynamic tasks acquired in the preliminary experiments. During the static grasps the subject was required to reach and grasp three objects without further interactions both while seated and standing. In the dynamic part the subject was instead asked to manipulate a subset of the objects to execute some actions while seated or standing.

Grasp	Static Tasks	Dynamic Tasks
medium wrap	take the **can** take the **door handle** take the **bottle**	drink from the **can** (standing) open and close the **door** (standing)
lateral	take the **key** take the zip of the **pencil case** take the **cup**	turn the **key** in the lock (standing) open and close the **jacket** (standing)
parallel extension	take the **plate** take the **book** take the **drawer**	lift the **plate** (standing)
tripod grasp	take the **bottle** take the knob of the **drawer** take the **cup**	open and close the cap of the **bottle** (standing) open and close the **drawer** (standing)
power sphere	take the **ball** take the **light bulb** take the **key**	move the **ball** to the right and back (standing)
precision disk	take the **jam jar** take the **light bulb** take the **ball**	open and close the lid of **jam jar** (seated) screw and unscrew the **light bulb** (seated)
prismatic pinch	take the **clothespin** take the **key** take the **can**	squeeze the **clothespin** (seated)
index finger extension	take the **remote control** take the **knife** take the **fork**	press a button on the **remote control** (seated) cut bread with the **knife** (seated)
adducted thumb	take the **screwdriver** take the **remote control** take the **wrench**	turn the **screwdriver** (seated)
prismatic four fingers	take the **fork** take the **knife** take the **wrench**	move the **fork** to the right and back (seated)

4.2.3 Test on Variability

We perform several classification experiments to study the deterioration of the accuracy when train and test data were acquired in the different experimental conditions listed in previous section. The aim of this analysis is to examine whether the sEMG recorded when performing a specific grasp results dependent on task-related factors, such as the limb position or the grasped object. The results should help to understand whether include or not these types of variability in the final protocol. To this goal we consider four settings, each of them represents a different train-test split of the data.

***Posture-split*:** Train and test data were split based on the posture of the subject, so that seated data were used to train a model and standing data to test it. This split is meant to evaluate the limb position effect.

***Dynamic-split*:** We included in the training the static tasks and in the test group the functional activities. This split makes it possible to understand whether and how wrist and limb movements influence the grasp.

***Object-split*:** We included two of the three objects used for each grasp type in the training set and the remaining one in the test set. We considered all possible combinations of the three objects and averaged the final accuracy results. The aim is to understand how object's characteristics influence the sEMG.

***Trial-split*:** We use three repetitions of each object-grasp combination to train a classifier and the remaining one to test it. All possible combinations of the four trials were considered and the obtained classification performance were averaged per subject. In this case train and test data have the same variability level. This is the standard approach considered in literature when multiple repetitions of the same movement are acquired, it is therefore provided as reference.

The hyperparameters of the classifiers are optimized using k-fold cross validation on the training set where each of the folds correspond to one of the grasp-repetitions in the training data. We adopted a dense grid search to choose the hyperparameters: $\lambda \in \{2^{-16}, 2^{-15}, \ldots, 2^{2}, 2^{3}\}$, and $\gamma_{rbf}, \gamma_{\chi^2} \in \{2^{-20}, 2^{-19}, \ldots, 2^{0}, 2^{1}\}$. For computational reasons the training data were downsampled with a factor 10, while the data used for hyperparameter optimization were downsampled with an additional factor 4.

In Figure 4.3, Figure 4.4, and Figure 4.5 the classification accuracy obtained in the *posture-split*, *dynamic-split*, and *object-split* experiments is compared against the *trial-split* setting. In each experiment the ten grasps (plus the rest posture) are classified when the input data are only sEMG, only accelerometer, and a combination of them. We report both the performance scored by each single subject (i.e., the five points in each column of the plots) and the mean value (i.e., the gray triangle).

The results of the *posture-split* experiment in Figure 4.3 clearly show poor performance when the subject's posture, and therefore arm's orientation, changes between train and test data. These results confirm the limb position effect already highlighted by several studies in the evaluation of grasps collected in multiple positions (Fougner et al. (2011), Geng et al. (2012), and Khushaba et al. (2014)).

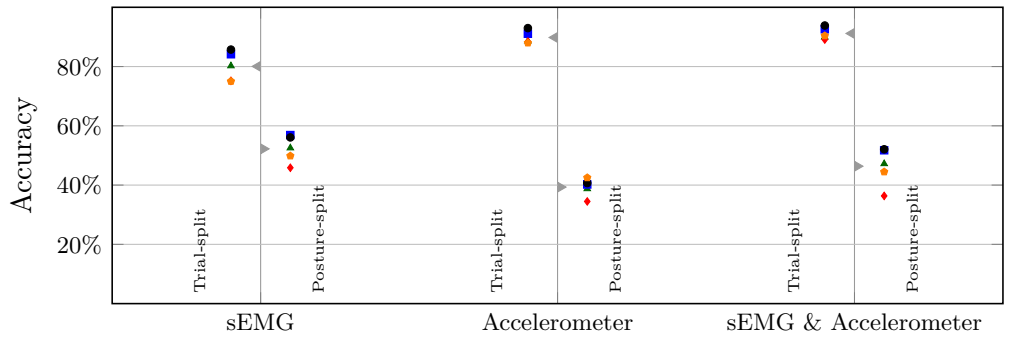


Figure 4.3. Comparison between the classification accuracy of the *trial-split* and *posture-split* settings. From left to right we report the performance in classifying the ten grasps and rest from sEMG, accelerometer, and a combination of them. Each symbol corresponds to a subject and the gray triangle indicates the average.

All these studies report an inter-position performance significantly lower than the intra-position one. Similarly, also the accuracy observed when passing from static to functional movements is lower than 60 % as the *dynamic-split* of Figure 4.4 highlights. Also Liu et al. (2014) showed that the absence of dynamic movements in the training data seriously impacts on the final performance. The classification error increases due to effects of wrist rotation (Peng et al. 2013), forearm orientation, and muscular contraction during a grasp (Khushaba et al. 2016). Similar results are also obtained when varying the objects in the *object-split* experiments of Figure 4.5, even if the degradation in performance is lower than previous cases. This means that factors like the object's weight or shape influence partially the sEMG and accelerometry during grasp.

This degradation disappears, however, in the *trial-split* setting where all the forms of variability are integrated both in the train and test data. In this case the average accuracy is greater than 80 %, for both the single modalities and their combination. The solution of pooling all data to enrich the training phase is not innovative and was also adopted by previous studies (Jiang et al. 2013; Boschmann and Platzner 2013; Masters et al. 2014). Whether this solution is better than other approaches, like hybrid method with classifiers trained in multiple positions (Radmand et al. 2014), is an open question.

4.3 The MeganePro Dataset

The results presented in the previous section show that factors like arm orientation, grasp dynamics, and object weight and shape dramatically influence the accuracy in posture classification. To obtain less domain-specific data we decided to include all the variability factors previously evaluated in the MeganePro dataset (Cognolato et al. 2019). The processed version of the acquired data is stored in Harvard's Dataverse. In the following we will present the final experimental protocol and setup, the subjects engaged in our experiments, the objects that they were required to manipulate, and the performed tasks.

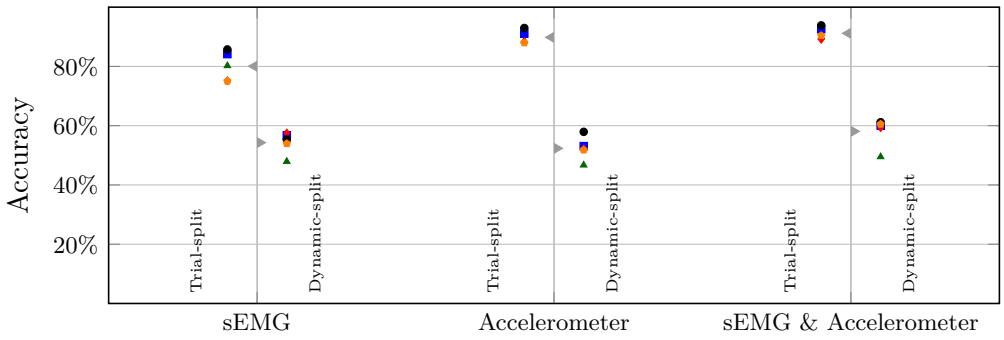


Figure 4.4. Comparison between the classification accuracy of the *trial-split* and *dynamic-split* settings. From left to right we report the performance in classifying the ten grasps and rest from sEMG, accelerometer, and a combination of them. Each symbol corresponds to a subject and the gray triangle indicates the average.

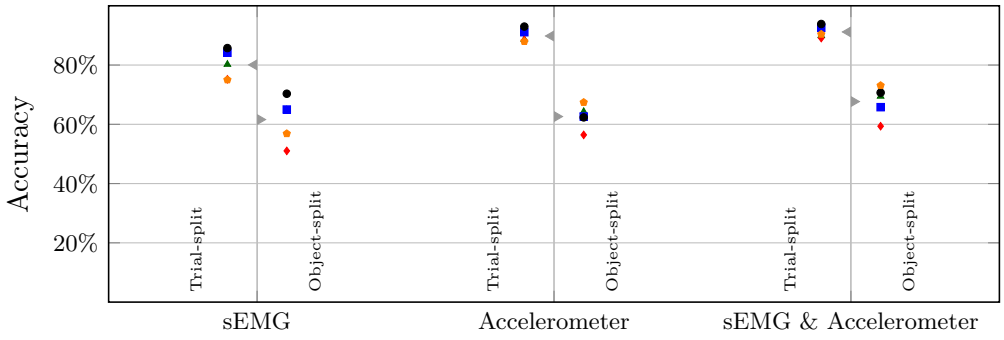


Figure 4.5. Comparison between the classification accuracy of the *trial-split* and *object-split* settings. From left to right we report the performance in classifying the ten grasps and rest from sEMG, accelerometer, and a combination of them. Each symbol corresponds to a subject and the gray triangle indicates the average.

4.3.1 Subject Recruitment

A total of 15 transradial amputees (13 M, 2 F; age: (47.13 ± 14.16) years) and 30 age- and gender-matched intact subjects (27 M, 3 F; age: (46.63 ± 15.11) years) participated in this study. In Table 4.2 we report the important characteristics of all the subjects and, for the first group, the information related to the amputation.

The experiment was designed and conducted in accordance with the principles expressed in the Declaration of Helsinki. The study was approved by the Ethics Commission of the canton of Valais in Switzerland and by the Ethics Commission of the Province of Padova in Italy. The subjects received a detailed written and oral explanation of the experiment and they were required to give informed consent for the participation.

4.3.2 Grasp Types and Objects

For the final acquisition we kept the same grasps used in the preliminary experiments (see Section 4.2.1). Also in this case each grasp was matched with three household objects that would naturally be manipulated with this grasp. After preliminary evaluations, some of the items used in previous acquisition have been replaced by others. For instance the worn jacket was replaced with a pencil case since, when opening and closing the zip of the former object, the subject's gaze often falls outside the field of view of the glasses preventing a good tracking. Moreover, in the case that an object is not usually found on a table (e.g., a door handle or a door lock), a custom made support was created. To avoid complications during the exercise, the key, bulb, lid of the jar, and the screw used with the screwdriver were modified such that they could not be completely removed from the support. As before, we took care to ensure a many-to-many relationship between objects and grasps. An overview of grasps and objects is shown in Table 4.3.

4.3.3 Acquisition Protocol

The acquisition started after the ethical requirements were fulfilled and the devices worn. As in the preliminary acquisition twelve electrodes were organized in two arrays around the forearm. The positioning is clarified in Figure 4.6. The subject was then asked to wear the Tobii glasses and to follow the standard one point target calibration procedure. This procedure is fast and generally requires only a few seconds. Once the calibration was completed we prepared an acquisition for the calibration assessment. This consisted in asking the subject to fixate on a black cross against a green background that was displayed on a monitor at a distance of about 1.3 m. The cross was in turn showed in five locations of the screen remaining fixed on a single one for 3 s.

After these preparation steps the data collection started. Prior to executing a grasp, videos in first and third person perspectives were shown to clarify the required movement and the part of the object that should be involved. The subject was however instructed to execute the movements as naturally as possible, rather than attempting to mimic the exact kinematics of the demonstration videos. The amputated subjects were required to execute the action with their missing limb, rather than just imagining it, as to elicit muscle activations in their residual limb.

Table 4.2. Participant characteristics. The table reports the ID of the subjects in the dataset, their age, their gender and their handedness. Clinical parameters about the amputation(s) are also reported for the transradial amputees. The rightmost column indicates the relative length of the residual limb with respect to the contralateral limb.

						Amputation				
	ID	Age	Gender	Handedness	Lang.	Side	Cause	Years	Prosthesis	Limb [%]
Intact Subjects	10	27	M	right	EN					
	11	63	M	right	FR					
	12	49	M	right	FR					
	13	32	M	left	FR					
	14	67	M	right	DE					
	15	68	M	right	DE					
	16	38	M	right	FR					
	17	63	M	ambidextrous	FR					
	18	55	M	right	FR					
	19	29	M	right	FR					
	20	48	M	left	FR					
	21	62	M	left	FR					
	22	39	M	right	FR					
	23	53	M	right	FR					
	24	29	M	right	FR					
	26	45	M	right	FR					
	27	68	M	right	FR					
	28	62	M	right	FR					
	29	58	M	right	FR					
	30	66	M	right	FR					
	31	39	M	right	FR					
	32	34	M	right	EN					
	33	69	M	right	FR					
	34	57	M	right	DE					
	35	29	F	ambidextrous	EN					
	36	28	M	right	IT					
	37	31	M	right	EN					
	38	29	F	right	EN					
	39	33	F	ambidextrous	EN					
	40	29	M	right	FR					
Transradial Amputees	101	52	M	right	IT	right	electrocution	2	cosmetic	60–80
	102	39	M	right	IT	right	electrocution	4	cosmetic	60–80
	103	63	M	ambidextrous	IT	right	trauma	3	myoelectric	60–80
	104	49	M	right	IT	right	trauma	18	myoelectric	80–100
	105	73	M	right	IT	right	trauma	6	body-powered	40–60
	106	70	M	left	IT	left	trauma	5	body-powered	80–100
	107	36	M	right	IT	left	trauma	7	body-powered	20–40
	108	35	M	right	IT	right	trauma	9	myoelectric	0–20
	109	65	M	right	IT	left	trauma	1	cosmetic	80–100
	110	38	M	right	IT	left	trauma	14	myoelectric	20–40
	111	38	M	right	IT	right	trauma	10	myoelectric	40–60
	112	33	F	right	IT	left	oncological	13	cosmetic	60–80
	113	28	M	right	IT	left	trauma	7	myoelectric	40–60
	114	52	M	right	IT	bilateral	trauma	35	myoelectric	n/a
	115	36	F	right	IT	left	burn	8	cosmetic	n/a

Table 4.3. Overview of the objects and the grasps of the MeganePro dataset. For each row we show three objects and the associated grasp with an illustrative picture. Figure credit for the second column: Atzori et al. 2014b.

Grasp		Objects		
medium wrap		bottle 	door handle 	can 
lateral		mug 	key 	pencilcase 
parallel extension		plate 	book 	drawer 
tripod grasp		bottle 	mug 	drawer 
power sphere		ball 	bulb 	key 
precision disk		jar 	bulb 	ball 
prismatic pinch		clothespin 	key 	can 
index finger extension		remote 	knife 	fork 
adducted thumb		screwdriver 	remote 	wrench 
prismatic four finger		knife 	fork 	wrench 

Figure 4.6. Position of the electrodes around the forearm of an amputated subject.

As in the preliminary version, we divided the acquisition in static and dynamic tasks. For each grasp, the former consist in reaching and grasping three objects without lifting or moving them. Each object-grasp combination (see Table 4.3) was repeated four times both while seated and standing. The latter tasks are instead composed of several functional activities. For each grasp two of the previously used objects were manipulated with the desired hand configuration as reported in Table 4.4. Each activity was repeated four times while seated or standing, depending on what posture seems more likely in daily life for the given task. After each movement the subjects were required to return to the rest position. A vocal command marked the beginning and the end of each grasp and announced the activity to perform. The rest and movement periods terminated respectively when the explanation of the required action and a specific “release" command were vocally completed. Therefore the duration of the movement and rest periods depends on the selected language for the vocal instructions. A grasp interval lasted on average 5.2 s, 5.7 s, 5.9 s, and 6.0 s for English, Italian, French, and German. A rest period followed for about 4.1 s, 4.7 s, 4.7 s, and 4.7 s for English, Italian, French, and German. The exact order of the objects within each repetition was randomized to avoid learning and habituation effects. Moreover, to encourage visual search, additional objects were placed on the table besides the ones involved in the task.

4.3.4 Data Processing

A number of processing steps were applied to the raw data acquired with the protocol described above. The objective of these steps was to sanitize the data, synchronize all modalities, and remove identifying information from the videos. In the following we describe all procedures in detail.

4.3.4.1 Timestamp Correction

Due to an unfortunate implementation error, during a number of acquisitions the modalities were assigned timestamps from individual clocks. To unify all timestamps in a shared clock, the offset of all clocks was estimated and corrected with respect to the clock of the sEMG modality using statistics of their relative timing collected during trial acquisitions. Validations on the remaining unaffected acquisitions

Table 4.4. Overview of the dynamic tasks. The vocal instruction in English indicates the task that had to be performed for each object-grasp pair. The last column indicates whether the subject performed the task while seated or standing.

Grasp	Vocal Instruction	Position
medium wrap	drink from the **can** open and close the **door handle**	standing
lateral	turn the **key** in the lock open and close the **pencil case**	standing
parallel extension	lift the **plate** lift the **book**	standing
tripod grasp	open and close the cap of the **bottle** open and close the **drawer**	standing
power sphere	move the **ball** to the right and back move the **keys** forwards and backwards	standing
precision disk	open and close the lid of **jar** screw and unscrew the **light bulb**	seated
prismatic pinch	squeeze the **clothespin** move the **keys** forwards and backwards	seated
index finger extension	press a button on the **remote control** cut bread with the **knife**	seated
adducted thumb	turn the **screwdriver** move the **wrench** to the right and back	seated
prismatic four finger	move the **knife** forwards and backwards move the **fork** to the right and back	seated

confirm that the maximum deviation of our estimate from the ground truth is less than 12 ms.

4.3.4.2 sEMG and Accelerometer Data

For computational efficiency, the sEMG and accelerometer streams from the electrodes were acquired and timestamped in batches. During post-processing, individual timestamps were assigned to each sample via piecewise linear interpolation. A new piece is created if the linear model would result in a deviation of more than 100 ms, which may happen if the fit is skewed due to missing or delayed data.

For the sEMG data, we furthermore filtered outliers by replacing samples that exceeded 30 standard deviations from the mean within a sliding window of 1 s with the preceding sample. The signals were subsequently filtered for power-line interference at 50 Hz (and its harmonics) using a Hampel filter (Allen 2009). Contrary to the more common notch filter, this method does not affect the spectrum if there is no interference.

4.3.4.3 Gaze Data

The data from the Tobii Pro glasses were acquired as individually timestamped JavaScript Object Notation (JSON) messages. During post-processing, these messages were decoded and separated based on their type. The messages that relate directly to gaze information, such as gaze points, pupil diameter and so on, were then grouped together based on their timestamps.

4.3.4.4 Stimulus

The text-to-speech engine that was used to give vocal instructions introduced noticeable delays in the corresponding changes of the stimulus. We measured these delays for all sentences and languages, and moved the stimulus changes forward by these amounts during post-processing. For each object also the more specific object-part involved in the exercise was calculated and added to the stimulus information.

4.3.4.5 Synchronization

All modalities were resampled at the original 1926 Hz sampling rate of the sEMG stream. For real-valued signals, this was done using linear interpolation, while for discrete signals we used nearest-neighbor interpolation. The signals that indicate the time and index of the MP4 video were handled separately using a custom routine, since they require to identify the exact change-point where one video transitions to the next.

4.3.4.6 Concatenation

The static and dynamic parts of the protocol were acquired independently and therefore produced separate sets of raw acquisition files. Furthermore, our acquisition protocol and software allowed to interrupt and resume the acquisition, either at request of the subject or to handle technical problems. After applying the previous

processing steps to the individual acquisition segments, they were concatenated to obtain a single data file per subject. During this merging, we incremented the timestamps and video counter to ensure that they are monotonically increasing. Furthermore, if part of the protocol was repeated due to interruptions or problems with acquisition software, when resuming the acquisition we took care to insert the novel segment at exactly the right place to avoid duplicate data.

4.3.4.7 Relabeling

The response of the subject and therefore the sEMG activation may not be aligned perfectly with the stimulus. As a consequence, the stimulus labels around the on- or offset of a grasp movement may be incorrect, resulting in an undue reduction in recognition performance. We addressed this shortcoming by realigning the stimulus boundaries with the procedure described by Kuzborskij et al. (2012). In short, this method optimizes the log-likelihood of a rest-grasp-rest sequence on the whitened sEMG data within a feasible window that spans from 1 s before until 2 s after the original grasp stimulus. As opposed to the uniform prior used in the earlier method, we instead adopted a smoothed variant of the original stimulus label as prior with $p = 0.6$. The recalculated stimulus boundaries have been saved in addition to the original ones.

4.3.4.8 Removing Identifying Information

All videos were checked manually for identifying information of anyone other than the experimenters. The segments of video that were marked as privacy-sensitive were subsequently anonymized with a Gaussian blur. In this procedure, we took care to re-encode only the private segments and to preserve the exact number and timestamps of all frames. In addition, the audio stream was removed from all videos for privacy reasons.

4.4 Technical Validation of the Dataset

The intended purpose for the dataset is the investigation of the sEMG and visual data and, eventually, their fusion. In this section we therefore concentrate on validating these two modalities. Some of these analyses are low-level to ensure the quality of the recorded signals, while others are meant to verify that the dataset can in fact be used for the motivations for which it was created.

4.4.1 Gaze Validation

The quality of gaze data primary depends on the correctness of the initial calibration phase recommended by manufacturers of the Tobii glasses before data collection. To validate this calibration we analyzed the dedicated data acquisitions in which a black cross was shown on a green background of a screen (see Figure 4.7 and Section 4.3.3) with the aim to compare the expected and the measured gaze position. Besides data quality estimation, we are further interested in verifying the naturalness of user's

gaze throughout an exercise. To this aim we also calculated distributions of fixations and saccades and compared them with previous studies.

4.4.1.1 Validation of Calibration

Validating the effectiveness of the calibration of the glasses consists in acquiring gaze data while the user is focused on a known target and subsequently comparing the measured gaze location with this ground truth (Holmqvist et al. 2012; Blignaut and Wium 2014). We used the data recorded during the calibration assessment described in Section 4.3.3 to evaluate the effectiveness of the calibration as well as possible accuracy degradation over time. These data were typically collected at the beginning and end of an exercise. If the exercise was interrupted, the procedure was shown again before resuming it. We determined the ground truth by manually locating the cross position in pixels at intervals of 0.2 s using custom software as shown in Figure 4.7. Since we also included calibration data for the other clinical exercises done jointly as part of the MeganePro project, a total of 498 acquisitions were processed in this manner.

The quality of eye tracking is often quantified in terms of accuracy and precision (Holmqvist et al. 2012; Reingold 2014; Blignaut and Wium 2014). For each axis, the former measures the systematic error, which is the mean offset between the actual and expected gaze locations. Precision, on the other hand, measures the dispersion around the gaze position and thus the random error of the gaze point. In Figure 4.8 these values are visualized with respect to the location within the video frame. This separation is intentional, as the eye tracking appears to be more accurate and precise in the center of the frame, namely (-3.5 ± 19.4) px and (-1.5 ± 29.6) px on the x and y axes. Moving away from the center, the gaze results systematically shift towards the borders of the frame and its random error increases. We only visualize regions where we acquired at least 40 validation samples. Pooling all data, the overall accuracy and precision is (-0.8 ± 25.8) px and (-9.9 ± 33.6) px on the horizontal and vertical axes. At a typical manipulation distance of 0.8 m, this corresponds to a real-world precision and accuracy of approximately (-0.4 ± 11.5) mm and (-4.4 ± 14.9) mm. This is deemed sufficiently accurate considering the size of the household objects used in our experimental protocol.

To establish whether the calibration deteriorated over time, we compared the accuracy and precision collected at the beginning of an acquisition with those taken at the end. In total, we considered 210 uninterrupted acquisitions in which there was a calibration validation routine both at the beginning and the end. We found no statistically significant difference in accuracy (sign test, $p = 0.95$ and $p = 1.0$ in the horizontal and vertical axes) or precision (sign test, $p = 0.24$ and $p = 0.37$), indicating that drift does not pose an issue for the gaze data.

4.4.1.2 Statistical Parameters

To statistically describe a user's gaze behavior during the exercises and validate it against related literature, we first identified fixations and saccades in our eye tracking data using the IVT method by Salvucci and Goldberg (2000) that was also described in Section 2.3.2. To ensure that we could calculate the angular velocity of both

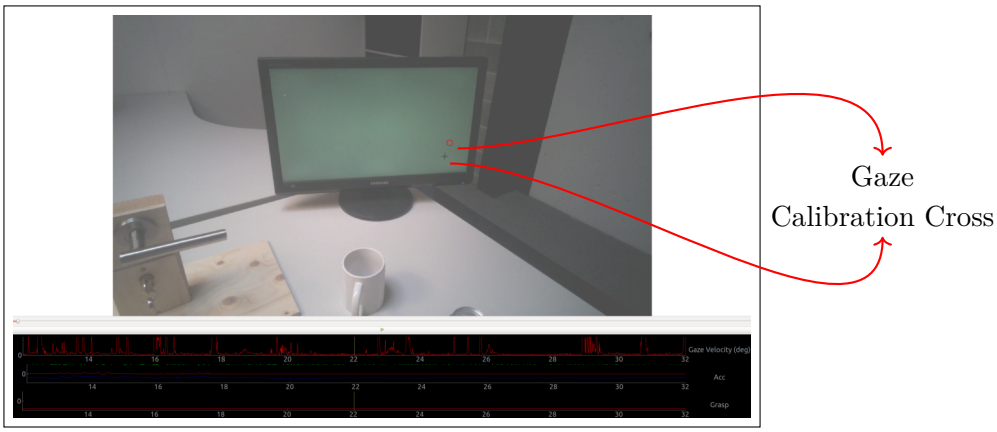


Figure 4.7. Custom software to manually acquire the cross position in frame coordinates. The red circle represents the 2-dimensional gaze position recorded by the Tobii glasses.

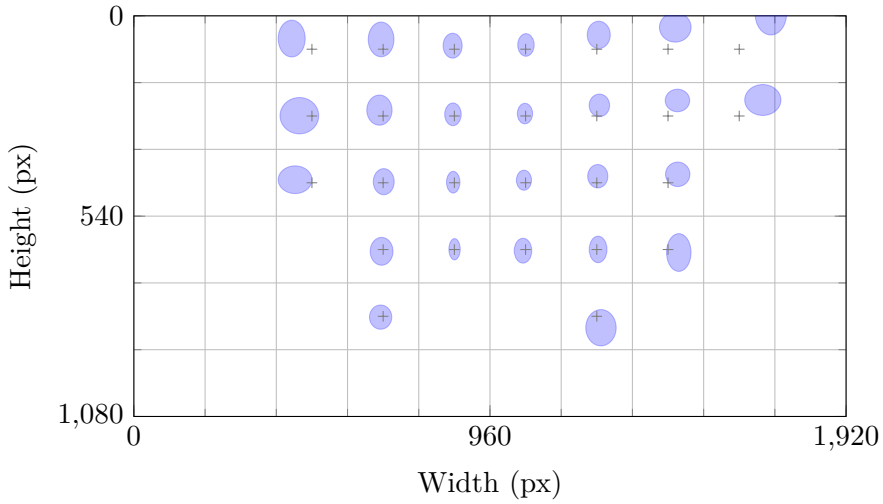


Figure 4.8. The accuracy and precision of the eye tracking with respect to the location within the video frame. For each patch, the shift of the ellipse center with respect to the cross indicates the accuracy in either axis of the gaze within that patch. The radii of the ellipse on the other hand indicates the precision.

eyes for a maximum number of samples, we linearly interpolated gaps of missing pupil data when shorter than 0.075 s (Olsen 2012). We used a threshold of 70 °/s to discriminate between fixations and saccades (Komogortsev et al. 2010). When the Tobii glasses failed to produce a valid eye-gaze point, even after interpolating small gaps, the corresponding sample was marked as *invalid*. Excluding one subject who had strabismus, the percentage of such invalid samples ranged between 1.7 % to 21.0 % and 4.3 % to 30.7 % for able-bodied and amputated subjects. Sequences of events of the same type were then merged into segments identified by a time range and processed following the approach described by Komogortsev et al. 2010. First, to filter noise or other disturbances, fixations separated by a short saccadic period of less than 0.075 s and 0.5° amplitude are merged. Second, fixations shorter than 0.1 s are marked as *invalid* and excluded from the analysis.

In the resulting sequence of gaze events, the majority of invalid data are located between two periods of saccades, namely (92.2 ± 2.7) % and (92.6 ± 3.9) % for able-bodied and amputated subjects. This indicates that the Tobii glasses fail predominantly to register high velocity data. Devices with sampling frequency lower than 250 Hz have indeed been categorized as "fixation pickers" (see Karn 2000) and often do not provide reliable results for saccades. For this reason in the following analysis, rather than considering many short saccades due to interruptions of invalid segments, we indicate as saccade the period ranging from the end of one fixation to the beginning of the following.

Figure 4.9 and Figure 4.10 show the distributions of fixation durations and of saccade amplitudes for both types of subjects. The characteristics of these distributions, summarized in Table 4.5, coincide with those described in analogous studies (Johansson et al. 2001; Kinsman et al. 2012; Duchowski 2017). Median values and the interquartile range are comparable with the results of Johansson et al. (2001), who report 0.286 s and 3.2° as median duration of fixations and median amplitude of saccades and 0.197 s to 0.536 s and 1.5° to 7.1° as range for the distributions between the same percentiles. Moreover, the mean value of the duration of fixations is similar to the mean duration of around 0.5 s reported by Hessels et al. (2017). Land et al. (1999) also indicated that saccades dealing with near objects range between 2.5° to 20°. This is coherent with the MeganePro protocol where all the objects were placed on a table in front of the subjects. These similarities confirm the quality of eye tracking data and highlight that the subjects maintained a natural gaze behavior throughout the exercise. Interestingly, we also do not note important differences between the distributions on intact and amputated subjects.

4.4.2 Myoelectric Signals

To assure the soundness of the recorded sEMG, we first analyzed the spectral properties and compared these with known results from literature. As a more high-level validation we verified that the sEMG signal can indeed be used to discriminate the grasp a subject was performing.

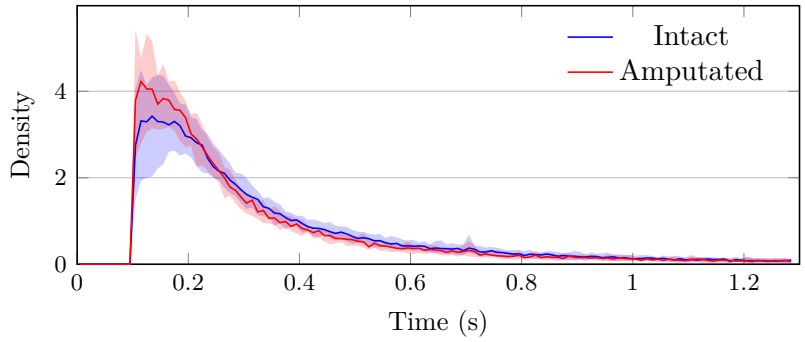


Figure 4.9. Distribution of the fixation length histogram for able-bodied (blue) and amputated (red) subjects. The shaded areas indicate the 10^{th} and 90^{th} percentiles, while the solid line represents the median.

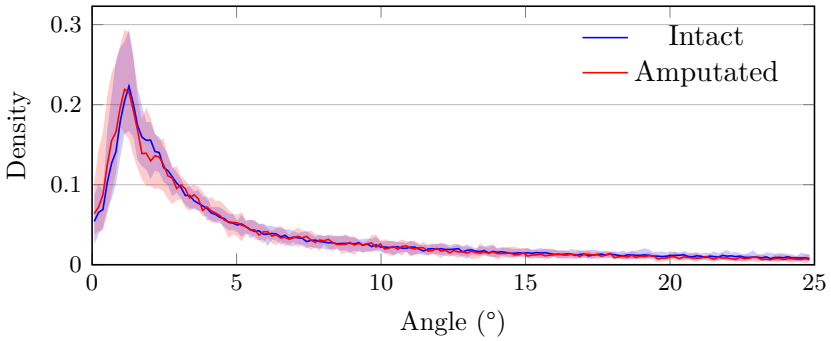
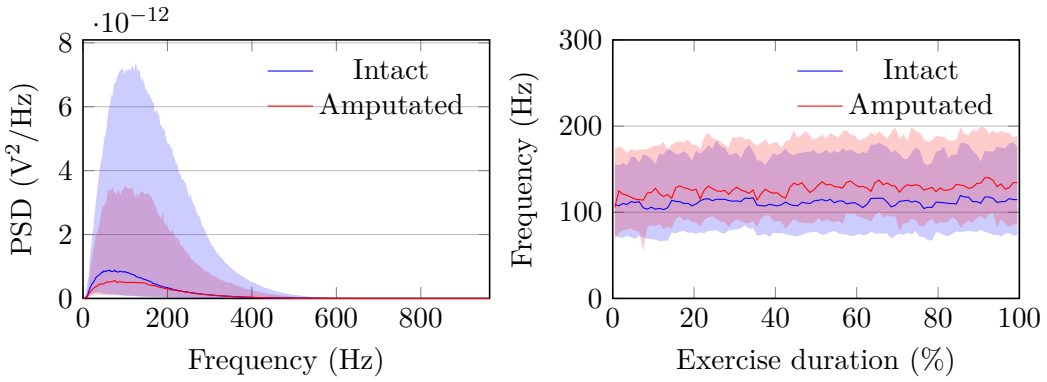


Figure 4.10. Distribution of the saccade amplitudes histogram for able-bodied (blue) and amputated (red) subjects. The shaded areas indicate the 10^{th} and 90^{th} percentiles, while the solid line represents the median.

Table 4.5. Statistical parameters of the duration of fixations and the amplitude of saccades for able-bodied and amputated subjects.

	Subjects	Mean	Percentiles		
			25th	50th	75th
Fixations	Intact	0.429 s	0.170 s	0.260 s	0.470 s
	Amputated	0.432 s	0.160 s	0.240 s	0.440 s
Saccades	Intact	7.754°	1.662°	3.883°	10.960°
	Amputated	7.377°	1.561°	3.720°	9.942°



(a) Power Spectral Density. **(b)** Median Frequency.

Figure 4.11. The distribution (a) of the power spectral densities and (b) the median frequency throughout the duration of the exercise. The solid line indicates the median over all electrodes for intact (blue) and amputated (red) subjects, while the shaded area indicates the 10^{th} and 90^{th} percentiles.

4.4.2.1 Spectral Analysis

For each subject and for each channel, we calculated the power spectral density via Welch's method with a Hann window of length 1024 (approximately 530 ms) and 50% overlap. From the distribution of these spectral densities, shown in Figure 4.11a, we note that nearly all of the energy of the signals is contained within 0 Hz to 400 Hz, as is typical for sEMG (Basmajian and De Luca 1985). Furthermore, there is no sign of powerline interference at 50 Hz or its harmonics, confirming the efficacy of the filtering approach detailed in the Section 4.3.4.2. In reference to the same figure we, however, observe a large variability of densities among subjects and electrodes. The spectrum and amplitude of sEMG signals depend on the positioning of an electrode over a muscle (De Luca 1997b). Since in our protocol none of the electrodes was positioned precisely on a muscle belly, the signal is variable and in some cases almost absent.

The same variability is also noticeable in Figure 4.11b, which reports the distribution of the median frequency over all electrodes for intact and amputated subjects throughout the entire exercise. The median frequencies we find are close to the approximately 120 Hz to 130 Hz typically reported for the flexor digitorum superficialis (Clancy et al. 2008; Kattla and Lowery 2010), which is one on the muscles we primarily recorded from with our electrode positioning. Finally, we note that the distribution of the median frequency remains relatively stable over time, indicating that there are no persistent down- or upward shifts in the spectrum.

4.4.2.2 Grasp Classification

To classify the sEMG signals we segment the data by employing the standard window-based approach (Englehart and Hudgins 2003) described in Section 2.2.3 with a window length of 400 samples (approximately 208 ms) and 95 % overlap between successive windows. As feature-classifier combinations we consider:

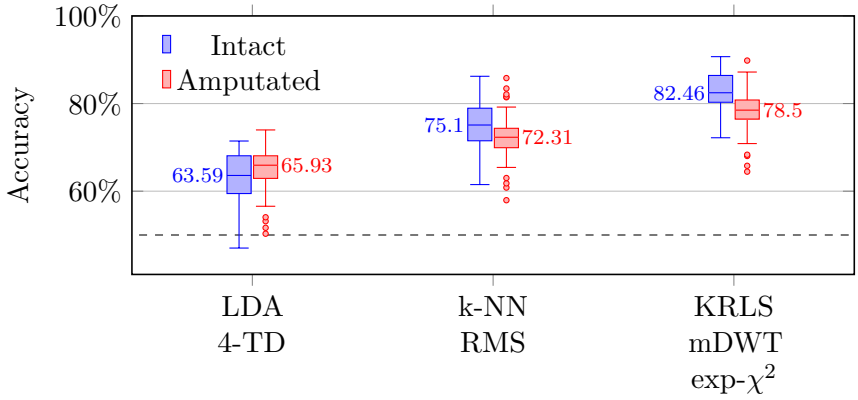


Figure 4.12. Classification accuracies for able-bodied and amputated subjects when predicting the grasp type with three different types of classifiers. The dashed line refers to the baseline accuracy a classifier would achieve by simply predicting the most frequent *rest* class.

- a (balanced) LDA classifier used with the popular four time-domain features (Englehart and Hudgins 2003);
- K-Nearest Neighbors (KNN) applied on RMS features (Atzori et al. 2014a); and
- KRLS with a nonlinear exponential χ^2 kernel and mDWT features (Gijsberts et al. 2014; Atzori et al. 2014a).

The classification accuracy is defined as the average accuracy of 4-fold cross validation. In each of these folds, three repetitions of each grasp-object-position combination were used as training data and the remaining repetition was employed as test. This division corresponds to the *trial-split* presented in Section 4.2.3. Similarly, any hyperparameters were optimized via nested 3-fold cross validation on the train repetitions. For all methods, the training data were downsampled with a factor 10 for computational reasons, while the data used for hyperparameter optimization were downsampled with an additional factor 4. All these steps were equally performed for the analyses on the preliminary dataset in Section 4.2.3.

The results of classification accuracy, reported in Figure 4.12, show accuracy between 63% and 82% for both intact and amputated subjects, depending on the classification method. A baseline classifier, which always predicts the most common grasp, would have achieved only approximately 50% accuracy. Indeed the *rest* follows each repetition of each grasp, it occurs then in about an half of the whole acquisition. The current results are similar to the results obtained in the preliminary evaluation (Section 4.2.3). Although a quantitative comparison with related work is of limited value due to discrepancies in experimental setup and protocol, the current results are a couple percentage points higher than those presented by Atzori et al. (2014b). The most likely explanation is the lower number of grasps (i.e., only 10 rather than 40), which inevitably boosts performance.

Chapter 5

Automated Analysis

The previous chapter introduced a dataset containing both visual and sEMG data. The Tobii glasses recorded a huge amount of visual data: the whole dataset is composed of more than 70 h of videos. Since the video is recorded at 25 frames per second, we collected more than 6 300 000 frames. Moreover, the eye tracker samples the gaze at 100 Hz providing on average four gaze coordinates per frame. These large numbers make it unpractical to manually annotate the gaze position in the scene to evaluate the attention of each subject. In this chapter, we present a method to automatically analyze these visual data.

In other psychophysical studies, the recorded videos are often manually annotated frame-by-frame to analyze where the focus of the subject is during the execution of the experiment (see for instance Land et al. 1999). This approach limits the amount of experimental data that can be analyzed and leaves room for subjective interpretation of these data. An alternative is to use the software provided by manufacturers of eye tracker devices, which may automate certain types of analyses (Bowman et al. 2009; Belardinelli et al. 2016). Those solutions however do not exploit the most recent developments in machine learning and computer vision. Consequently, they still require a considerable amount of manual effort per video and the overall solution is not fully automated.

We propose a method that minimizes the required human intervention by automatically extracting all information of interest from the recorded data (Gregori et al. 2019). It achieves this by leveraging over state-of-the-art deep learning techniques to detect and segment all objects of interest (see Table 4.3) from all videos. The first step of this procedure consisted in an efficient method to collect a few dozen segmentations per object class from a sequence of frames. Rather than generate these manually, we instead used a deep learning algorithm to facilitate the creation of a dataset of segmented items. These data were used to fine-tune a pretrained object detector, which is later employed to identify and segment all known objects in the entire video sequence. Finally, these segmentations are combined with the gaze coordinates to calculate metrics relating gaze with the objects and the user's own limb.

In Section 5.1, we present the method used to build the training dataset and in Section 5.2 we explain how this was used to fine-tune a deep neural network for

inference on all the MeganePro videos. In Section 5.3, we will describe the metrics that this approach allowed us to calculate.

5.1 Creation of the Training Dataset

To collect a training dataset containing binary segmentations of the MeganePro objects, we made use of a recent algorithm for real time video tracking. In the following, we first present this method and we later show how this was embedded in a custom application for data collection.

5.1.1 Introduction on SiamMask

The semi-supervised SiamMask method has recently been proposed by Wang et al. (2019) to track the position of one or more arbitrary objects in a video sequence. By marking just a bounding box around an object in one frame, this deep convolutional algorithm (1) segments the object from the background and (2) tracks it in the following frames.

The majority of approaches for object tracking have tackled the problem of identifying the target object by means of a bounding box around it (Wu et al. 2013; Kristan et al. 2018; Valmadre et al. 2018). For our purposes, a shortcoming of this approach is that the bounding box includes part of the background as well. Object segmentation methods address this problem by supplementing the bounding box with a binary segmentation mask, which for each pixel indicates whether it belongs or not to the object of interest. Such pixel-level accuracy of course requires more computational effort, which is why many proposed methods are not able to process a video stream in real-time (Wen et al. 2015; Tsai et al. 2016; Yang et al. 2018; Bao et al. 2018).

SiamMask has been proposed with the aim to provide the community with a fast video object segmentation method by overcoming both previous limitations. Its architecture takes inspiration from Siamese networks (Bromley et al. 1994). In general, a Siamese architecture consists of two parallel neural networks, each receiving one input image, and a final shared part of the network that computes the similarity between the two images. Bertinetto et al. (2016) followed this approach to address a tracking problem by feeding the network with two images: an annotation of the target object provided by the user as reference, and a new frame in which this reference should be located. These two images and the Siamese part of the network are shown on the left hand side of Figure 5.1. This part of the network produces a measure of similarity ($^\star d$) for all possible candidate regions within the new frame with respect to the reference image. The output is then used to train three tasks jointly, as shown on the right hand side of Figure 5.1. For each region, the objective of the three tasks is respectively to assign a score, a refined bounding box, and a binary mask that segments the object. The final prediction during inference is then given by the candidate region that has the maximum score.

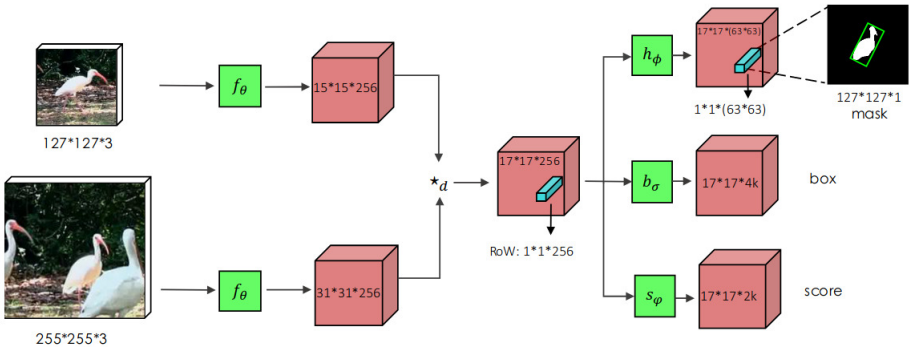


Figure 5.1. Schematic architecture of SiamMask. The network takes as input the frame cropped around the target (top left) and a new frame (bottom left) and computes their similarities ($\star d$). The last part of the network is composed of three branches: two for bounding box identification and refinement, and one for the binary mask proposal. Figure credit: Wang et al. (2019).

5.1.2 Application on the MeganePro Dataset

In the present study we opted to use the SiamMask method for its advantages with respect to other available approaches. First, to precisely estimate where the subject is focused, we prefer a binary segmentation of the object rather than a simple bounding box. Second, with real-time performance the building of the dataset is fast and can interactively be supervised by the user. Although it may seem tempting to run this tracking algorithm on an entire video annotating each object only at its first occurrence, in practice it does not work sufficiently reliably on such long time scales and when objects may exit and enter the visual scene. We therefore used this method to amplify our manual annotations; with just a single bounding box annotation per object, we obtain 10 to 20 times as many binary segmentation masks for our training set.

We embedded the official implementation of SiamMask[1] in a custom application. The network is composed of a ResNet-50 (He et al. 2016) until the final convolutional layer of the 4$^{\text{th}}$ stage combined with convolutional layers on the top. The software that we implemented to build the dataset with segmented objects works as follows.

1. The user selects a frame from a video recorded by the Tobii glasses.

2. This frame becomes the input image for SiamMask and it is manually annotated by the user by drawing a bounding box around the objects of interest and by selecting their class identity.

3. Based on this initialization, SiamMask processes the input image and subsequent frames one by one and presents the output segmentations to the user for validation.

4. For each frame the user can either accept and store, or refuse the proposed segmentations and proceed with the processing of the following frame (3-4).

[1] https://github.com/foolwood/SiamMask

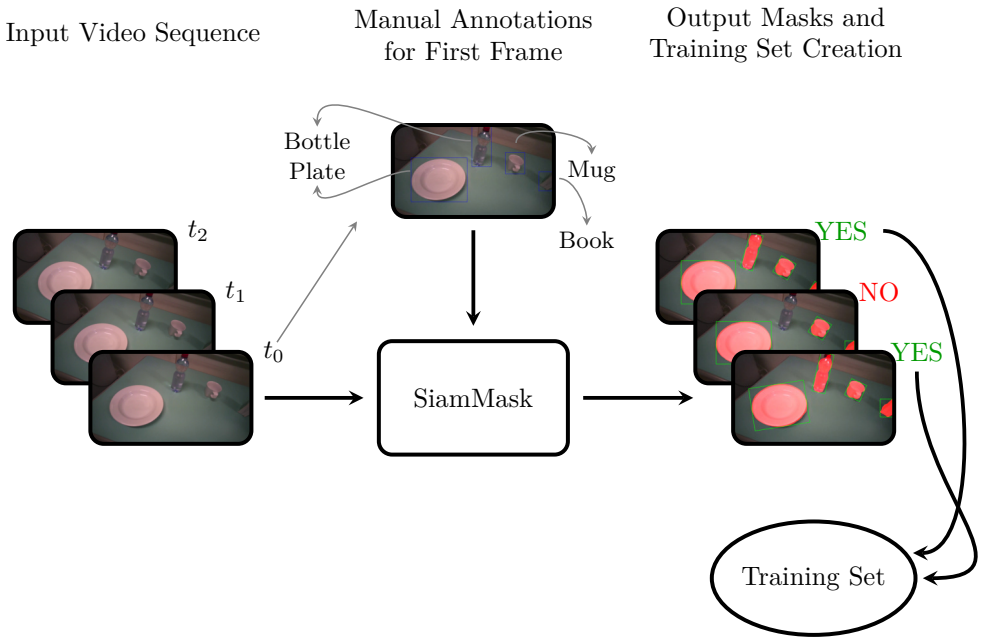


Figure 5.2. The procedure to acquire the training set of segmentation masks. We first select an arbitrary frame from a video and annotate each object with its bounding box and object identity. This information is passed to SiamMask, which produces segmentation masks for this initial frame and the subsequent frames in the video sequence. At each frame, the user can choose whether or not to include the frame and its segmentations in the training set or to move to a new initial frame.

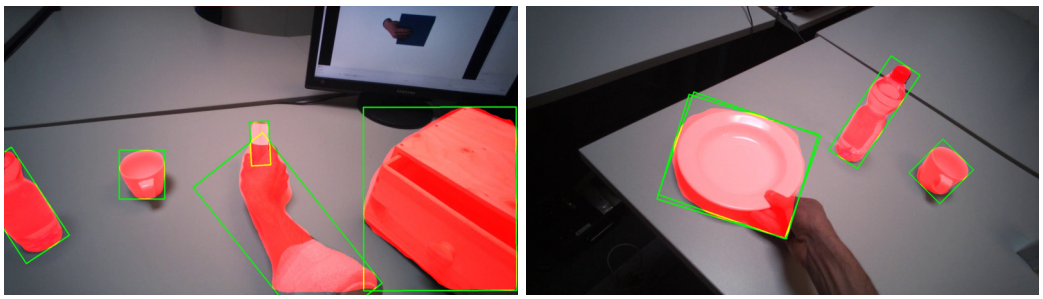
Alternatively, the user may stop the procedure for the current sequence of frames and start at step (1) with a newly selected initial image.

This procedure is also schematically shown in Figure 5.2.

In practice, for each selected initial frame we accepted sequences up to around 15 frames. Applying this procedure repeatedly, we segmented in total 2422 frames with 11726 segmented object instances chosen from 15 subjects. To include as much variability as possible in our dataset, we captured the objects from different perspectives, with different backgrounds, and while partially occluded. Furthermore, besides the eighteen objects in Table 4.3, we also included segmentations for a "person" class, which will primarily be used to detect the subjects' forearm. Two frames with the segmented instances are presented as example in Figure 5.3.

5.2 Object Segmentation

The collected dataset of masks was then used to fine-tune a deep neural network to segment the objects in all MeganePro videos. In the following we will describe the used framework and how we modified it for our purposes.



(a) Good segmentation. **(b)** Wrong segmentation.

Figure 5.3. Example of two frames extracted from the MeganePro videos and segmented using SiamMask. In (a) all the objects are well segmented, in this case the output masks were all stored for the training dataset. In (b) the arm of the subject is not recognized, its mask overlays instead with the plate's mask. In this case all the masks were discarded.

5.2.1 Introduction on Mask R-CNN

Mask R-CNN is a multi-task framework for object classification, detection, and segmentation proposed by He et al. (2017). The detection and segmentation tasks consist in identifying the position of an object respectively via a bounding-box and with a binary mask at the pixel-level. Whereas SiamMask tracks an arbitrary annotated object in a sequence of frames, Mask R-CNN instead segments and classifies all instances of objects that it recognizes in a single frame.

The structure of Mask R-CNN is schematically shown in Figure 5.4. Mask R-CNN is an extension of Faster R-CNN, which was proposed by Ren et al. (2015) for object detection. Faster R-CNN is composed of two stages. Initially, the network proposes candidate object bounding boxes for an input image. For each of these, the network extracts some features to both propose a classification label and refine the original bounding box. Mask R-CNN adds to the previous architecture a third distinct parallel branch that is trained to create a binary segmentation for each identified object. Like SiamMask, the architecture is therefore composed of parallel branches, each with a specific role. This task decoupling differentiates the current approach from previous slower and less accurate methods (Pinheiro et al. 2016; Dai et al. 2016).

Multiple architectures have been considered on Mask R-CNN. In particular the backbone, which is the first part of the network for feature extraction, was implemented with ResNet-50 (He et al. 2016), ResNeXt (Xie et al. 2017), and ResNet-50-FPN (Lin et al. 2017). The head, which is the top part of the network, was instead taken from Faster R-CNN with the addition of a small Fully Convolutional Network (FCN) (Long et al. 2015) for segmentation.

5.2.2 Inference on the MeganePro Dataset

The data we acquired via SiamMask were used to train Mask R-CNN on our objects of interest (see Table 4.3). Rather than training a model "from scratch", we instead bootstrapped from a model that was supplied with the implementation of Mask

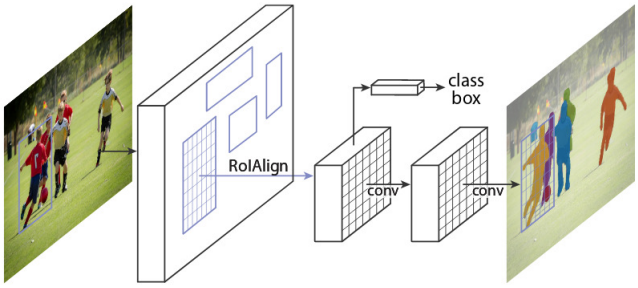


Figure 5.4. Schematic architecture of Mask R-CNN. The first part of the network proposes several regions of the input image as candidate object bounding boxes. Later each region is individually processed. A branch of the network proposes the final bounding box and the object's class, another branch outputs the segmentation. Figure credit: He et al. (2017).

Table 5.1. Comparison of Mask R-CNN's detection accuracy on the COCO dataset and the accuracy of our finetuned model on the MeganePro dataset. The AP is the average precision over IoU from 0.5 to 0.95 evaluated at steps of 0.05. AP50 and AP75 represent the average precision when the threshold of IoU is 0.5 or 0.75. A detailed description of these metrics can be found on the website of the COCO dataset (`http://cocodataset.org`).

Dataset	AP [%]	AP50 [%]	AP75 [%]
MeganePro	77.5	92.7	87.6
COCO (He et al. 2017)	33.6	55.2	35.3

R-CNN by Massa and Girshick (2018). This model used a ResNet-50-FPN backbone and was pretrained on the Common Objects in COntext (COCO) dataset (Lin et al. 2014), a large scale generic dataset for object detection, segmentation, and classification. As is common with fine-tuning, we replaced the final classification layer of the model with a random initialization and then performed additional training iterations with a reduced learning rate of 0.0025 to tailor the model to our custom dataset. The data of ten subjects were used for training, while the validation set consisted of the data of the remaining five subjects, which were chosen to be as representative as possible for the entire dataset. We chose to use the model that minimized the loss on the validation set (i.e., early stopping), which was obtained after just 4000 iterations. The performance of this model is compared in Table 5.1 with the Average Precision (AP) metrics of the pretrained model on the original COCO dataset. Note that, due to the limited domain of our dataset and the smaller number of classes, our performance compares favorably to the larger COCO dataset. After training, we employed the model in inference mode to detect and segment objects in all videos of all subjects, as shown graphically in Figure 5.5.

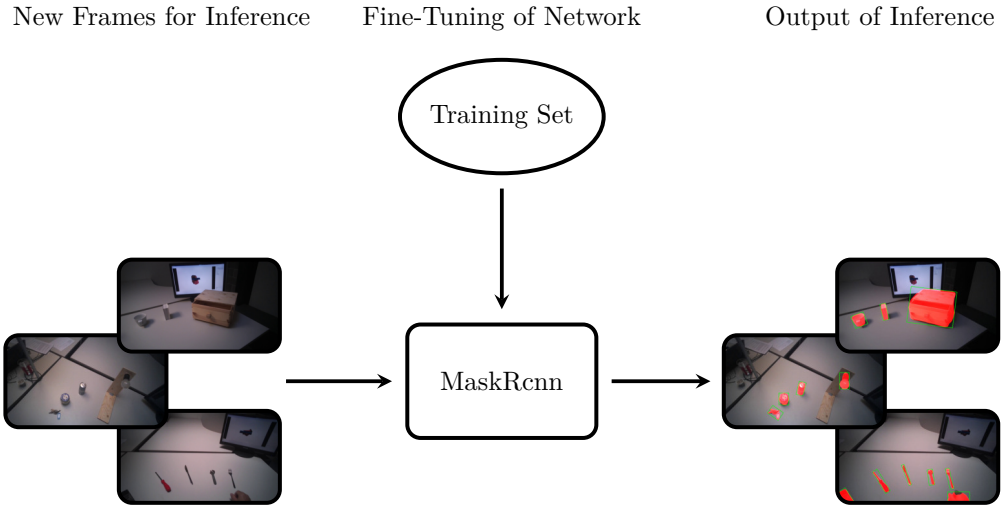
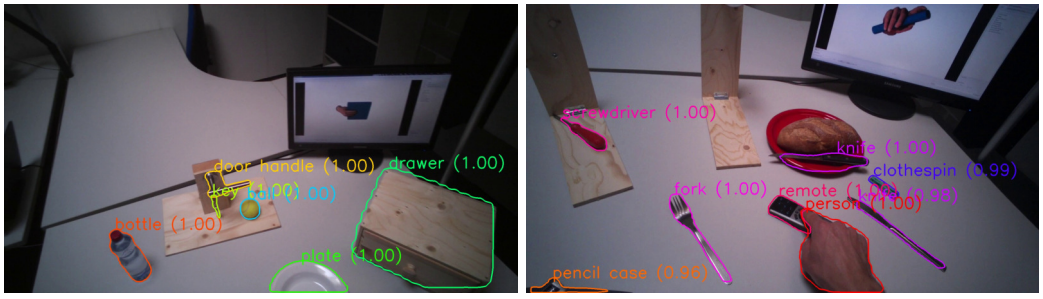


Figure 5.5. Procedure for the segmentation the whole dataset. By means of the previous selected training set we fine-tune the Mask R-CNN model. Later we feed the network with new frames and the network provides the segmented and classified objects as output.



(a) Good segmentation. **(b)** Wrong segmentation.

Figure 5.6. Example of two frames extracted from the MeganePro videos and segmented by Mask R-CNN. In (a) all the objects are well segmented and classified. Even the plate, which is only partially captured, was recognized with a high score (i.e., 1.0) like the other items. Also in (b) the majority of the objects are well segmented and classified, except for two objects. The wrench is well segmented but confused with a pencil case, while the segmentation mask of the screwdriver does not include the entire object.

5.3 Distances

The segmentation masks for all videos were stored to disk and then combined with the gaze data to calculate various distances in pixels. In the following, we restrict ourselves to segmentations that were recognized with a certainty score of at least 0.8. The distances that are of interest for our analyses are the following.

- The *gaze-target* distance, which is the distance between the gaze point in frame coordinates and the target object for a grasp trial, if visible in the frame. If multiple instances of the same target class were recognized, then we chose the largest in terms of area.
- The *gaze-limb* distance denotes the distance between the gaze point and the hand or residual limb of the participant, if visible. We only consider instances identified as "human" that fall in the lower half of the image frame and again prefer the largest one.
- When applicable, the *limb-target* distance indicates the distance between the subject's hand or residual limb and the target object, as defined in the previous two distances.

Note that with the term "distance" we intend the minimum Euclidean distance in pixels between a point and the contour of a binary mask or between the contours of two binary masks. If these overlap, then the distance is 0.

Chapter 6

Visuomotor Coordination of Amputated and Intact Subjects

The previous chapter presented an automatic framework to determine which object a subject is fixating, if any, and the distance of the limb from this object. In the following we will employ this information to answer two main questions (Gregori et al. 2019).

1. Can we determine the window of opportunity in which gaze provides useful information for intent recognition?

2. How does the visuomotor coordination of amputees compare with that of intact subjects?

After introducing the modalities involved in the analyses in Section 6.1, we present the results of the visuomotor coordination both in the reach-to-grasp and manipulation phases. These findings are discussed more thoroughly in Section 6.3 including analogies with related studies, the comparison between intact and amputated subjects, and comments regarding the potential prosthetic application.

6.1 Experimental Setup

In the following analyses we evaluate the gaze behavior by considering the previous calculated *gaze-target* distance (see Section 5.3) and the gaze velocity (see Section 2.3.3). The head movements are studied by means of the velocity provided by the gyroscope of the Tobii glasses (see Section 4.1). The activity of the hand and the forearm are collected by the electrodes placed on the arm and the embedded accelerometers (see Section 4.1). Moreover the hand's movements and positioning can be monitored by the *gaze-limb* and the *limb-target* distances (see Section 5.3). Given the scope of the present analyses, we only use sEMG from the second and seventh electrode, which were placed approximately on the extensor and flexor digitorum superficialis muscles (see Figure 6.1). Besides having relatively high activations, these electrodes also indicate roughly whether the hand was opening or closing. To aid visualization, both channels were rectified with a sliding RMS of approximately

Figure 6.1. The electrodes considered for the visuomotor analysis (circled) were placed approximately on the extensor and flexor digitorum superficialis muscles.

29 ms (i.e., 57 samples) (Merletti 1999). With respect to accelerometry, we note that the accelerations of all electrodes were highly correlated due to their positioning around the forearm. We therefore use accelerations only from the first electrode, which were normalized with respect to the inertial frame of the initial position in each movement repetition (Tundo et al. 2013).

6.1.1 Events

The profile of the distances described in Section 5.3 and the modalities described previously were used to determine the timing of visuomotor events with respect to the stimulus, such as the first fixation on the target object or the onset of the arm movement. These events allow us to quantitatively describe the time interval between the activation of the eyes, head, and limb. The analysis window for each trial ranges from 2 s before until 2.5 s after the end of the corresponding vocal instruction with a resolution of 20 ms. We define the following events.

- The first *fixation* is defined as the first of at least two successive samples where the *gaze-target* distance is less than 20 px. This threshold was chosen to accommodate for some systematic error in the gaze tracking and is roughly twice the average gaze tracking accuracy (Cognolato et al. 2020). The requirement for two successive samples that fall below the threshold is to ignore occasional outliers.

- The *saccade* to the target object is assumed to initiate at the last sample where the gaze velocity was less than $70\,^\circ\,\mathrm{s}^{-1}$ (Komogortsev et al. 2010), starting from 500 ms prior to the target *fixation*. This definition in terms of the last preceding fixation rather than the first saccade makes it robust against missing data from the eye tracker during saccades. Furthermore, we require this saccade to start from a *gaze-target* distance of at least 100 px to avoid occasional trials where the subject was already fixating the target object.

- The start of the *head* movement is defined as the first of two successive samples where the Euclidean norm of the angular velocity vector of the Tobii glasses

exceeds $12\,^{\circ}\,s^{-1}$. This threshold was chosen manually to be insensitive to systematic errors in the measurements of the gyroscope in the Tobii glasses.

- The movement of the *arm* starts at the first of two successive samples where the Euclidean norm of the three-axis accelerations exceeds 0.07 g. Also in this case the threshold was tuned manually to be insensitive to the baseline level of noise of the accelerometers.

- The activation of the *forearm muscles* starts when either of the myoelectric signals exceeds 4 times its baseline level for two successive samples. This baseline level is taken as the average activation in the rest period from 2 s to 1 s before the vocal instruction ended.

- Finally, the first *grasp* occurs when there are two successive samples where the *limb-target* distance is less than 5 px. This threshold was chosen to allow for a small error margin in the detected segmentation masks.

Whenever the conditions for an event were not satisfied it was marked as missing for the corresponding trial. Furthermore, we invalidate all events that were found within the first 100 ms of the analysis window, as it implies that the subject was not in a rest position or was already fixating the target object.

6.2 Eye, Head, and Hand Coordination

In the following we relate movements of the eyes and head with that of the forearm in response to the grasp stimulus during the reach-to-grasp and manipulation phases. With grasp stimulus we intend the vocal command that instructs the participants which task to execute. These analyses were performed on all the 30 intact subjects and 14 amputees collected in the MeganePro dataset. One of the amputated subjects suffered from strabismus and was therefore excluded from the study. Indeed this condition made it impossible to obtain reliable eye tracking data since the left and right gaze directions do not intersect neither pass one near the other (see Section 2.3). Before moving to these analyses, we verified that the subjects effectively looked at the target object during a grasp trial. Thanks to the deep learning approach described in Chapter 5, we determined that in 95.9 % of the trials the subjects looked at least once at the target (i.e., a *gaze-target* distance less than 20 px). Manual evaluation of the remaining 4.1 % of the trials revealed that these were caused by a low accuracy of the Tobii glasses rather than lack of subject engagement. Note that in this and following analysis the threshold of 20 px was chosen on the base of previous results of accuracy and precision of the Tobii glasses shown in Section 4.4.1.1.

6.2.1 Statistical Analysis

One of the objectives of this thesis is to determine the window of opportunity in which gaze can provide useful information about an upcoming grasp. Table 6.1 shows that for intact subjects there is a median interval of 561 ms between the *fixation* event and the subsequent *grasp* event. The same interval increases to more than a second for amputated subjects, although this difference is because they did not

Table 6.1. Statistical description of the intervals in seconds between various events. The count refers to the number of trials where both events were recognized, out of a total of 9703 trials for intact and 4482 trials for amputated subjects, respectively. For the Kolmogorov-Smirnov test the intervals were averaged per subject to guarantee independent samples. We highlight in bold the statistically significant differences between the two groups.

Interval	Intact				Amputated				Significance
	#	Q1	Med.	Q3	#	Q1	Med.	Q3	
		[s]				[s]			
fixation → grasp	8144	0.321	0.561	0.842	1942	0.581	1.042	1.644	$KS = 0.724, p = \mathbf{2.602 \times 10^{-5}}$
saccade → fixation	5625	0.080	0.160	0.301	2522	0.060	0.140	0.281	$KS = 0.190, p = 0.811$
saccade → head	5419	−0.301	0.020	0.160	2367	−0.461	−0.020	0.140	$KS = 0.338, p = 0.173$
head → arm	7929	0.020	0.120	0.301	3507	0.000	0.140	0.371	$KS = 0.262, p = 0.447$
arm → muscles	7907	−0.020	0.080	0.401	3576	0.200	0.581	1.042	$KS = 0.829, p = \mathbf{4.524 \times 10^{-7}}$

physically interact with the objects and the *limb-target* distance therefore did not as often converge to within the 5 px threshold. Not surprisingly, a Kolmogorov-Smirnov test on the average interval per subject indicated that this difference between both subject groups was statistically significant. This is in contrast to the coordination between the initial saccade, the head, and the arm movements, for which we fail to find a significant difference between both groups. The saccade to the target object leads to its fixation in approximately[1] 150 ms. Concurrently with the eyes, also the head starts to move. This head movement is then followed by acceleration of the arm around 130 ms later. In intact subjects, the activation of the forearm muscles comes only 80 ms after the onset of the arm movement in the median case. This interval is more than half a second longer for amputated subjects and this difference is found to be statistically significant.

6.2.2 Reach-to-Grasp Phase

The coordination during the reaching phase of all "static" and "functional" grasps is reported in Figure 6.2 for both intact and amputated participants. Whereas the previous statistical analysis was intended to provide a quantitative assessment of the relative timings in eye-hand coordination, this figure instead complements those numbers by demonstrating how this coordination evolves over time. It does so by showing the median and quartiles of the distribution over all trials from all subjects in either group from 1.5 s before to 2.5 s after the conclusion of the vocal instruction. For both types of subjects, we observe an increase in gaze velocity from approximately −0.5 s to 1 s. This increase also marks a sharp decrease in the distance between the gaze and the target object, which leads to a fixation soon after. From this moment on, the subjects retain their fixation on the object of interest for the grasp. Based on the median profiles, we see again that the onset of the head movement starts around the same time as the eye movement and continues for approximately 1.5 s.

The initiation of the arm movement follows the onset of the eye movements, as

[1] This is likely a slight overestimation, considering our definition of the *saccade* and missing values in the gaze data from the Tobii glasses.

shown by the median profile of the forearm's acceleration in Figure 6.2. Shortly after the arm starts to move, we also observe an increase in sEMG activity, with initially an emphasis on the extensor and later on the flexor. For able-bodied subjects, the profile of the *limb-target* distance confirms our earlier finding that the limb arrives at the object approximately 500 ms after its fixation. Although this result is not directly comparable with that for amputated subjects, we observe that the convergence between their residual limb and the target object appears more gradual and is characterized by a much larger variability.

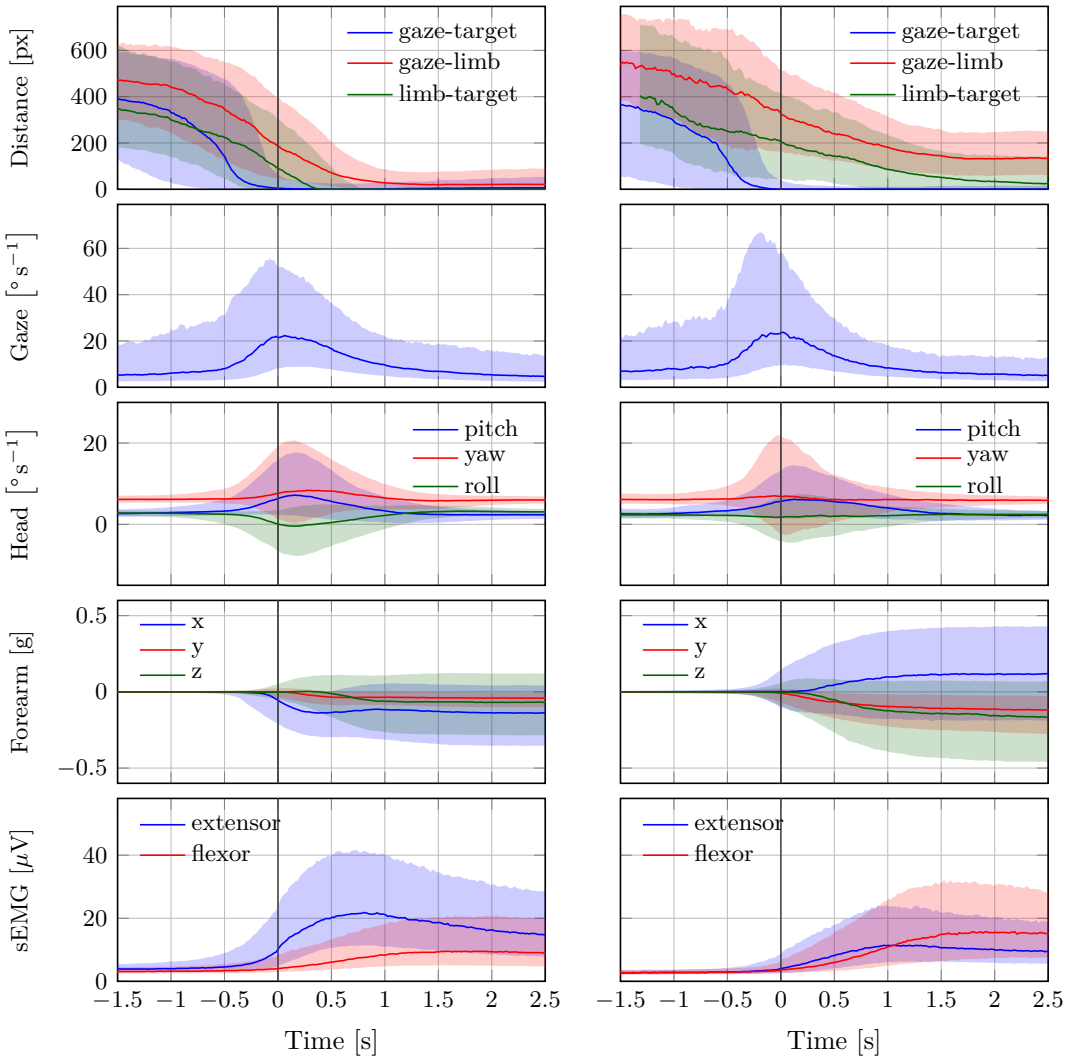
A noteworthy observation is that the activation of the eyes always preceded the end of the vocal stimulus. The reason is that subjects could typically deduce the target object already before the end of the instruction, as we can observe in the vocal commands of Table 4.4. This does however not affect our results, since we are interested in the relative delay between eyes, head, and forearm rather than reaction times to the stimulus. The differences in reaction time to the vocal instructions do increase however the dispersion of the distributions. We also note that the relative contribution among the three axes of the acceleration profile differs between able-bodied and amputated subjects. The reason is that we normalized this profile with respect to the initial position of the forearm, which is typically different for both types of subjects. In the present study, we use accelerometry to determine *when* the arm starts to move and rely on the *limb-target* distance to measure its convergence to the target object.

6.2.3 Manipulation Phase

In this section we focus on the behavior of intact and amputated subjects during the functional tasks to further investigate the similarities in gaze strategy. These figures start from 2 s before the vocal instruction and cover the entire manipulation action. For this analysis we group the MeganePro activities in three categories based on the type of task and the associated visual behavior, as shown in Table 6.2. These categories are *in place* manipulation actions, *lifting* actions, and finally *displacement* actions.

6.2.3.1 In Place Actions

The *in place* actions concern manipulation tasks that do not require moving the object, like opening an object, cutting bread, or pressing a button of the remote control. The aggregated profiles of all modalities for these actions are shown in Figure 6.3a for able-bodied subjects and in Figure 6.3b for amputees. During this type of action, the gaze remains fixed on the target object throughout the entire duration of the manipulation, as can also be seen in the example in Figure 6.4 that overlays gaze and object segmentations on representative frames of the first person video. As expected, the hand remains on the target for the entire duration in case of able-bodied subjects, whereas for amputees there remains a constant subject-dependent distance between the residual limb and the target. Head movements are limited to the initial reach-to-grasp phase to center the object in the field of view, after which the head remains fixed until the end of the manipulation.

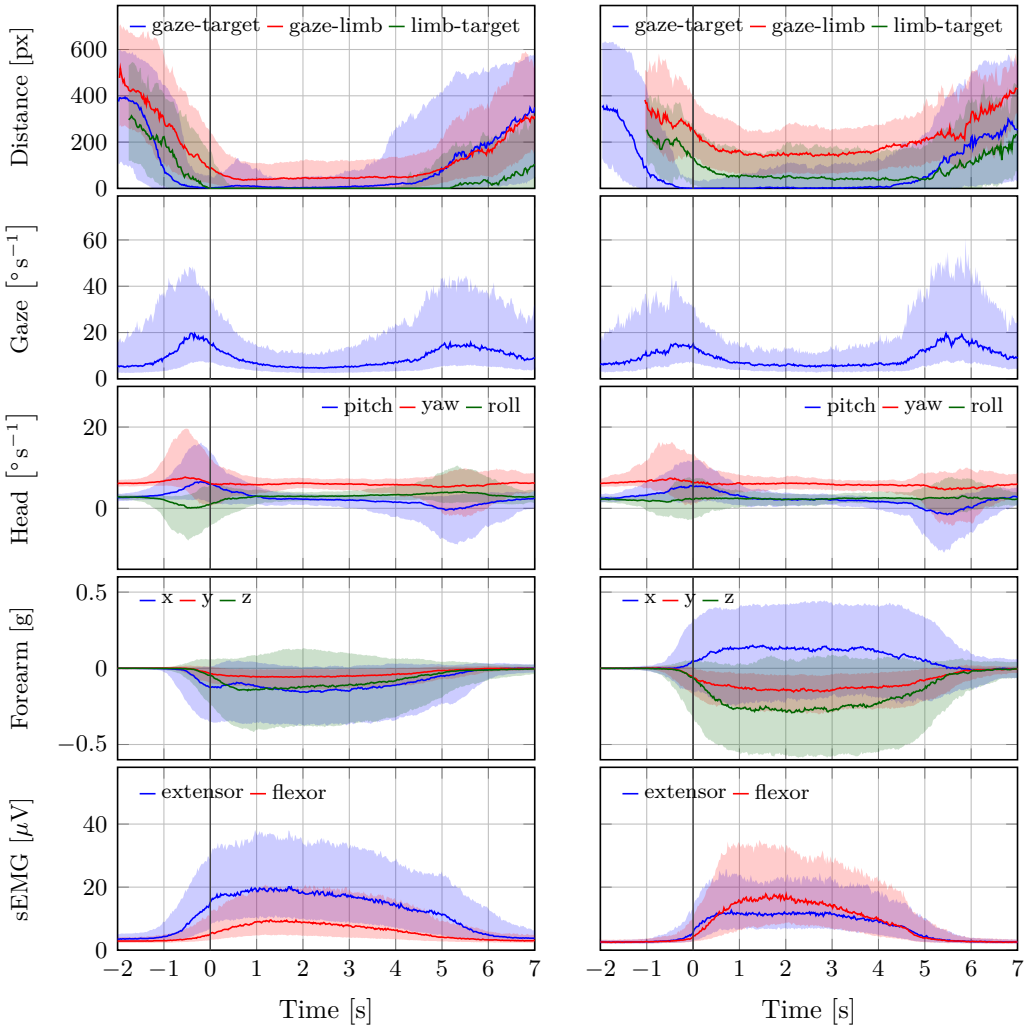


(a) Intact Subjects. **(b)** Amputated Subjects.

Figure 6.2. The trend of each modality in the reach-to-grasp phase for (a) intact and (b) amputated subjects. The zero corresponds to the end of the vocal instruction indicating the required manipulation. The solid line represents the median over all trials from all subjects, whereas the shaded areas indicate the 25$^{\text{th}}$ and 75$^{\text{th}}$ percentiles. Segments with more than 90 % missing data were omitted.

Table 6.2. The functional activities performed in the dynamic part of the MeganePro protocol are grouped in three categories: *in place*, *lifting*, and *displacement* actions. In the first group we consider manipulation activities that do not require moving the object. The second group concerns those tasks in which the subject was required to lift an object up and then place it back in its initial position. In the last case the subject is asked to horizontally move an object between two positions on the table.

Category	Instruction
In place	open and close the door handle
	turn the key in the lock
	open and close the pencil case
	open and close the cap of the bottle
	open and close the drawer
	open and close the lid of jar
	screw and unscrew the light bulb
	squeeze the clothespin
	press a button on the remote control
	cut bread with the knife
	turn the screwdriver
Lifting	drink from the can
	lift the plate
	lift the book
Displacement	move the ball to the right and back
	move the keys forwards and backwards
	move the keys forwards and backwards
	move the wrench to the right and back
	move the knife forwards and backwards
	move the fork to the right and back



(a) *In place* tasks of intact subjects. **(b)** *In place* tasks of amputated subjects.

Figure 6.3. The trend of each modality in the *in place* functional tasks for (a) intact and (b) amputated subjects. The zero corresponds to the end of the vocal instruction indicating the required manipulation. The solid line represents the median over all trials from all subjects, whereas the shaded areas indicate the 25$^{\text{th}}$ and 75$^{\text{th}}$ percentiles. Segments with more than 90 % missing data were omitted.

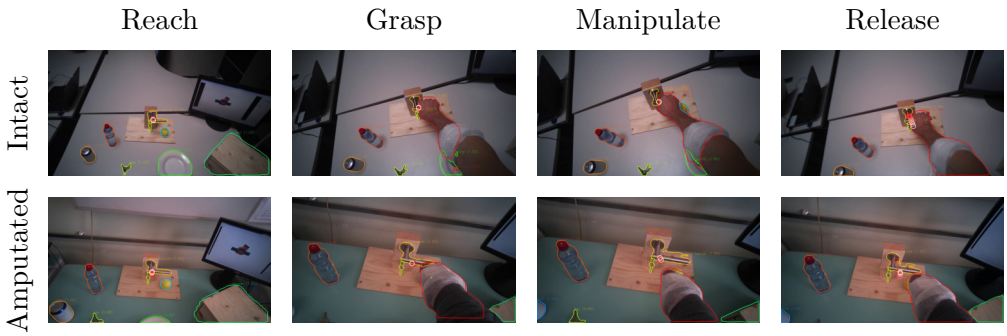


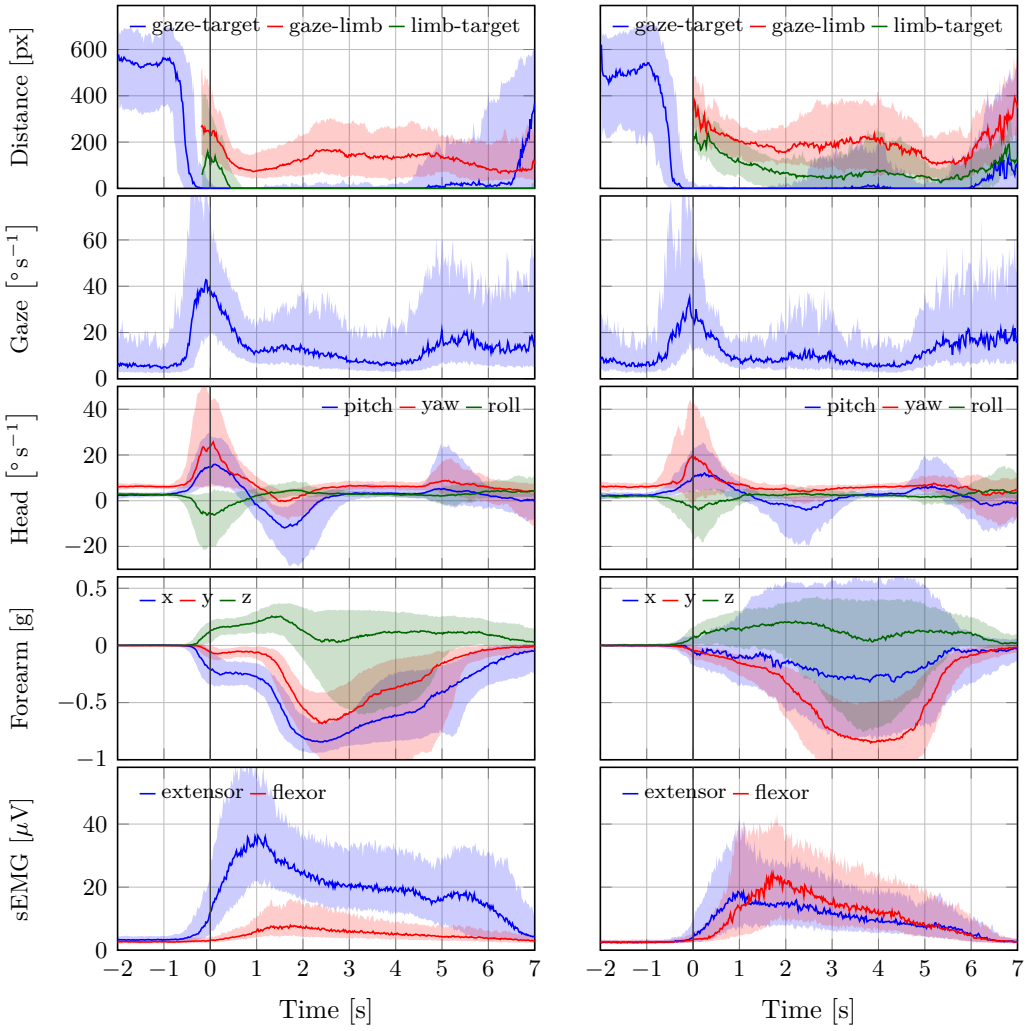
Figure 6.4. Example of the visuomotor behavior of an intact (first row) and amputated (second row) participant while opening a door handle. The gaze trail is represented by the circles from the current gaze position (red) to ten samples later (white). Both subject groups direct the gaze on the object during the reaching phase (first column). The eyes then remain focused on the target object during the grasping and manipulation phases (second and third columns). In both cases, the motor behavior of the arm is similar for intact and amputated subjects. During the release phase the gaze shifts away from the object (fourth column).

6.2.3.2 Lifting Actions

The second group is composed of *lifting* actions, in which the subject was required to lift an object up and then place it back in its initial position. As can be seen in Figure 6.5a and Figure 6.5b, also in this case the gaze anticipates head and forearm movements. More interestingly, we see a clear movement in the pitch orientation of the head. Since these actions are executed while standing, the subjects first lower their head to locate the target object on the table. Then, when they have located and grasped the object, they raise their head again with a peak pitch velocity at approximately 1.7 s for able-bodied subjects and slightly later for amputated subjects. This head movement coincides with a modestly increased gaze velocity and is due to the tracking motion of the lifting action. In some cases, this tracking strategy even caused an amputated subject's *gaze-target* distance to increase, as can also be seen in the example in Figure 6.6. Finally, the subjects lower their head again when tracking the release of the object at the end of the trial.

6.2.3.3 Displacement Actions

The final category are the so-called *displacement* actions. During these tasks, the subjects had to grasp the objects, move them horizontally to another position, and then place them back in the initial position. We note that the gaze and motor behavior starts earlier with respect to the vocal instruction. For this category of tasks, the name of the object happens to appear at the beginning of the instruction (see Table 4.4), thus allowing subjects to initiate the task early. For intact subjects, we see in Figure 6.7a that, approximately 200 ms before the hand reaches the object, the *gaze-target* distance starts to increase again. The gaze, in this case, shifts already to the destination position for the displacement action, as demonstrated in the second panel in Figure 6.8. Although less pronounced, the same pattern repeats itself at around 1.5 s when the subject initiates the return movement. The profiles



(a) *Lifting* tasks of intact subjects. **(b)** *Lifting* tasks of amputated subjects.

Figure 6.5. The trend of each modality in the *lifting* functional tasks for (a) intact and (b) amputated subjects. The zero corresponds to the end of the vocal instruction indicating the required manipulation. The solid line represents the median over all trials from all subjects, whereas the shaded areas indicate the 25$^{\text{th}}$ and 75$^{\text{th}}$ percentiles. Segments with more than 90 % missing data were omitted.

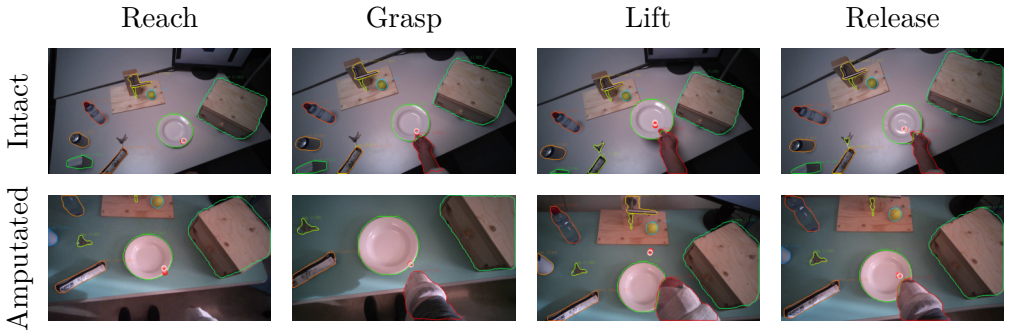


Figure 6.6. Example of the visuomotor behavior of an intact (first row) and amputated (second row) participant lifting a plate. The gaze trail is represented by the circles from the current gaze position (red) to ten samples later (white). The eyes focus on the manipulation point to plan the hand's approach (first and second columns). During the lifting phase, the eyes move away from the reaching point and the amputee's gaze even exceeds the mask boundary of the plate (third column). The object is fixated again during the release (fourth column).

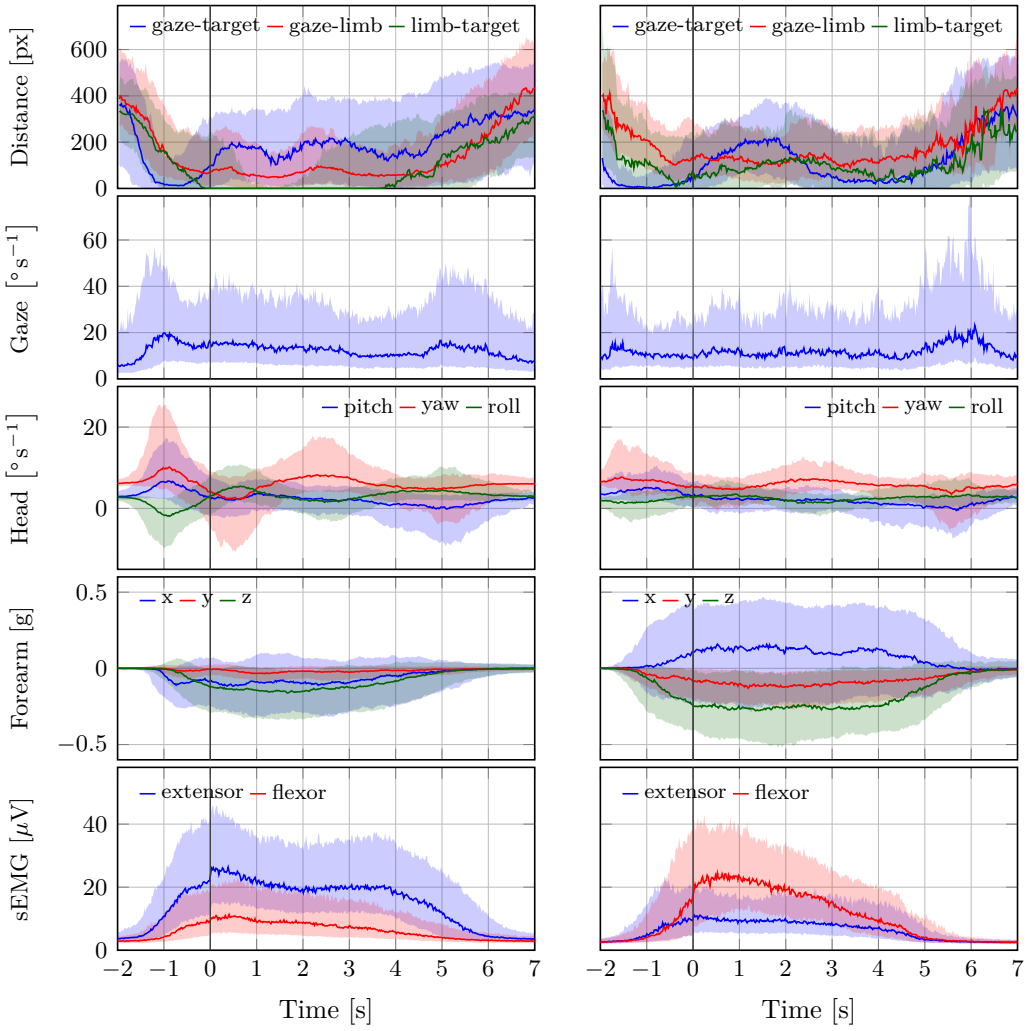
for the amputated subjects in Figure 6.7b show different behavior, with an overall increase in *gaze-target* distance throughout the entire duration of the movement. As intact subjects did, their gaze anticipates the path of the hand rather than the path of the object, which is not physically displaced. This strategy is demonstrated clearly in the bottom row of Figure 6.8.

6.3 Discussions

The aim of these analyses was to determine the window of opportunity for exploiting gaze as contextual information in decoding the manipulation intent of amputees. A related question is to which extent natural gaze behavior of amputees and able-bodied subjects are similar. After comparing our results with related work, we discuss both these topics. Finally, we argue for the use of recent developments in deep learning in the analysis of large-scale visuomotor studies.

6.3.1 Comparison with Related Work

In Section 6.2.2, we presented the results of eye, head, and limb coordination during reaching and grasping. The eyes are the first to react to the vocal stimulus by exhibiting an increasing saccade-related activity, leading to a fixation on the target in about 150 ms. When the eyes start moving, also the head follows almost immediately. Such short delays between movement of the eyes and the head have been reported in the literature, ranging from 10 ms to 100 ms during a block-copying task (Smeets et al. 1996) or in reaction to visual stimuli (Di Cesare et al. 2013; Goldring et al. 1996). This behavior is however strongly dependent on the experimental setting and even small variations therein can change the outcome. For instance, Pelz et al. (2001) found that depending on the exercise's instruction the head may both precede (by about 200 ms) or follow the eyes (by about 50 ms) in the same block-copying task.



(a) *Displacement* tasks of intact subjects.

(b) *Displacement* tasks of amputated subjects.

Figure 6.7. The trend of each modality in the *displacement* functional tasks for (a) intact and (b) amputated subjects. The zero corresponds to the end of the vocal instruction indicating the required manipulation. The solid line represents the median over all trials from all subjects, whereas the shaded areas indicate the 25$^{\text{th}}$ and 75$^{\text{th}}$ percentiles. Segments with more than 90 % missing data were omitted.

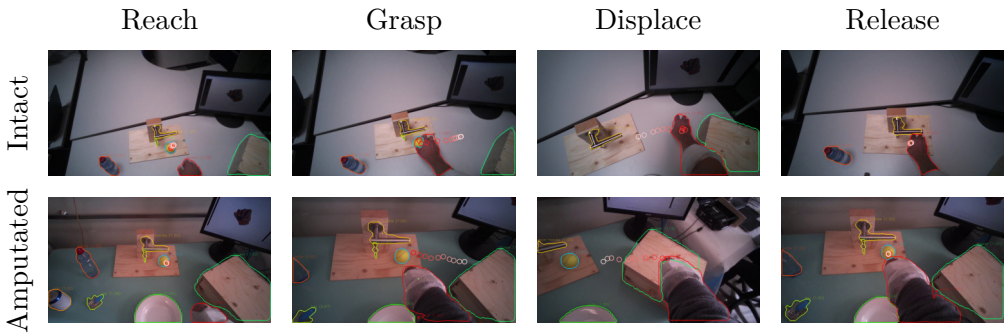


Figure 6.8. Example of the visuomotor behavior of an intact (first row) and amputated (second row) participant while moving a ball. The gaze trail is represented by the circles from the current gaze position (red) to ten samples later (white). The gaze focuses on the object until the hand's arrival (first column), when the grasping phase begins the eyes shift away toward the destination (second column). When the hand reaches the destination the gaze shifts back to the initial location (third column) to release the target (fourth column).

After the activation of the eyes and the head we observe the movement onset of the arm approximately 130 ms later. Similar values ranging from 170 ms to 300 ms were also reported by Smeets et al. (1996) and Pelz et al. (2001) in a block-copying task and by Belardinelli et al. (2016) in a pick and place task. Land et al. (1999) instead found a median delay of 0.56 s during a tea-making activity. Rather than movement onset, the time the hand takes to reach the target is more interesting for our scope. For the intact subjects, the hand typically starts to occlude the target object around 500 ms after the first fixation. Although occlusion does not necessarily already imply a completed grasp, especially given the first person perspective, we do expect the grasp to follow not much later. These results confirm that visual attention on objects anticipates manipulation. In previous studies concerning displacements (Johansson et al. 2001; Belardinelli et al. 2016; Lavoie et al. 2018) and grasping activities (Brouwer et al. 2009), a variable delay ranging from 0.53 s to 1.3 s was found between the eye and hand. Also in these cases, the exact value of the delay depends on the characteristics of the experiment.

In Section 6.2.3 we concentrated on the visuomotor strategy adopted by amputated and able-bodied subjects to interact with the objects during three groups of functional tasks. We can characterize the strategies associated with these groups in terms of the types of fixations defined by Land et al. (1999) and Land and Hayhoe (2001), namely *locating*, *directing*, *guiding*, and *checking*. A fixation to *locate* is typically done at the beginning of an action, to mentally map the location of objects that will be used. Instead, a fixation to *direct* is meant to detect an object that will be used immediately after. Fixations to *guide* are usually multiple and occur when the gaze shifts among two or more objects that are approaching each other. Finally, there are long *checking* fixations to monitor the state of an action waiting for its completion.

The visual strategy of the *in place* actions is relatively straightforward. In these tasks, subjects initiate with a fixation to *direct* the attention to the target object. Subsequently, their fixation remains on the manipulated object to *check* the correct

execution of the task. Note that this visual attention seems focused on the target object rather than the subject's hand, as can be seen comparing the *gaze-target* and *gaze-limb* distances in Figure 6.3a and Figure 6.3b. Indeed, Land et al. (1999) noted that the hands themselves are rarely fixated.

Also the *lifting* actions start with a *directing* fixation to locate the object of interest. However, whereas the initial fixation is focused on the intended grasp location (cf. the left column in Figure 6.6), the gaze shifts upwards when the hand has grasped the object. This coincides with the transition from the *directing* fixation to visually *checking* the lifting action. This is in line with observations by Voudouris et al. (2018), who noted that people may fixate higher when grasping and lifting an object to direct their gaze to where the object will be in the future.

Finally, *displacement* actions are the ones most investigated in the literature. Previous studies on pick and place tasks (Belardinelli et al. 2016; Lavoie et al. 2018) and on the block-copying task (Smeets et al. 1996; Pelz et al. 2001) fall in this category. In this case, we observe in Figure 6.7a that the *gaze-target* and *gaze-limb* distances have three minima for intact subjects, namely at the initial pick-up, the destination, and at the release again at the initial position. All three minima indicate fixations that are meant to *direct* the approach of the hand, either for (1) grasping the object, (2) displacing it, or finally (3) releasing it. This behavior can clearly be seen for both intact as well as amputated subjects in the example in Figure 6.8. We also notice that the eyes did not wait for the completion of the pick-up action, moving instead toward the position of the destination around 200 ms in advance. This proactive role of the eyes was highlighted by Land et al. (1999), who measured the gaze moving on to the next object between 0 s to 1 s before the current object manipulation was terminated. Also Pelz et al. (2001) observed the eyes departing from the target object 100 ms to 150 ms before the arrival of the hand.

6.3.2 Comparison between Intact and Amputated Subjects

One of the aims of this work was to understand if a transradial amputation has introduced important changes in the visuomotor behavior of amputees. During the reach-to-grasp phase, the overall behavior of intact and amputated subjects is comparable. Even if the coordination timeline between eyes, head, and limb are similar, there are some minor discrepancies between the two groups. The main observed difference concerns the delayed activation of the forearm muscles during the reaching phase for amputated subjects, which was found to be statistically significant. Similarly, during the lifting tasks we noted slower pitch movements of the head. It is likely that some subjects interpreted the instruction to perform the grasp with their missing limb by activating their phantom limb. Such movements executed with the phantom limb are known to be slower than those executed with the intact hand (Raffin et al. 2012a; Graaf et al. 2016).

Throughout the manipulation phase, we observe a striking similarity in visuomotor strategy between the amputated subjects and the control group. The differences that we noted in the results are not due to an alternative gaze strategy, but rather the impossibility to physically interact with the objects. For instance, in the *lifting* task visualized in Figure 6.5b we saw an increase in *gaze-target* distance in the range from 2 s to 5 s. This increase was due to an upward shift in the gaze location to

track the execution of the lifting action. Similarly, during the displacement task in Figure 6.7b we do not observe a minimum in *gaze-target* distance at around 1.5 s when arriving at the intermediate destination, as was the case for intact subjects (see Figure 6.5a). Instead, around the same time we observe a peak for the amputated subjects, solely because the target object is still at its original position. The examples for these gaze strategies in Figure 6.5 and Figure 6.7 demonstrate how similar intact and amputated subjects behaved.

6.3.3 Integration of Vision in Prostheses to Improve Intent Recognition

The estimated time interval from *fixation* to *grasp* in Section 6.2.1 shows that the window of opportunity is approximately 500 ms for intact subjects. This interval cannot be accurately determined for amputated subjects, as they executed the movement with their missing limb and therefore lacked physical contact with the target object. Although Figure 6.2b suggests that this window will at least be as long for amputated users, one may argue that this result is not representative for movements performed with a prosthesis. However, previous studies showed without exception that prosthetic users still fixate the target object for the majority of the reaching phase (Bouwsema et al. 2012; Sobuh et al. 2014; Chadwell et al. 2016; Hebert et al. 2019; Parr et al. 2019), albeit alternating it more often with fixations on the hand (i.e., the "switching" strategy). Moreover, this reaching phase may actually take more than twice as long as compared to the same movement performed with the anatomical limb (Sobuh et al. 2014; Hebert et al. 2019). These findings suggest that the target object will still be fixated proactively by a prosthetic user and that the window of opportunity will more likely be longer than shorter.

Exploiting this anticipatory gaze behavior is appealing because it comes naturally and therefore does not require specific attention from the user. The success of this approach relies however on the ability to distinguish informative fixations from those that are not necessarily related to the manipulation intent. In a preliminary study we attempted to address this problem by including the onset of the arm movement as an additional condition, which we have shown here to shorten the window of opportunity (Gigli et al. 2018). In the next chapter we describe instead a recent approach that we have developed to maximize the inclusion of gaze related information. Thanks to the frame-by-frame segmentations, we could accurately and instantaneously recognize object fixations by measuring the distance between the object's segmentation mask and the gaze point. In contrast, common fixation classifiers, such as IVT (Salvucci and Goldberg 2000), define a fixation simply as the lack of eye movement. In reality, gaze shifts more commonly involve not only eye movement, but also head and sometimes even trunk movements (Morasso et al. 1973; Land 2006). When the head moves, the optokinetic and vestibulo-ocular reflexes cause the eyes to counteract the head movement to maintain a stable gaze point (Lappe and Hoffmann 2000). It is exactly due to such coordinated gaze movements that the initial object fixation in Figure 6.2 actually coincides with a *peak* in gaze velocity. The need to detect fixations as early as possible therefore implies a detection method that uses more information than eye movement alone. Whether this is best done by compensating for head movements (Kinsman et al.

2012; Larsson et al. 2014) or by comparing the visual object at the gaze point as in the present study is an open question.

6.3.4 Advantages of Automatic Analysis

Without the deep learning approach described in the previous chapter it would have been extremely labor intensive to analyze approximately 70 h of video and data from 44 subjects. Manufacturers of eye tracking devices often provide applications for semi-automatic analyses, but these do not allow the level of automation nor precision as the procedure described here. Although the object segmentations produced by Mask R-CNN were occasionally mistaken, the segmentations seen in the examples of Figure 6.4, Figure 6.6, and Figure 6.8 are representative for the overall performance. It may easily be overlooked that data from research studies, such as the present, often contain much less variability than the datasets on which these algorithms are trained and evaluated. With minimal finetuning efforts, it is therefore likely to obtain levels of performance that considerably exceed those reported in the literature, as was seen in Table 5.1.

Our current approach ignores the temporal relationship between consecutive frames. Considering that the input data come from the recorded videos of the Tobii glasses, it is likely that the identity and location of the segmented objects will be very similar from one frame to the next. Mask R-CNN, like CNNs in general, was not designed to accept a sequence of data as input. The model could be extended to allow such inputs by employing a so-called Recurrent Neural Networks (RNNs) on top, which is a type of network proposed in 80's specifically for modeling time series (Rumelhart et al. 1986). The "recurrent" in the name refers to its characteristic that the output at one time step becomes part of the input to the network at the next step, thus giving it a sense of memory. In the context of video sequences, this would allow to improve the predictions of a new frame by also considering the information and predictions for the previous frames.

Chapter 7

Proof of Concept

The study on visuomotor coordination presented in the previous chapter shows that a fixation on the target object typically precedes the subsequent grasp by at least 500 ms. Moreover, depending on the functional activity to be performed, both intact and amputated subjects remain focused on the target also during the whole manipulation or, at least, part of it. Since vision actively participates in such tasks it seems possible to merge visual and forearm information to aid movement recognition.

In a preliminary study, we proposed a method for modality integration based on multiple steps (Gigli et al. 2018). In the first step, a stable and relevant fixation was detected by combining a low amount of eye activity with an increase in muscular activity; this allowed to select fixations that precede a grasp. The object on which the gaze was directed at that point in time was segmented employing the active segmentation method proposed by Mishra et al. (2012). High-level features were then extracted from the cropped image patch using a CNN and integrated, at the kernel level, in a grasp classifier together with sEMG features. Under the assumption that the object of interest remains the same during the whole grasp, the CNN features associated to a fixation were propagated until the next fixation. This method was evaluated on five intact subjects using the preliminary data collection described in Section 4.2.1. Despite an improvement in classification accuracy when integrating visual information, there were several shortcomings that should be addressed. First, the method to detect stable fixations evaluates gaze-velocity related quantities. As argued in Section 6.3.3, such approaches delay the recognition of the fixation due to the presence of head movements. Second, the algorithm used for object segmentation requires the gaze to fall inside the target object, which may not always happen due to calibration problems with the Tobii glasses (see Section 4.4.1.1). Finally, the propagation of the same visual information until the next fixation is unrealistic and increased the classification errors in the rest period, during which gaze and vision do not yield any useful information.

In this chapter we present a proof of concept that extends the previous approach and addresses these problems. Rather than detecting a stable fixation and then segmenting whichever object is at the gaze point, it instead uses the approach described in Chapter 5 to segment all the known objects in a frame. The object of interest is then simply defined as the one *nearest* to the gaze position, regardless

of any information on arm movement. Moreover, we add an additional term that automatically regulates the contribution of the visual information based on the distance between the gaze point and the object. Also in this case the sEMG and visual modalities are merged inside a classifier via a kernel combination. This method is explained in Section 7.1 and tested on the subjects of the MeganePro dataset in Section 7.2.

7.1 Multimodal Integration

In the following we first present the strategy that was used to combine the sEMG- and gaze-related information and we conclude with the details of the classifier used for movement recognition.

7.1.1 Kernel Combination

To process the sEMG data we followed the approach by Englehart and Hudgins (2003), already used in Section 4.4.2.2, by computing features in a sliding window of 400 samples (i.e., 208 ms) with an increment of 20 samples (i.e., 10 ms). We define as $\boldsymbol{x}_e^i$ the mDWT features extracted from the i^{th} window. However for simplicity in the following calculation we remove the index of the window indicating the vector as only $\boldsymbol{x}_e$. The relevant visual information is obtained by combining the segmentation and classification results of Mask R-CNN with the gaze position. For each frame, among all the detected objects we only consider the object with the minimum distance d from the gaze coordinates. The class of the nearest object is then encoded in a feature vector $\boldsymbol{x}_{id}$ using the one-hot representation. Each vector has eighteen dimensions as the number of classes (i.e., eighteen MeganePro objects) and all the elements are 0 except for the position corresponding to the object's class, which is 1. We excluded the class "person" from these analyses since during manipulation we want avoid to recognize the hand as nearest object rather than the target object itself.

After the features have been extracted both for sEMG and visual cues, this information was used to train a grasp classifier. Among the possible techniques, we opted for a mid-level multimodal integration (Tommasi et al. 2008). In this case the fusion happens at the kernel level by means of a weighted sum of cue-specific kernels:

$$k(\boldsymbol{x}, \boldsymbol{y}) = w_e \cdot k_e(\boldsymbol{x}_e, \boldsymbol{y}_e) + w_g \cdot k_{id}(\boldsymbol{x}_{id}, \boldsymbol{y}_{id}) \ , \tag{7.1}$$

where $\boldsymbol{x} = (\boldsymbol{x}_e, \boldsymbol{x}_{id})$ and $\boldsymbol{y} = (\boldsymbol{y}_e, \boldsymbol{y}_{id})$ are two couples of sEMG and gaze feature vectors, and w_e and w_g are the weights of the sEMG and gaze contributions. Based on the results presented in Section 4.4.2.2, we use the exponential χ^2 kernel to express the similarity between the sEMG features in $\boldsymbol{x}_e$ and $\boldsymbol{y}_e$ (see Section 2.2.2). For the visual information we use instead a linear kernel, noting that the evaluation between two $\boldsymbol{x}_{id}$ vectors is 1 if they indicate the same object and 0 otherwise. Since the output of first kernel ranges in $(0, 1]$ and the output of the linear kernel belongs to the set $\{0, 1\}$, we can ensure that also their combination is in the same range. For simplicity we can set $w_e = 1$ and leave as free hyperparameter to tune only w_g, which describes the importance of the visual cue with respect to the sEMG cue. To

regulate the gaze contribution we add to the visual cue another term that takes into account the distance of the gaze from the nearest object. Indeed the closer the gaze is to the object, the more importance the information has in the kernel evaluation. The farther the gaze is from the object, the less the visual information should be regarded. To consider this variable contribution, we add a linear kernel,

$$k_d(x_d, y_d) \ . \tag{7.2}$$

This kernel uses a feature x_d that decreases exponentially with the distance d between the gaze and the nearest object:

$$x_d = e^{-g \cdot \max\{0, d-\epsilon\}}, \quad \text{with} \quad g = 0.01, \quad \epsilon = 20\,\text{px}. \tag{7.3}$$

This function is maximum, $x_d = 1$, when the gaze is at most 20 px from the nearest object, which coincides with the threshold used to indicate the beginning of a fixation in Chapter 6. As the *gaze-object* distance increases, the function drops to 0 with a decay constant g. This property bounds the kernel in the $(0, 1]$ interval, therefore making its range comparable with that of the previous kernels without the need of any scaling factors. Adding the modifications, Equation 7.1 becomes:

$$k(\boldsymbol{x}, \boldsymbol{y}) = k_{\chi^2}(\boldsymbol{x}_e, \boldsymbol{y}_e) + w_g \cdot k_{lin}(x_d, y_d) \cdot k_{lin}(\boldsymbol{x}_{id}, \boldsymbol{y}_{id}) \ , \tag{7.4}$$

where we indicate with k_{χ^2} and k_{lin} the exponential χ^2 and linear kernels described in Section 2.2.2.

7.1.2 Classifier

The kernel in Equation 7.4 was used with the KRLS classifier, which was used previously in Section 4.4.2.2 and introduced in Section 2.2.2. For the proof of concept, the evaluation was repeated over the four possible splits of the train and test data, where in turn one of the four grasp repetitions was used in the test phase and the remaining three were employed to train the model. The final prediction accuracy is given by the average over the four splits. The same k-fold process, but with just the three training repetitions, was also adopted for the inner cross validation to optimize the hyperparameters. This optimization involves the regularization parameter λ of the classifier, the exponential χ^2 kernel parameter γ_{χ^2} , and the weight w_g used in the kernel combination. The value of each parameter was selected from a dense grid search: $\lambda \in \{2^{-16}, 2^{-15}, \ldots, 2^2, 2^3\}$, $\gamma_{\chi^2} \in \{2^{-20}, 2^{-19}, \ldots, 2^0, 2^1\}$, and $w_g \in \{0.01, 0.1, 1, 10\}$. Given the results from preliminary analyses, the decay constant g was instead kept constant at 0.01. For computational reasons, the training data were downsampled with a factor 10, while the data used for hyperparameter optimization were downsampled with an additional factor 4. The classification accuracy used later for the analyses is defined as the proportion of windows in the test data that was classified correctly by the classifier.

7.2 Grasp Recognition with Visual Information

The aim of the proof of concept is to experimentally verify whether the grasp recognition could benefit from the integration of visual information. In this section,

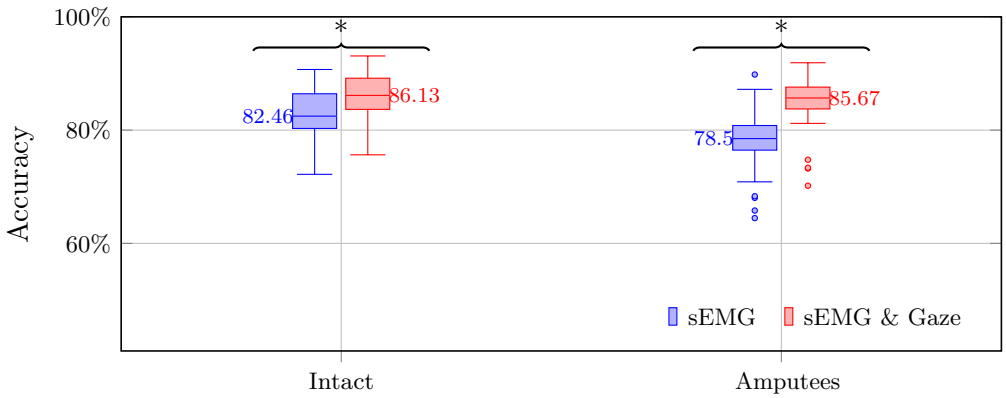


Figure 7.1. Classification accuracies for intact (left) and amputated (right) subjects when predicting the grasp type using only sEMG and while integrating also the visual information. For both types of subjects, the classification accuracy increases when integrating the visual modality.

we first present the classification performance obtained with the inclusion of gaze and then analyze these results more deeply.

7.2.1 Classification Performance

Figure 7.1 shows the classification accuracy for both intact and amputated subjects. We compare for each group the results obtained with the standard classifier that just utilizes sEMG and the proposed method that integrates gaze and visual information. The integration of vision with muscular information increases the performance by about 4 % and 7 % for intact subjects and amputees, respectively. When using only sEMG data, the performance difference between the two groups of subjects is approximately four percentage points. This is usually attributed to a deterioration of the myoelectric signal quality due to the amputation. However, the inclusion of gaze and visual information, which we have shown to be practically unaffected by amputation, lowers this difference to less than one percentage point. Moreover, as reported in Table 7.1, this improvement is consistent for all subjects and statistically significant (sign test, $p = 5.684 \times 10^{-14}$). Even the subject affected by strabismus (114) shows a small gain, meaning that the performance increases when the gaze is tracked. These results confirm the hypothesis that the visual cue holds important information that can help to improve grasp recognition without asking the subjects to alter their natural behavior during grasp and manipulation activities.

7.2.2 Analysis of Improvements

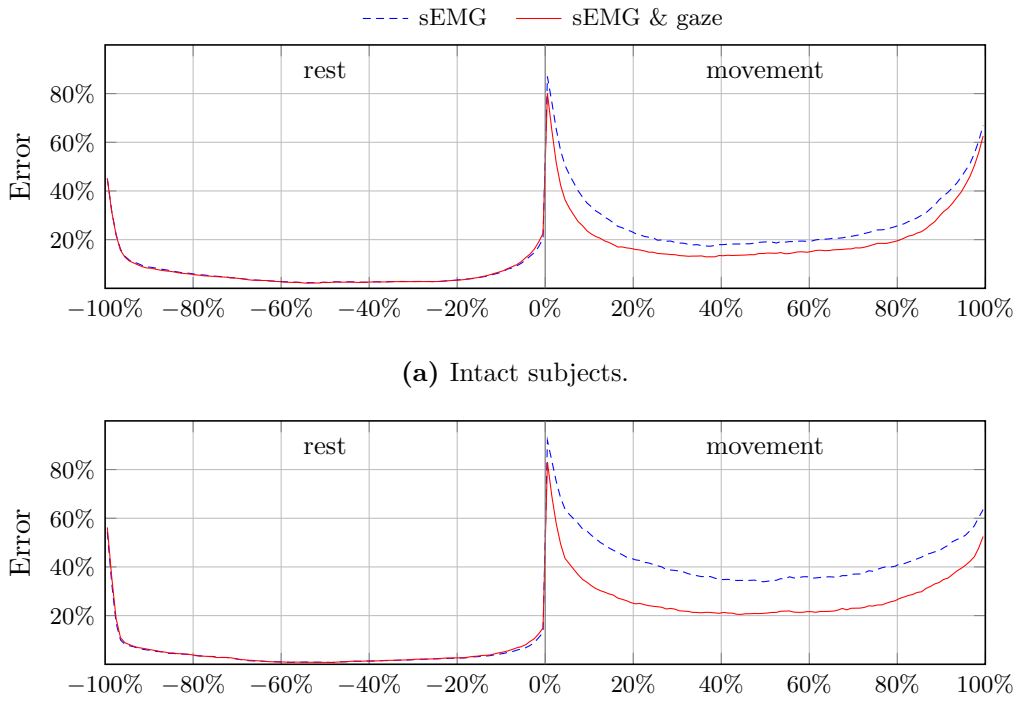
Besides showing an increase in classification performance, we are also interested to understand *where* this improvement has been gained. In Figure 7.2 we report the trend of the prediction error during the rest and grasp periods averaged over all subjects and trials. Since each movement has a (slightly) different duration, we normalized both the rest phase (−100 % to 0 %) and subsequent grasp (0 % to 100 %) to make trials comparable. During the grasp the inclusion of visual information

Table 7.1. Classification accuracy per subject. The table reports the ID of the subjects in the dataset (first column), and the classification accuracy in each of the four train-test splits (S) using only sEMG (second-fifth columns) and the combination of sEMG and gaze (sixth-ninth columns). For each split the highest performance is highlighted in bold.

		sEMG [%]				sEMG & gaze [%]			
	ID	S1	S2	S3	S4	S1	S2	S3	S4
Intact	10	82.38	84.86	83.26	82.51	**86.36**	**87.92**	**87.09**	**86.19**
	11	73.79	73.83	74.77	72.89	**76.85**	**75.63**	**77.62**	**75.76**
	12	82.12	86.51	86.28	85.43	**85.79**	**89.95**	**89.10**	**89.33**
	13	85.92	86.51	86.71	85.24	**88.37**	**89.33**	**90.03**	**87.91**
	14	80.75	82.23	81.45	79.49	**83.24**	**84.42**	**84.38**	**82.40**
	15	80.77	81.91	82.41	80.87	**81.61**	**82.26**	**84.00**	**82.54**
	16	79.29	81.00	80.54	81.14	**84.37**	**85.82**	**86.07**	**86.67**
	17	86.13	88.05	87.41	88.85	**88.74**	**90.06**	**89.55**	**90.10**
	18	77.75	80.99	80.99	79.20	**81.20**	**84.85**	**85.45**	**82.44**
	19	79.72	80.74	78.64	75.55	**84.09**	**83.91**	**82.62**	**80.85**
	20	90.38	90.62	89.90	89.80	**91.88**	**92.04**	**91.13**	**91.72**
	21	78.77	79.52	79.43	78.19	**82.98**	**82.39**	**82.38**	**81.83**
	22	83.08	84.79	86.20	85.63	**85.63**	**86.82**	**87.74**	**87.41**
	23	85.10	87.19	86.83	84.82	**86.96**	**88.95**	**88.72**	**86.73**
	24	83.39	81.80	81.72	80.18	**89.65**	**88.54**	**87.66**	**87.17**
	26	86.50	88.27	88.20	86.45	**89.44**	**90.63**	**90.87**	**88.92**
	27	82.00	82.66	83.57	81.99	**85.99**	**85.24**	**85.68**	**83.88**
	28	82.95	85.16	85.39	85.29	**85.35**	**87.22**	**87.92**	**88.33**
	29	88.99	90.70	90.38	90.05	**91.06**	**92.43**	**91.61**	**92.11**
	30	88.34	88.21	90.00	89.03	**90.77**	**90.11**	**91.98**	**91.11**
	31	84.53	87.02	86.50	83.53	**87.74**	**90.35**	**89.94**	**87.57**
	32	80.44	81.24	80.71	79.65	**83.79**	**85.04**	**84.67**	**83.71**
	33	74.04	76.70	76.51	75.11	**77.67**	**80.14**	**79.13**	**78.91**
	34	83.03	85.56	84.49	83.10	**85.83**	**87.29**	**87.39**	**85.70**
	35	77.77	80.33	81.53	79.84	**83.34**	**84.81**	**85.67**	**83.70**
	36	82.50	86.40	84.93	81.48	**87.16**	**90.57**	**88.62**	**85.89**
	37	86.69	90.02	89.29	87.69	**90.41**	**93.09**	**91.94**	**91.53**
	38	79.11	81.93	80.85	79.05	**80.75**	**83.55**	**83.00**	**81.23**
	39	74.24	73.76	72.19	72.62	**80.69**	**78.80**	**77.89**	**78.98**
	40	79.40	82.43	81.42	80.78	**84.94**	**87.47**	**85.59**	**85.67**
Amputees	101	75.36	77.95	78.62	76.52	**84.30**	**86.12**	**86.01**	**85.24**
	102	70.86	73.54	73.91	71.76	**82.99**	**85.73**	**85.76**	**83.89**
	103	77.29	78.02	79.35	77.28	**87.50**	**89.12**	**89.40**	**87.90**
	104	76.51	80.65	80.34	78.14	**85.31**	**86.91**	**86.80**	**85.01**
	105	75.74	79.10	76.31	75.86	**82.25**	**83.73**	**82.03**	**81.71**
	106	78.38	81.65	78.01	76.10	**84.09**	**86.12**	**84.11**	**82.60**
	107	77.63	78.33	79.38	76.52	**82.36**	**83.06**	**84.54**	**81.18**
	108	76.97	79.13	80.34	77.16	**85.95**	**87.73**	**88.98**	**85.67**
	109	78.61	79.06	80.63	79.47	**83.82**	**84.14**	**85.76**	**84.57**
	110	81.30	82.05	81.93	77.65	**88.01**	**88.06**	**87.81**	**85.68**
	111	76.24	81.88	82.48	79.06	**83.87**	**88.71**	**87.76**	**86.13**
	112	82.92	86.51	87.20	82.23	**89.93**	**91.89**	**91.79**	**87.55**
	113	76.24	82.14	79.07	78.75	**81.52**	**87.83**	**84.29**	**83.73**
	114	82.87	87.03	89.81	86.80	**83.45**	**87.49**	**90.10**	**87.10**
	115	68.32	68.04	65.79	64.46	**74.76**	**73.35**	**73.22**	**70.17**

contributes to lower the error consistently throughout the entire duration. This means that vision both compensates for the level of noise in the myoelectric signals during movement transitions (i.e., 0 % to 20 % and 80 % to 100 % intervals of Figure 7.2) and, if the target object is fixated, helps to stabilize the grasp during manipulation (i.e., 20 % to 80 %). In the rest phase it is interesting to note that the error rate remains practically unchanged for both groups of subjects. As long as the subject's gaze is not near to one of the objects of interest, the kernel in Equation 7.2 cancels the visual contribution in Equation 7.4, which relies exclusively on the myoelectric information. This is the major improvement that the proposed integration strategy has with respect to the approach used in our previous work (Gigli et al. 2018), in which the visual feature associated to a certain fixation was maintained until the next detected fixation. This propagation often "spilled" into the subsequent rest period, leading to an increase in misclassifications.

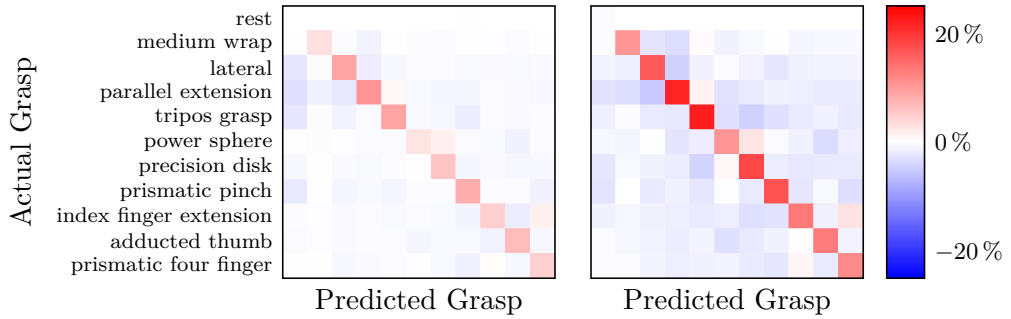
To estimate whether the improvement in performance concerns all the acquired grasps we evaluate the confusion matrices with and without the inclusion of the visual information. In general a confusion matrix provides for each class both its recognition score (diagonal elements) and the misclassification with the other classes (elements outside the diagonal). To estimate how much each class has gained from the introduction of vision, the confusion matrix of the standard sEMG-only classifier has been subtracted to the one of the proposed multimodal classifier. This difference is shown in Figure 7.3 for both intact and amputated subjects. The positive values on the diagonal indicate a uniform improvement of the classification accuracy over all the ten grasps. For the rest in position $(1, 1)$, no improvement nor worsening is observed. This is in line with the effect already observed in the rest phase of Figure 7.2. For the amputees, as already shown with classification accuracy results, the improvement in the diagonal is bigger than for intact subjects. This is moreover highlighted by the presence of negative values outside the diagonal, indicating that the number of misclassifications is decreasing.



(a) Intact subjects.

(b) Amputated subjects.

Figure 7.2. Normalized error of sEMG (blue dashed) and sEMG+gaze (red) classifiers. The integration of visual information with sEMG reduces the prediction error during the grasp both for (a) intact and (b) amputated subjects. The recognition over the rest was already good with the only sEMG modality and does not change when including gaze. The figure is best viewed in color.



(a) Intact subjects. **(b)** Amputated subjects.

Figure 7.3. Difference between the confusion matrices of the sEMG+gaze and sEMG classifiers. Positive values on the diagonal indicate better recognition of the relative classes when integrating visual information to sEMG both for (a) intact and (b) amputated subjects. The figure is best viewed in color.

Chapter 8

The Difficulty of Grasp Recognition during ADLs

The previous section demonstrated that gaze may hold useful information for the recognition of grasp intent. However, the evaluation was intentionally referred to as a proof-of-concept, since the study was conducted in a controlled laboratory environment and analyzed offline. The experimental setting involved objects that were already prepared on a table in front of the subjects, interactions with multiple objects were not included in the protocol, and a grasp was always followed by a rest period. Moreover, the participants were not engaged in ADLs and even the functional tasks, which are more elaborate than the static counterparts, can only be considered as a small part of more complex activities. These conditions are far from everyday life situations and inhibit completely natural behaviors.

Good academic results often do not translate to tangible improvements for prosthetic end-users. For instance, some studies have started to argue that offline performance metrics commonly reported in academic studies are only weakly related to the actual controllability of a prosthesis (Lock et al. 2005; Ortiz-Catalan et al. 2013; Jiang et al. 2014) or applicability in everyday life (Farina et al. 2014). Among the first to investigate the relationship between offline classification accuracy and online usability were Lock et al. (2005) and Hargrove et al. (2007a). They trained models for a number of intact subjects and subsequently evaluated these on a clothespins relocation task in a virtual environment. The primary result was that classification accuracy is only weakly related to online usability, as measured by the number of pins a subject could move. Other studies reported similar results when simply requiring subjects to match a target posture with a virtual hand (Li et al. 2010; Simon et al. 2011; Ortiz-Catalan et al. 2013; Jiang et al. 2014). Also Vujaklija et al. (2017) show that offline accuracy does not reflect performance obtained in clinical tests, like the popular Southampton Hand Assessment Protocol (SHAP) (Light et al. 2002) and the box-and-blocks test. Moreover both online and offline evaluation are at risk of overestimating real-life performance if not all sources of variability are taken into account. Failing to include the influence of limb position (Jiang et al. 2013; Khushaba et al. 2014; Boschmann and Platzner 2013), forearm rotation (Peng et al. 2013), and static and functional movements (Peng et al. 2013) can cause serious

degradation during the test evaluation.

The analyses presented in this chapter continue along these lines investigating whether a rich training phase is sufficient to detect movements executed during ADLs (Gregori et al. 2018). We take a data-centric approach to understand a possible degradation of performance by means of an sEMG dataset from two intact subjects in two sessions each. The first session was composed of standard laboratory tasks, and the second of ADLs. Unfortunately it was not possible to integrate the gaze data in this study due to the high level of invalid samples, particularly during the executions of ADLs. Since both parts were acquired in immediate succession and with intact subjects, we ignore signal variability due to electrode shift, fatigue, or muscle changes. The entire evaluation is performed offline, meaning that the eventual reactivity and adaptation of the user based on the immediate feedback they may receive in online testing is not addressed here. Instead, the objectives of this study are:

1. to provide a best-case analysis on whether grasps can be recognized if they are part of a composite, goal-oriented manipulation action such as an ADL;

2. to understand the eventual cause of a degradation by investigating the data distributions throughout the acquisition.

The acquired dataset and the preprocessing steps are described in Section 8.1. In Section 8.2 we present the results of the classification accuracy obtained from models trained on the only sEMG modality when multiple types of variability are considered. These findings are discussed and clarified with further analyses in Section 8.3.

8.1 Data Collection and Processing

The dataset we collected to perform our investigation consists of two parts, namely a set of grasps acquired in a laboratory setting with various sources of variability and the same grasps in a domestic environment when performing five ADLs. In the following we describe the acquisition protocol, the preprocessing routines, and the used classifiers for the sEMG data. The gaze data are analyzed independently, where we also motivate why we could not include them in these analyses.

8.1.1 Dataset

Two intact right-handed subjects (1M, 1F) participated in two experimental trials each, for a total of four data acquisitions, and each experiment was divided in a laboratory and a home part. The laboratory acquisition followed a standard experimental procedure, which was very similar to the ones used for the preliminary dataset and the final MeganePro dataset described in Chapter 4. The subject was in front of a table with a set of objects while a laptop provided audiovisual instructions on the required grasp and object. Table 8.1 lists the set of tasks and related grasps, which were selected based on their importance for amputees (Peerdeman et al. 2011) and their relevance for the subsequent ADLs. We first acquired so-called static movements, in which the subject was asked to simply grasp the object without performing any further manipulation action. These movements were acquired both

while seated and standing to vary the limb position when reaching the objects on the table. Then we enriched the dataset by performing the grasps as part of functional actions, such as moving or opening an object. In this case, we executed the movement either seated, standing, or both depending on which would be more likely in real life. All 51 movements in this laboratory phase were repeated six times to limit the total duration of the acquisition while still maintaining a sufficient number of repetitions to split the dataset during the analyses. Each repetition lasted approximately 4 s and they were separated by a rest period of around 3 s.

While the first part of the acquisition is similar to the MeganePro protocol, the second part was much less regulated and involved the subjects executing the five ADLs described in Table 8.2 in a real home environment. In particular, the previous grasps and functional movements from the laboratory acquisition were combined to compose more complex activities. As in a real situation, each grasp can be followed either by another grasp or by a rest posture, as opposed to the laboratory acquisition in which movements are always separated by the rest posture. The instructions in Table 8.2 were meant as a rough guide for the participants; they were free to make small deviations or interruptions and perform the activity with the speed that felt most natural.

The setup involved in this experiment is the same as the one adopted in the other acquisitions described in Section 4.1. However fourteen electrodes, instead of twelve, were used to acquire the sEMG signal: eight were equally spaced around the forearm at the height of the radio-humeral joint, while the remaining six were positioned in a similar equidistant configuration approximately 5 cm lower. A laptop stored the data acquired from the electrodes, from the Tobii glasses, and, only for the laboratory acquisitions, the label of the grasp that the subject was required to perform in each trial.

8.1.2 Processing and Classification

The videos recorded by the Tobii glasses were used to manually label grasps during the home acquisition, which could not be done automatically due to the unconstrained nature. A custom software application was developed to visualize this video together with the myoelectric signals, allowing us to classify the grasps and to determine the exact movement boundaries. When a grasp could not be reliably determined or in between ADLs, the corresponding part of the acquisition was marked as invalid and excluded from the analysis. An example of this process is shown in Figure 8.1. In total we obtained about 12 min of usable data per session in the home acquisitions.

The data preprocessing and classification procedures were inspired by the approach taken by Gijsberts et al. (2014). The preprocessing steps are a preliminary version of the final method developed for the MeganePro dataset (see Section 4.3.4). Initially, powerline interference was filtered from the myoelectric signals using a Hampel filter and, only for the laboratory acquisitions, the labels were realigned with actual sEMG activity. The usable data in the laboratory part are approximately 42 min. For the sEMG data all fourteen channels were standardized to have zero mean and unitary standard deviation, based on statistics calculated exclusively on the training set. Furthermore, we segmented the data using a sliding window of 200 ms and an increment of 10 ms (i.e., 20 samples).

Table 8.1. Overview of the tasks performed in the laboratory acquisition. The static activities were executed both seated and standing, while the functional movements were accomplished either seated, standing, or both.

Posture	Description
seated & standing (static)	Perform a static tripod grasp on the cap of a bottle Perform a static medium wrap on a bottle Perform a static stick grasp on a screwdriver Perform a static index finger extension on a knife in cutting position Perform a static lateral grasp on a key Perform a static prismatic pinch on a keyring
seated & standing (functional)	Move a little cup from a distant position in front of the subject and back (tripod grasp) Stir with a teaspoon inside a cup (tripod grasp) Move a bottle from a distant position in front of the subject and back (medium wrap) Move a pencil case from the right in front of the subject and back (medium wrap) Smear with a knife and put it back on the table (stick grasp) Scoop with a knife from a jar and put it back on the table (stick grasp) Cut with a knife and put it back on the table (index finger extension) Move a little cup from the handle from the right in front of the subject and back (lateral grasp) Pour from a little cup (lateral grasp) Move an upside down bottle cap from the right in front of the subject and back (prismatic pinch) Pick a sheet of paper from a distant position in front of the subject and back (prismatic pinch) Pour from a bottle (medium wrap)
seated (functional)	Open and close the cap of a bottle (tripod grasp) Move a pencil case from the floor in front of the subject and back (medium wrap) Open and close a horizontal zipper of a pencil case (lateral grasp) Move a power cable from the floor in the front of the subject and back (lateral grasp) Pick a cookie, bring it to mouth and put it back on the table (prismatic pinch) Open and close a book (prismatic pinch)
standing (functional)	Open and close a lid of a jam jar (tripod grasp) Move a medium jam jar from a cupboard in front of the subject and back (medium wrap) Make opening action with a door handle (medium wrap) Brush with a toothbrush near the subject's mouth and put it back on the table (stick grasp) Move a ruler resting on a pencil case up and down (lateral grasp) Open and close a horizontal zipper of backpack (lateral grasp) Open and close a vertical zipper of jacket (lateral grasp) Rotate a key and put it back on table (lateral grasp) Bring to mouth a little cup held from the handle and put it back on the table (lateral grasp)

Table 8.2. Overview of the ADLs performed in the home acquisition. The instructions to accomplish each task are reported with the suggested grasp type.

Task	Description
Prepare bread with jelly (standing)	Take bread (medium wrap) Take a knife (index finger extension) Cut a slice (index finger extension) Take a jar of jelly (medium wrap) Open the jar (tripod grasp) Take a knife (stick grasp) Scoop jelly with the knife (stick grasp) Smear the jelly on a slice of bread (stick grasp) Put the knife back (stick grasp) Bring the slice to the mouth (prismatic pinch)
Prepare coffee (standing)	Open the coffee machine (tripod grasp) Take out the filter (prismatic pinch) Place the filter on the table (prismatic pinch) Let the tap run (lateral grasp) Fill the base with water under the tap (medium wrap) Close the tap (lateral grasp) Put the filter in the coffee machine (prismatic pinch) Take a teaspoon (stick grasp) Fill the filter with coffee (stick grasp) Close the coffee machine (tripod grasp) Put the coffee machine on the stove (lateral grasp) Take a cup (tripod grasp) Take sugar (medium wrap) Take a teaspoon (stick grasp) Put sugar in the cup (stick grasp) Pour coffee (lateral grasp) Take a bottle of milk (medium wrap) Open the milk (tripod grasp) Pour the milk (medium wrap) Close the bottle of milk (tripod grasp) Stir the coffee with milk and sugar (tripod grasp) Bring the cup to the mouth (lateral grasp)
Open the door (standing)	Open a bag's zipper (lateral grasp) Take the keys from the bag (lateral grasp) Open the lock with a key (lateral grasp) Put the keys back in the bag (lateral grasp) Open the door with the handle (medium wrap)
Brush teeth (standing)	Open the toothpaste (tripod grasp) Take a toothbrush (stick grasp) Hold the toothbrush while applying toothpaste (stick grasp) Brush teeth (stick grasp) Clean the toothbrush under tap (stick grasp) Put the toothbrush back (stick grasp)
Do homework (seated)	Open a bag's zipper (lateral grasp) Take a pencil case from the bag (medium wrap) Open the zipper of the pencil case (lateral grasp) Take a piece of paper (prismatic pinch) Take a pen from the pencil case (stick grasp) Place a pen on the paper (stick grasp) Put the pen back in the pencil case (stick grasp) Plug a power cable in an outlet (lateral grasp) Open the laptop (prismatic pinch) Take a bottle from the right (medium wrap) Open the bottle (tripod grasp) Take a glass from the right side of the table (medium wrap) Pour water in the glass (medium wrap) Close the bottle (tripod grasp)

As feature representation and classifier we combined mDWT features (see Section 2.2.3) with a KRLS classifier with exponential χ^2 kernel for non-linearity (see Section 2.2.2). This setup is the same chosen for previous experiments in Chapter 4, Chapter 7, and in similar studies (Gijsberts et al. 2014). In some experiments, however, we compare its performance with the popular LDA (see Section 2.2.1) applied on top of standard RMS features (see Section 2.2.3). This classifier-features couple is considered as a baseline since LDA is a simple linear algorithm and RMS represents the less sophisticated features we can extract from the signal. For each of the following experiments the data were divided on the base of grasp repetitions so that approximately two thirds were used to train the models and the remaining one third to test it. For computational reasons, we subsampled the training data with a factor of 10. The regularization parameter λ of KRLS and its kernel bandwidth γ_{χ^2} were optimized using 4-fold cross validation over a grid with $\lambda \in \{2^{-16}, 2^{-15}, ..., 2^4\}$ and $\gamma_{\chi^2} \in \{2^{-20}, 2^{-19}, ..., 2^4\}$. The data were also in this case subsampled with an additional factor of 4 to speed up the optimization phase. The cross validation split was based on movement repetitions, such that available training repetitions were assigned to one of the folds in a round-robin fashion. We adopted this strategy to use a constant number of folds in each experiment and to guarantee at least one repetition per grasp in each fold. The classification accuracy used in the remainder of this chapter is defined as the proportion of windows in the test data that was classified correctly.

8.1.3 Analysis of Gaze Data

Contrary to the previous analysis in Chapter 7, the classifier used in this study takes into account only sEMG without including any visual information. Although it would be very interesting to understand whether and how vision helps grasp recognition especially during ADLs, in this case the gaze data could not be included due to the high level of invalid samples. Table 8.3 reports the percentage of invalid gaze-related data for each subject, session, and acquisition. In all but one session the percentage of invalid samples is lower than 20 % during the laboratory exercises. In contrast, during the home tasks this value increases to about 50 % in most cases, except for the second subject in the last session. This high quantity of invalid samples, which suggests that the Tobii glasses do not provide a good tracking in truly unconstrained situations, makes the inclusion of gaze data impossible in the following analyses. The presence of invalid samples is probably related to the combined effects of the standing posture, in which the majority of the exercises were performed, and the position of the manipulated objects, right near the subject. In these conditions the participants tended to look under the lenses of the glasses, but not through them, preventing the correct eye tracking. Moreover the eye tracking resulted difficult also when the attention rapidly switch from one object to another. This situation does not occur during the laboratory acquisitions since the movement to movement transitions were not taken into consideration.

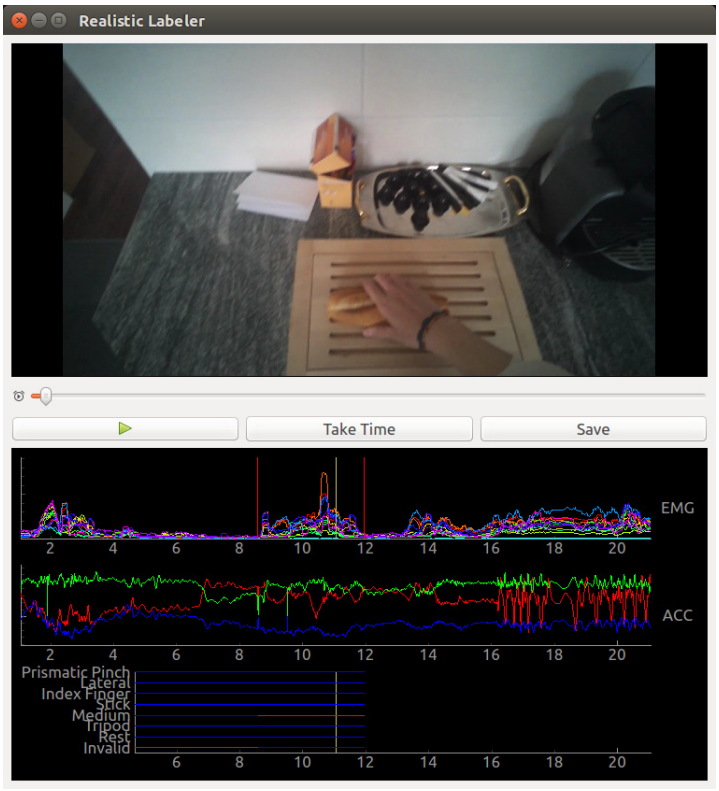


Figure 8.1. The graphical user interface of the software that was used to label the grasps executed during the home acquisitions. The video recorded by the scene camera of the Tobii glasses is shown on the top to visualize the action. The sEMG channels and three axes of one accelerometer are plotted to visualize the arm's activity. On the bottom the seven classes and an "invalid" option are listed. The user, by means of the "Take time" button, chooses the boundaries of the movement (red vertical lines overlaid on the sEMG channels) and one of the proposed classes (red horizontal line on the labels). The current video time is shown by the yellow horizontal line. The labeling process can be paused and the acquired labels can be saved to persistent storage.

Table 8.3. The proportion of invalid samples as recorded by the Tobii glasses for each subject, session, and acquisition. In particular, the 2-dimensional and 3-dimensional gaze position, and the pupil center position of the left and right eye are compared between the laboratory and home acquisitions. As highlighted by the bold notation, the data recorded during the home part always present an higher percentage of invalid samples.

		Invalid [%]							
		Gaze 2D		Gaze 3D		Pupil center left		Pupil center right	
Subject	Session	Lab	Home	Lab	Home	Lab	Home	Lab	Home
1	1	12.47	**56.70**	12.54	**56.71**	12.57	**39.85**	18.94	**45.29**
1	2	34.46	**64.52**	34.49	**64.54**	11.50	**53.69**	17.47	**66.67**
2	1	19.80	**51.58**	19.84	**51.69**	17.57	**40.89**	16.74	**27.31**
2	2	20.22	**25.32**	20.26	**25.33**	13.65	**19.69**	15.65	**21.86**

8.2 Classification Accuracy of sEMG

To provide insight on how well a laboratory procedure can be employed to predict grasps in an unconstrained home environment we evaluated multiple splits between the training and testing data, capturing different types of variability. In Figure 8.2a we report the classification accuracy per subject when training and testing are limited to the laboratory environment. We clearly observe poor performance when varying the subject's posture between the train and test phases, confirming the limb position effect reported by several previous studies and also in the *posture-split* analyzed in Section 4.2.3. Performance drops also when transferring from static to functional movements, which is due to the change in arm and hand dynamics when purposefully interacting with objects. This corresponds to the *dynamic-split* already analyzed in Section 4.2.3. Both detrimental effects can be compensated, however, by simply incorporating these types of variability in the training set. Indeed, when we align the train and test data by splitting the entire laboratory acquisition over repetitions (i.e., *trial-split* of Section 4.2.3), we obtain over 80 % classification accuracy.

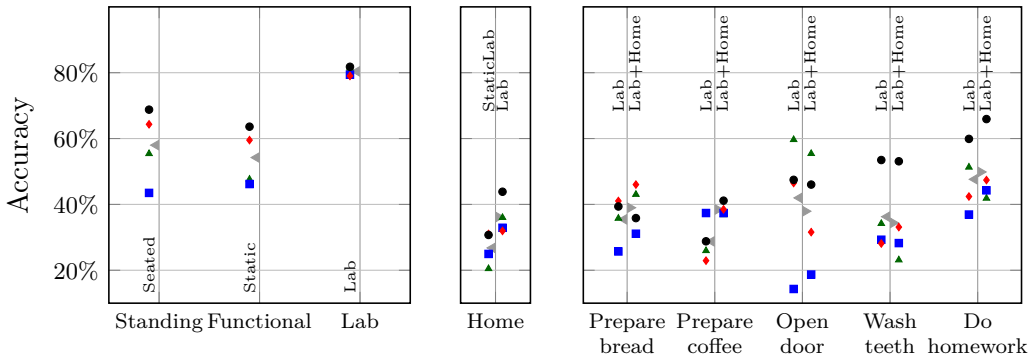
This last quantity is representative for the numbers and methodology often reported in studies on machine learning to recognize hand movements, it is indeed similar to results reported in our previous analyses in Section 4.2.3 and Section 7.2. Figure 8.2b shows that it unfortunately does not reflect real-life performance; a model trained in the laboratory setting obtains less than 40 % accuracy during the ADLs. Furthermore, the inclusion of functional movements does not significantly improve this result, even though they were selected specifically to be relevant for our ADLs. This is in agreement with the common observation that results in a controlled laboratory setting do not transfer to real life. We also investigated the potential gain of using the ADLs themselves as additional training data; while testing on one ADL we included the data from the remaining ones in the training set. However, Figure 8.2c shows that this does not lead to any consistent improvement.

8.3 Discussion

The previous results show that a model trained in a laboratory setting performs poorly when applied in a home environment. Here we further analyze this problem and provide an explanation of this degradation from a machine learning point of view.

8.3.1 Analysis of Misclassifications

To provide further insight we investigate in Figure 8.3 the change in misclassifications for various grasp types when moving from the laboratory to the home setting. As done for the analyses presented in Section 7.2.2, we subtract the confusion matrix of the classifier tested on the laboratory data from the one of the classifier tested on the home data. The negative values on the diagonal show that the decrease in performance is consistent for all types of grasps in the home environment. Interestingly, we also observe that movements are less often confused with "rest" (i.e., negative value in the first column of the matrix). This observation is caused by the rest-movement-rest sequence in our laboratory acquisition protocol: movement transitions, which are



(a) Laboratory **(b)** Home **(c)** Home Tasks

Figure 8.2. Classification accuracy per session with the KRLS classifier in different settings. First, we (a) compare train-test splits over the posture (i.e., *posture-split* from seated to standing) or the functional context (i.e., *dynamic-split* from static to functional) with a test on all laboratory data split over repetitions (i.e., *trial-split*). Then we (b) compare performance on home tasks when the classifier is trained either on the static movements or on the entire laboratory acquisition. Finally, we (c) test each ADL individually with models trained on just the laboratory acquisition and models that were enriched with data from the remaining ADLs. Each symbol corresponds to a subject and a trial. The gray triangle indicates the average over the sessions. The indications aligned with each vertical line and at the bottom of it specify respectively the training and test set.

typically difficult to predict due to their ambiguous nature (Hargrove et al. 2007a), always involve the rest posture. During ADLs, on the other hand, grasps can also immediately succeed other grasps, causing confusion in between these grasp types.

The importance of the transient phase of movements is demonstrated in Figure 8.4, which reports the error rate over the normalized duration of rest (−100 % to 0 %) and movements (0 % to 100 %). Errors in the laboratory setting are concentrated during transitions between rest and movement (−20 % to 20 %), and vice versa (80 % to 100 % and −100 % to −80 %). In contrast, during the home acquisition the error rate remains consistently high throughout the entire movement. This result applies both when the model is trained only on the static laboratory tasks or on the entire laboratory acquisition, showing only a small improvement in the second case. Together with the former result, this shows that the poor grasp recognition in real tasks is irrespective of the grasp type or the phase within a movement.

8.3.2 Domain Divergence

The drastic degradation we observe can be attributed to a mismatch in the domains of training and test data. Machine learning techniques typically assume that both datasets come from the same distribution. A violation of this assumption implies that the model is being evaluated on a different problem than it has been trained on, with poor performance as logical consequence. To estimate the difference between our laboratory and home distributions we adopt the measure of domain divergence proposed by Ben-David et al. (2007). The idea behind this divergence is that if two distributions are different, then a classifier should be able to distinguish them.

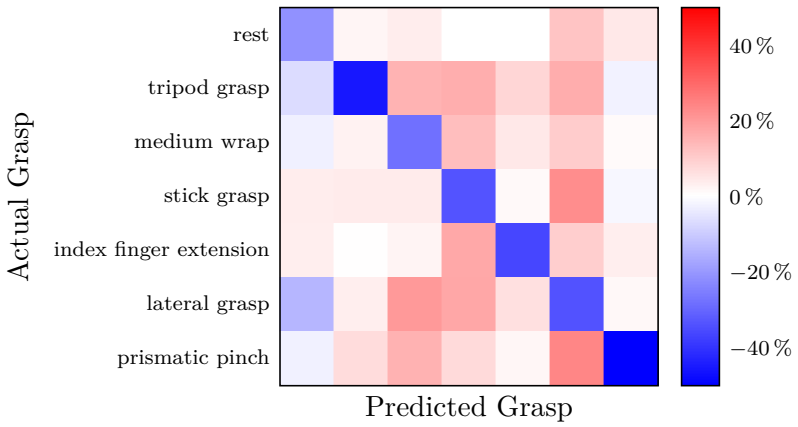


Figure 8.3. Difference between the confusion matrices when testing on the laboratory or home acquisition. Negative values on the diagonal indicate a decrease in correct classifications when passing from the laboratory to the home environment. The figure is best viewed in color.

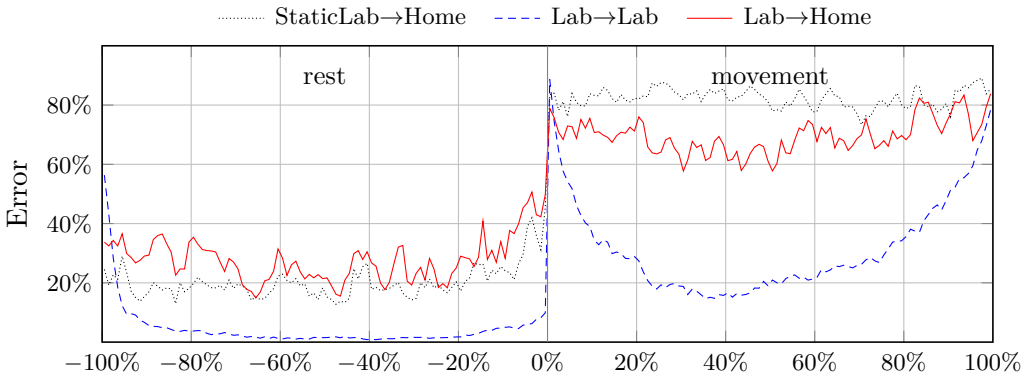


Figure 8.4. Distributions of errors with respect to normalized movement or rest duration. The blue line represents error behavior when training and testing on laboratory data, the black line when training on static laboratory data and testing on home data, and the red line when training on all laboratory data and testing on home data.

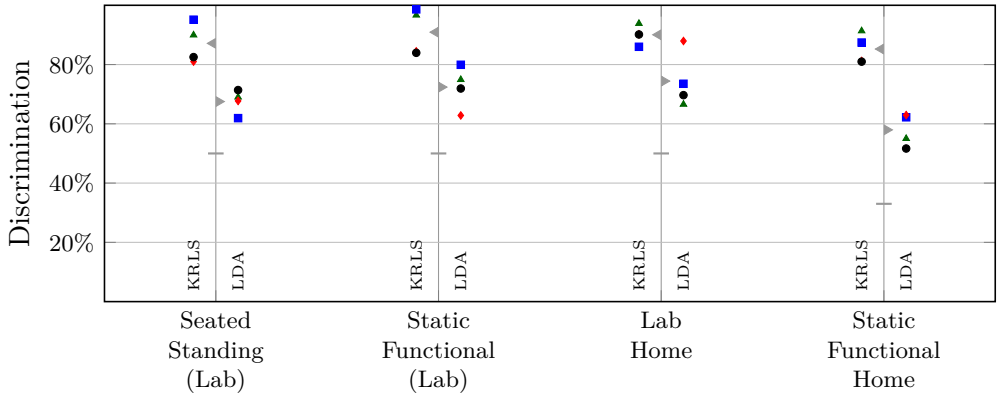


Figure 8.5. Balanced classification accuracy per session with balanced variants of the KRLS+mDWT and LDA+RMS classifier-feature combinations. In the first two columns we trained classifiers to distinguish between seated and standing, and between static and functional movements exclusively from the laboratory acquisition. In the last two cases the classifiers were trained to discriminate the environment (i.e., laboratory or home) and the movement "complexity" (static, functional and home). Each symbol corresponds to a subject and a trial. The gray triangle indicates the average over the sessions and the horizontal line marks chance level accuracy.

Furthermore, the higher the classifier accuracy, the larger is the divergence between the two domains.

In Figure 8.5 we apply this strategy to approximate the divergence between various subsets of our datasets. This check is implemented by simply training a classifier on a number of classes that correspond to the different variability factors we are considering. As shown, a balanced variant of our standard KRLS classifier can discriminate with high accuracy between movements executed while seated or standing, or also between static and functional movements. More interestingly, it also discriminates movements executed as part of the laboratory acquisition from those of the home acquisition with 90 % accuracy. Even a simple LDA with rectified RMS features is able to distinguish both settings with much higher accuracy than the chance level one would expect if both distributions were identical. This confirms that the high discrimination observed is independent from the non linearity introduced by the classifier.

8.3.3 Variability of Movements during ADLs

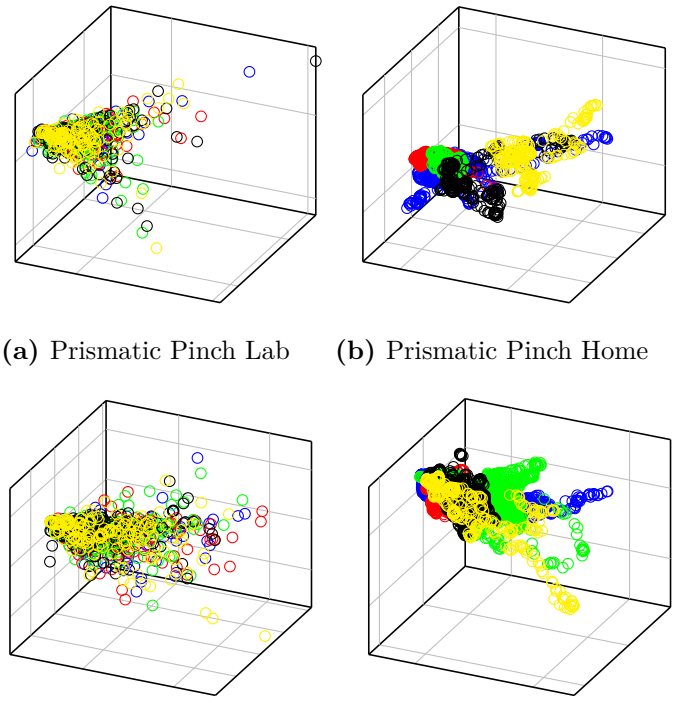
Besides demonstrating that movements are different when executed as part of ADLs, we are also interested in knowing how they are different. In Table 8.4 we report the average Euclidean intraclass distances for the grasps during the static, functional and home parts of the acquisitions. We observe that the variability increases from static to functional movements and similarly from laboratory to home movements. Differences in the context of the movement, such as the object weight, the arm dynamics, and the goal of the movement, are a probable explanation for this observation.

The difficulty with these sources of variability is that each ADL will have its own

Table 8.4. Average intraclass distances on RMS features. The distance increases as the considered variability and movement complexity increase.

	Intraclass distance ($\times 10^{-5}$)		
Grasp	Lab Static	Lab Functional	Home
Tripod Grasp	7.0	11.5	17.7
Medium Wrap	6.2	22.7	16.5
Stick Grasp	6.3	17.9	38.6
Index Finger Extension	6.3	14.3	15.2
Lateral Grasp	6.6	22.6	28.5
Prismatic Pinch	3.6	16.3	16.2

characteristic influence on myoelectric signals. This is observed in Figure 8.6, which visualizes the first three principal components of the rectified myoelectric signals of multiple repetitions of two example movements. As expected, repetitions in the laboratory setting are indistinguishable from one another. In the home acquisition, on the other hand, the repetitions of the same grasp correspond to different tasks or even different ADLs and we can easily distinguish them. Considering that our home acquisition contains only a tiny fraction of the set of activities a normal person engages in, it would be nearly impossible to construct a training set that sufficiently incorporates all possible variations.



(a) Prismatic Pinch Lab **(b)** Prismatic Pinch Home

(c) Lateral Grasp Lab **(d)** Lateral Grasp Home

Figure 8.6. Comparison of multiple repetitions of the same movement via the first three principal components of the rectified sEMG signals in the (a,c) laboratory and (b,d) home environments. Each color corresponds to one of the repetitions. The figure is best viewed in color.

Chapter 9

Conclusions

The motivation for this work was to investigate whether the integration of gaze behavior could improve grasp recognition in myoelectric prostheses. To study this idea, sEMG, gaze, and visual data were collected from 15 upper limb amputated subjects during the execution of grasps and manipulation tasks on household objects. For comparison also 30 able bodied subjects were engaged in the same experiments. This dataset was collected following standard procedures in a laboratory environment, asking the subjects to interact with several objects placed on a table in front of them with ten predefined hand configurations. The participants were equipped with electrodes to collect the sEMG signals from the forearm and with eye tracking glasses to record first person videos and the gaze position.

To understand *whether* and, if so, *when* the gaze can provide useful information for the recognition of the grasp, we thoroughly analyzed the coordination of eye, head, and hand movements of the participants. The main goal of these analyses was to verify the anticipatory role of gaze in the visuomotor strategy to estimate the window of opportunity in which an intelligent prosthesis could exploit the visual information to better understand the intended grasp. Furthermore, we were interested to understand whether the amputation has introduced important differences in the visuomotor coordination as compared to intact subjects. To perform these analyses, we devised an approach based on recent developments in deep learning to identify and segment in all videos the objects that were used in the acquisitions as well as the limb of our participants. The acquired segmentations were later combined with the gaze position within the image frame to automatically understand the fixated and grasped objects. This approach not only minimized the manual effort required to annotate the data, but also allowed to recognize the beginning of a fixation without delay, even in the presence of eye movements that serve to counteract head and body movements.

We found that a fixation on the target object typically preceded the subsequent grasp by approximately 500 ms. Moreover, the visuomotor strategies of amputees were similar to those of intact subjects both during the reach-to-grasp phase as well as during functional manipulation tasks. The only observed differences were related to the impossibility of the amputees to physically displace or interact with the objects due to the absence of a prosthetic device. We used this knowledge in a

proof of concept to combine vision with sEMG data in a multimodal kernel-based classifier for grasp recognition. Contrary to previous studies, we proposed a natural and simultaneous integration of sEMG and gaze data with the possibility to ignore the visual modality when it did not contain relevant information for the grasp recognition. The integration of vision improved the classification accuracy for all the subjects, making the average performance of the intact and amputated groups almost equal.

Finally, we investigated if similar results also hold outside the constrained laboratory setting, where an action or grasp is part of a complex goal-oriented activity performed in a home environment. In particular, we were interested to understand whether methods trained on a set of isolated movements can recognize these same movements during ADLs. To provide insight into this question we acquired data from two intact subjects in two independent trials each. The initial part of this acquisition consisted of repetitions of a set of six grasps in a variety of contexts and executed according to a typical rest-movement-rest protocol in a laboratory environment, as done for previous experiments. The second part of the acquisition was performed in a real home environment where the subjects completed five ADLs combining the previous grasps in unconstrained sequences. This study only involved sEMG data; gaze data unfortunately had to be excluded due to the large percentage of invalid samples in this unconstrained situation. Our analysis using machine learning methods showed that even when trained on a rich set of laboratory data we are not able to achieve satisfactory accuracy on the ADLs. The reason for this degradation is a divergence in the distribution of myoelectric data between the laboratory and home settings. This divergence violates the underlying assumption in many machine learning methods, namely that train and test data come from the same distribution. We further show that the trivial solution of simply adding some more data from some ADLs to predict other such activities is not sufficient to solve this problem. These results indicate that myoelectric data depend strongly on the context in which a grasp is performed, such as the goal of the movement, the size and weight of the manipulated object, and the arm dynamics.

9.1 Future Work

There are a number of possible directions that can be considered for future work. From the point of view of visuomotor coordination it would be interesting to investigate the gaze behavior in everyday life situations, as the practical engagement in daily living tasks would result in more natural behavior. Despite the difficulty of analyzing data from unconstrained acquisitions, this would give the possibility to study sequences of multiple activities where a subject is required to interact with several objects and to plan subsequent actions. Some studies in literature were already oriented in this direction, however the majority of them involved only a few subjects, probably due to difficulties in analyzing big amount of data (Land and Lee 1994; Land et al. 1999; Land and McLeod 2000; Patla and Vickers 2003; Hayhoe et al. 2012). To this aim it would be useful to test or extend the proposed segmentation framework in unconstrained situations to facilitate and automate the data processing.

Both in laboratory and everyday life situations, the visuomotor coordination of amputated subjects is altered if a prosthetic arm is used. It would be interesting to understand how the findings of Chapter 6 relate to the eye-hand coordination when using a prosthetic device. Previous studies (Sobuh et al. 2014; Bouwsema et al. 2012) have underlined that prosthetic users are more fixated on guiding the current manipulation, rather than planning the follow-up action. The prosthetic users declared that a high level of visual attention is required when performing certain functions with the prosthetic arm (Atkins et al. 1996). This behavior is most likely caused by the fact that amputated people rely almost exclusively on visual feedback. However, since only a small number of subjects were engaged in the previous studies more research will be needed to fully understand the disruption of the visuomotor strategy. In particular, whether or not this strategy improves when the user develops trust in the prosthesis (Chadwell et al. 2016) merits attention. Another equally interesting question is to which extent the proactive gaze behavior can be restored by integrating tactile or proprioceptive feedback in the prosthesis (Cipriani et al. 2011; Markovic et al. 2018; Marasco et al. 2018, among others).

Another problem that should be addressed is that the Tobii glasses result uncomfortable when used on a daily basis. The unattractive appearance of many prosthetic arms is one of the cause of abandonment of some of these devices (Atkins et al. 1996; Biddiss and Chau 2007a), so the addition of another obtrusive and cumbersome device would almost surely not be accepted by the vast majority of end-users. A solution to maintain the acquisition of "visual information" without burdening the user would be to embed a camera in the prosthesis. Although a few studies have already reported interesting results in this direction (Došen et al. 2010; Došen and Popović 2011; DeGol et al. 2016; Taverne et al. 2019), the integration of sEMG with visual information is often absent or unnatural. Moreover, a disadvantage of having the camera in the palm is that the object is only seen once the reaching phase is almost terminated, whereas the eyes already fixate the object even before the arm starts to move. Therefore, this anticipatory nature would be missing and the window of opportunity for the integration of visual information would be shorter.

Regardless of the devices used to acquire data, it seems necessary to explore performance metrics beyond the mere offline evaluation. In the last chapter of this thesis we observed that the performance of grasp recognition drastically decreases when they are part of ADLs performed in home environments. To address this issue, we should investigate whether this train and test discrepancy can be addressed using incremental learning, in which training becomes an integral part of the daily operation of a prosthesis. This implies (1) the life-long collection of new data every time deemed necessary by the user (Strazzulla et al. 2016; Patel et al. 2017) and (2) to pass from offline to online evaluation where the reaction and the adaptation of the user is not only taken into account but even embraced (Hahne et al. 2015; Hahne et al. 2017).

Bibliography

[1] Henny Admoni and Siddhartha Srinivasa. "Predicting User Intent Through Eye Gaze for Shared Autonomy". In: *2016 AAAI Fall Symposia*. Arlington, Virginia, USA, Nov. 2016.

[2] Edgar D Adrian and Detlev W Bronk. "The discharge of impulses in motor nerve fibres: Part II. The frequency of discharge in reflex and voluntary contractions". In: *The Journal of physiology* 67.2 (1929), pp. 9–151.

[3] David P. Allen. "A frequency domain Hampel filter for blind rejection of sinusoidal interference from electromyograms". In: *Journal of Neuroscience Methods* 177.2 (2009), pp. 303–310.

[4] Reuben M. Aronson and Henny Admoni. "Gaze for Error Detection During Human-Robot Shared Manipulation". In: *Fundamentals of Joint Action workshop, Robotics: Science and Systems*. 2018.

[5] Reuben M. Aronson, Thiago Santini, Thomas C. Kübler, Enkelejda Kasneci, Siddhartha Srinivasa, and Henny Admoni. "Eye-hand behavior in human-robot shared manipulation". In: *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM. 2018, pp. 4–13.

[6] Diane J Atkins, Denise CY Heard, and William H Donovan. "Epidemiologic overview of individuals with upper-limb loss and their reported research priorities". In: *JPO: Journal of Prosthetics and Orthotics* 8.1 (1996), pp. 2–11.

[7] Manfredo Atzori, Arjan Gijsberts, Henning Müller, and Barbara Caputo. "Classification of hand movements in amputated subjects by sEMG and accelerometers". In: *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Aug. 2014, pp. 3545–3549.

[8] Manfredo Atzori, Arjan Gijsberts, Claudio Castellini, Barbara Caputo, Anne-Gabrielle Mittaz Hager, Simone Elsig, Giorgio Giatsidis, Franco Bassetto, and Henning Müller. "Electromyography Data for Non-Invasive Naturally-Controlled Robotic Hand Prostheses". In: *Scientific Data* 1 (Dec. 2014). Data Descriptor.

[9] James R. Augustine. *Human neuroanatomy*. Academic Press, 2008.

[10] Robert W. Baloh, Andrew W. Sills, Warren E. Kumley, and Vicente Honrubia. "Quantitative measurement of saccade amplitude, duration, and velocity". In: *Neurology* 25.11 (1975), p. 1065.

[11] Linchao Bao, Baoyuan Wu, and Wei Liu. "CNN in MRF: Video object segmentation via inference in a CNN-based higher-order spatio-temporal MRF". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 5977–5986.

[12] John V. Basmajian and Robert Blumenstein. *Electrode placement in EMG biofeedback*. Williams & Wilkins Baltimore, 1980.

[13] John V. Basmajian and Carlo J. De Luca. *Muscles alive*. 5th ed. Baltimore: Williams & Wilkins, 1985.

[14] Anna Belardinelli, Madeleine Y. Stepper, and Martin V. Butz. "It's in the eyes: Planning precise manual actions before execution". In: *Journal of vision* 16.1 (2016), p. 18.

[15] Joseph T. Belter, Jacob L. Segil, Aaron M. Dollar, and Richard F. Weir. "Mechanical design and performance specifications of anthropomorphic prosthetic hands: A review." In: *Journal of Rehabilitation Research & Development* 50.5 (2013).

[16] Shai Ben-David, John Blitzer, Koby Crammer, and Fernando Pereira. "Analysis of representations for domain adaptation". In: *Advances in neural information processing systems*. 2007, pp. 137–144.

[17] Daniel A. Bennett and Michael Goldfarb. "IMU-based wrist rotation control of a transradial myoelectric prosthesis". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 26.2 (2017), pp. 419–427.

[18] Massimo Bergamasco, Antonio Frisoli, Marco Fontana, Claudio Loconsole, Daniele Leonardis, Marco Troncossi, Mohammad Mozaffari Foumashi, and Vincenzo Parenti-Castelli. "Preliminary results of BRAVO project: brain computer interfaces for Robotic enhanced Action in Visuo-motor tasks". In: *2011 IEEE International Conference on Rehabilitation Robotics*. IEEE. 2011, pp. 1–7.

[19] Luca Bertinetto, Jack Valmadre, Joao F. Henriques, Andrea Vedaldi, and Philip H. S. Torr. "Fully-convolutional siamese networks for object tracking". In: *European conference on computer vision*. Springer. 2016, pp. 850–865.

[20] Elaine Biddiss and Tom Chau. "Upper-limb prosthetics: critical factors in device abandonment". In: *American journal of physical medicine & rehabilitation* 86.12 (2007), pp. 977–987.

[21] Elaine Biddiss, Dorcas Beaton, and Tom Chau. "Consumer design priorities for upper limb prosthetics". In: *Disability and Rehabilitation: Assistive Technology* 2.6 (2007), pp. 346–357.

[22] Elaine A Biddiss and Tom T Chau. "Upper limb prosthesis use and abandonment: a survey of the last 25 years". In: *Prosthetics and orthotics international* 31.3 (2007), pp. 236–257.

[23] Christopher M. Bishop. *Pattern recognition and machine learning*. springer, 2006.

[24] Amy Blank, Allison M Okamura, and Katherine J Kuchenbecker. "Identifying the role of proprioception in upper-limb prosthesis control: Studies on targeted motion". In: *ACM Transactions on Applied Perception (TAP)* 7.3 (2010), p. 15.

[25] Pieter Blignaut and Daniël Wium. "Eye-tracking data quality as affected by ethnicity and experimental design". In: *Behavior research methods* 46.1 (2014), pp. 67–80.

[26] A. Boschmann and M. Platzner. "Reducing the limb position effect in pattern recognition based myoelectric control using a high density electrode array". In: *ISSNIP Biosignals and Biorobotics Conference (BRC)*. Feb. 2013, pp. 1–5.

[27] Hanneke Bouwsema, Peter J. Kyberd, Wendy Hill, Corry K. Van Der Sluis, and Raoul M. Bongers. "Determining skill level in myoelectric prosthesis use with multiple outcome measures". In: *Journal of Rehabilitation Research and Development* 49.9 (2012), pp. 1331–1347.

[28] Miles C Bowman, Roland S Johannson, and John Randall Flanagan. "Eye–hand coordination in a sequential target contact task". In: *Experimental brain research* 195.2 (2009), pp. 273–283.

[29] Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah. "Signature verification using a" siamese" time delay neural network". In: *Advances in neural information processing systems*. 1994, pp. 737–744.

[30] Anne-Marie Brouwer, Volker H. Franz, and Karl R. Gegenfurtner. "Differences in fixations between grasping and viewing objects". In: *Journal of Vision* 9.1 (2009), pp. 18–24.

[31] Ian M. Bullock, Joshua Z. Zheng, Sara De La Rosa, Charlotte Guertler, and Aaron M. Dollar. "Grasp frequency and usage in daily household and machine shop tasks". In: *IEEE Transactions on Haptics* 6.3 (2013), pp. 296–308.

[32] Helena Burger. "Return to work after amputation". In: *Amputation, prosthesis use, and phantom limb pain*. Springer, 2009, pp. 101–114.

[33] Stephanie L Carey, Derek J Lura, and M Jason Highsmith. "Differences in myoelectric and body-powered upper-limb prostheses: Systematic literature review." In: *Journal of Rehabilitation Research & Development* 52.3 (2015).

[34] Roger H. S. Carpenter. *Movements of the Eyes*. Pion Limited, 1988.

[35] Claudio Castellini and Georg Passig. "Ultrasound image features of the wrist are linearly related to finger positions". In: *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2011, pp. 2108–2114.

[36] Claudio Castellini and Giulio Sandini. "Gaze tracking for robotic control in intelligent teleoperation and prosthetics". In: Jan. 2006, pp. 73–77.

[37] Claudio Castellini and Patrick van der Smagt. "Surface EMG in advanced hand prosthetics". In: *Biological Cybernetics* 100.1 (2009), pp. 35–47.

[38] Claudio Castellini, Panagiotis Artemiadis, Michael Wininger, Arash Ajoudani, Merkur Alimusaj, Antonio Bicchi, Barbara Caputo, William Craelius, Strahinja Dosen, Kevin Englehart, Dario Farina, Arjan Gijsberts, Sasha B. Godfrey, Levi Hargrove, Mark Ison, Todd Kuiken, Marko Marković, Patrick M. Pilarski, Rüdiger Rupp, and Erik Scheme. "Proceedings of the first workshop on Peripheral Machine Interfaces: going beyond traditional surface electromyography". In: *Frontiers in Neurorobotics* 8 (2014), p. 22.

[39] Alix Chadwell, Laurence Kenney, Sibylle Thies, Adam Galpin, and John Head. "The Reality of Myoelectric Prostheses: Understanding What Makes These Devices Difficult for Some Users to Control". In: *Frontiers in Neurorobotics* 10 (2016), p. 7.

[40] Xiang Chen, Xu Zhang, Zhang-Yan Zhao, Ji-Hai Yang, Vuokko Lantz, and Kong-Qiao Wang. "Hand gesture recognition research based on surface EMG sensors and 2D-accelerometers". In: *2007 11th IEEE International Symposium on Wearable Computers*. IEEE. 2007, pp. 11–14.

[41] Christian Cipriani, Marco Controzzi, and Maria Chiara Carrozza. "The SmartHand transradial prosthesis". In: *Journal of neuroengineering and rehabilitation* 8.1 (2011), p. 29.

[42] Edward A. Clancy, Mark V. Bertolina, Roberto Merletti, and Dario Farina. "Time- and frequency-domain monitoring of the myoelectric signal during a long-duration, cyclic, force-varying, fatiguing hand-grip task". In: *Journal of Electromyography and Kinesiology* 18.5 (2008), pp. 789 –797.

[43] Francesco Clemente, Strahinja Dosen, Luca Lonini, Marko Markovic, Dario Farina, and Christian Cipriani. "Humans can integrate augmented reality feedback in their sensorimotor control of a robotic hand". In: *IEEE Transactions on Human-Machine Systems* 47.4 (2016), pp. 583–589.

[44] Coapt, LLC. *Coapt Engineering*. `https://www.coaptengineering.com`. 2015. URL: `https://www.coaptengineering.com`.

[45] Matteo Cognolato, Arjan Gijsberts, Valentina Gregori, Gianluca Saetta, Katia Giacomino, Anne-Gabrielle Mittaz Hager, Andrea Gigli, Diego Faccio, Cesare Tiengo, Franco Bassetto, Barbara Caputo, Peter Brugger, Manfredo Atzori, and Henning Müller. "Gaze, Visual, Myoelectric, and Inertial Data of Grasps for Intelligent Prosthetics". In: *medRxiv* (2019).

[46] Matteo Cognolato, Arjan Gijsberts, Valentina Gregori, Gianluca Saetta, Katia Giacomino, Anne-Gabrielle Mittaz Hager, Andrea Gigli, Diego Faccio, Cesare Tiengo, Franco Bassetto, Barbara Caputo, Peter Brugger, Manfredo Atzori, and Henning Müller. "Gaze, Visual, Myoelectric, and Inertial Data of Grasps for Intelligent Prosthetics". In: *Scientific Data* 7 (2020), p. 43.

[47] Elaine A. Corbett, Nicholas A. Sachs, Konrad P. Körding, and Eric J. Perreault. "Multimodal decoding and congruent sensory information enhance reaching performance in subjects with cervical spinal cord injury". In: *Frontiers in Neuroscience* 8 (2014), p. 123.

[48] Elaine A. Corbett, Konrad P. Kording, and Eric J. Perreault. "Real-time fusion of gaze and EMG for a reaching neuroprosthesis". In: *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. 2012, pp. 739–742.

[49] Alain Courteville, Tijani Gharbi, and Jean-Yves Cornu. "MMG measurement: A high-sensitivity microphone-based sensor for clinical use". In: *IEEE transactions on biomedical engineering* 45.2 (1998), pp. 145–150.

[50] Beau Crawford, Kai Miller, Pradeep Shenoy, and Rajesh Rao. "Real-Time Classification of Electromyographic Signals for Robotic Control". In: *Proceedings of AAAI*. 2005, pp. 523–528.

[51] Eleanor Criswell. *Cram's introduction to surface electromyography*. Jones & Bartlett Publishers, 2010.

[52] David J. Curcie, James A. Flint, and William Craelius. "Biomimetic finger control by filtering of distributed forelimb pressures". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 9.1 (2001), pp. 69–75.

[53] Mark R. Cutkosky. "On grasp choice, grasp models, and the design of hands for manufacturing tasks". In: *IEEE Transactions on Robotics and Automation* 5.3 (1989), pp. 269–279.

[54] Jifeng Dai, Kaiming He, and Jian Sun. "Instance-aware semantic segmentation via multi-task network cascades". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 3150–3158.

[55] Carlo J. De Luca. "The use of surface electromyography in biomechanics". In: *Journal of applied biomechanics* 13.2 (1997), pp. 135–163.

[56] Carlo J. De Luca. "The use of surface electromyography in biomechanics". In: *Journal of applied biomechanics* 13.2 (1997), pp. 135–163.

[57] Carlo J. De Luca and Roberto Merletti. "Surface myoelectric signal cross-talk among muscles of the leg". In: *Electroencephalography and clinical neurophysiology* 69.6 (1988), pp. 568–575.

[58] Joseph DeGol, Aadeel Akhtar, Bhargava Manja, and Timothy Bretl. "Automatic grasp selection using a camera in a hand prosthesis". In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2016, pp. 431–434.

[59] C. Scotto Di Cesare, D. Anastasopoulos, L. Bringoux, Pei-Yun Lee, M. J. Naushahi, and A. M Bronstein. "Influence of postural constraints on eye and head latency during voluntary rotations". In: *Vision research* 78 (2013), pp. 1–5.

[60] N. A. Dimitrova, G. V. Dimitrov, and O. A. Nikitin. "Neither high-pass filtering nor mathematical differentiation of the EMG signals can considerably reduce cross-talk". In: *Journal of Electromyography and Kinesiology* 12.4 (2002), pp. 235–246.

[61] Strahinja Došen and Dejan B. Popović. "Transradial prosthesis: artificial vision for control of prehension". In: *Artificial organs* 35.1 (2011), pp. 37–48.

[62] Strahinja Došen, Christian Cipriani, Miloš Kostić, Marco Controzzi, Maria C. Carrozza, and Dejan B. Popović. "Cognitive vision system for control of dexterous prosthetic hands: experimental evaluation". In: *Journal of neuroengineering and rehabilitation* 7.1 (2010), p. 42.

[63] Andrew T. Duchowski. *Eye Tracking Methodology: Theory and Practice*. 3rd ed. Berlin, Heidelberg: Springer-Verlag, 2017.

[64] Kevin Englehart and Bernard Hudgins. "A robust, real-time control scheme for multifunction myoelectric control". In: *IEEE transactions on biomedical engineering* 50.7 (2003), pp. 848–854.

[65] Kevin Englehart, Bernard Hudgins, Philip A. Parker, and Maryhelen Stevenson. "Classification of the myoelectric signal using time-frequency based representations". In: *Medical engineering & physics* 21.6-7 (1999), pp. 431–438.

[66] Julio Fajardo, Victor Ferman, Amparo Muñoz, Dandara Andrade, Antonio Ribas Neto, and Eric Rohmer. "User-Prosthesis Interface for Upper Limb Prosthesis Based on Object Classification". In: *2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE)*. IEEE. 2018, pp. 390–395.

[67] Dario Farina and Francesco Negro. "Accessing the neural drive to muscle and translation to neurorehabilitation technologies". In: *IEEE Reviews in biomedical engineering* 5 (2012), pp. 3–14.

[68] Dario Farina, Corrado Cescon, and Roberto Merletti. "Influence of anatomical, physical, and detection-system parameters on surface EMG". In: *Biological cybernetics* 86.6 (2002), pp. 445–456.

[69] Dario Farina, Ning Jiang, Hubertus Rehbaum, Aleš Holobar, Bernhard Graimann, Hans Dietl, and Oskar C Aszmann. "The extraction of neural information from the surface EMG for the control of upper-limb prostheses: emerging avenues and challenges". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 22.4 (2014), pp. 797–809.

[70] Dario Farina, Roberto Merletti, and Roger M. Enoka. "The extraction of neural strategies from the surface EMG". In: *Journal of applied physiology* 96.4 (2004), pp. 1486–1495.

[71] Thomas Feix, Roland Pawlik, Heinz-Bodo Schmiedmayer, Javier Romero, and Danica Kragic. "A comprehensive grasp taxonomy". In: *Robotics, Science and Systems: Workshop on Understanding the Human Hand for Advancing Robotic Manipulation*. Vol. 2. 2.3. 2009, pp. 2–3.

[72] Thomas Feix, Javier Romero, Heinz-Bodo Schmiedmayer, Aaron M Dollar, and Danica Kragic. "The grasp taxonomy of human grasp types". In: *IEEE Transactions on Human-Machine Systems* 46.1 (2015), pp. 66–77.

[73] J Randall Flanagan and Roland S Johansson. "Action plans used in action observation". In: *Nature* 424.6950 (2003), p. 769.

[74] Anders Fougner, Erik Scheme, Adrian DC Chan, Kevin Englehart, and Øyvind Stavdahl. "Resolving the limb position effect in myoelectric pattern recognition". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 19.6 (2011), pp. 644–651.

[75] Antonio Frisoli, Claudio Loconsole, Daniele Leonardis, Filippo Banno, Michele Barsotti, Carmelo Chisari, and Massimo Bergamasco. "A new gaze-BCI-driven control of an upper limb exoskeleton for rehabilitation in real-world tasks". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42.6 (2012), pp. 1169–1179.

[76] Marcus Gardner, Richard Woodward, Ravi Vaidyanathan, Etienne Bürdet, and Boo Cheong Khoo. "An unobtrusive vision system to reduce the cognitive burden of hand prosthesis control". In: *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*. IEEE. 2014, pp. 1279–1284.

[77] Purushothaman Geethanjali. "Myoelectric control of prosthetic hands: state-of-the-art review". In: *Medical Devices (Auckland, NZ)* 9 (2016), p. 247.

[78] Yanjuan Geng, Ping Zhou, and Guanglin Li. "Toward attenuating the impact of arm positions on electromyography pattern-recognition based motion classification in transradial amputees". In: *Journal of neuroengineering and rehabilitation* 9.1 (2012), p. 74.

[79] Marcus Georgi, Christoph Amma, and Tanja Schultz. "Recognizing Hand and Finger Gestures with IMU based Motion and EMG based Muscle Activity Sensing." In: *Biosignals*. 2015, pp. 99–108.

[80] Ghazal Ghazaei, Ali Alameer, Patrick Degenaar, Graham Morgan, and Kianoush Nazarpour. "Deep learning-based artificial vision for grasp classification in myoelectric hands". In: *Journal of neural engineering* 14.3 (2017).

[81] Andrea Gigli, Valentina Gregori, Matteo Cognolato, Manfredo Atzori, and Arjan Gijsberts. "Visual Cues to Improve Myoelectric Control of Upper Limb Prostheses". In: *IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*. 2018, pp. 783–788.

[82] Arjan Gijsberts, Manfredo Atzori, Claudio Castellini, Henning Müller, and Barbara Caputo. "Movement error rate for evaluation of machine learning methods for sEMG-based hand movement classification". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 22.4 (2014), pp. 735–744.

[83] Francesca Giordaniello, Matteo Cognolato, Mara Graziani, Arjan Gijsberts, Valentina Gregori, Gianluca Saetta, Anne-Gabrielle Mittaz Hager, Cesare Tiengo, Franco Bassetto, Peter Brugger, Barbara Caputo, Henning Müller, and Manfredo Atzori. "Megane Pro: Myo-Electricity, Visual and Gaze Tracking Data Acquisitions to Improve Hand Prosthetics". In: *IEEE International Conference on Rehabilitation Robotics (ICORR)*. 2017, pp. 1148–1153.

[84] Joseph H. Goldberg and Jack C. Schryver. “Eye-gaze-contingent control of the computer interface: Methodology and example for zoom detection”. In: *Behavior research methods, instruments, & computers* 27.3 (1995), pp. 338–350.

[85] Michael S. Goldenberg, H. John Yack, Frank J. Cerny, and Harold W. Burton. “Acoustic myography as an indicator of force during sustained contractions of a small hand muscle”. In: *Journal of applied physiology* 70.1 (1991), pp. 87–91.

[86] Jenny E. Goldring, Michael C. Dorris, Brian D. Corneil, Peter A. Ballantyne, and Douglas R. Munoz. “Combined eye-head gaze shifts to visual and auditory targets in humans”. In: *Experimental brain research* 111.1 (1996), pp. 68–78.

[87] Jozina B. De Graaf, Nathanaël Jarrassé, Caroline Nicol, Amélie Touillet, Thelma Coyle, Luc Maynard, Noël Martinet, and Jean Paysant. “Phantom hand and wrist movements in upper limb amputees are slow but naturally controlled movements”. In: *Neuroscience* 312 (2016), pp. 48–57.

[88] Daniel Graupe, A. A. Beex, William J. Monlux, and Ian Magnussen. “A multifunctional prosthesis control system based on time series identification of EMG signals using microprocessors.” In: *Bulletin of prosthetics research* 10.27 (1977), pp. 4–16.

[89] Valentina Gregori, Arjan Gijsberts, and Barbara Caputo. “Adaptive learning to speed-up control of prosthetic hands: A few things everybody should know”. In: *IEEE International Conference on Rehabilitation Robotics (ICORR)*. 2017, pp. 1130–1135.

[90] Valentina Gregori, Matteo Cognolato, Gianluca Saetta, Manfredo Atzori, The MeganePro Consortium, and Arjan Gijsberts. “On the Visuomotor Behavior of Amputees and Able-Bodied People During Grasping”. In: *Frontiers in Bioengineering and Biotechnology* 7 (2019), p. 316.

[91] Valentina Gregori, Barbara Caputo, and Arjan Gijsberts. “The Difficulty of Recognizing Grasps from sEMG during Activities of Daily Living”. In: *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*. 2018, pp. 583–588.

[92] Jing-Yi Guo, Yong-Ping Zheng, Hong-Bo Xie, and Terry K. Koo. “Towards the application of one-dimensional sonomyography for powered upper-limb prosthetic control using machine learning models”. In: *Prosthetics and orthotics international* 37.1 (2013), pp. 43–49.

[93] Janne M. Hahne, Sven Dähne, Han-Jeong Hwang, Klaus-Robert Müller, and Lucas C. Parra. “Concurrent adaptation of human and machine improves simultaneous and proportional myoelectric control”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 23.4 (2015), pp. 618–627.

[94] Janne M. Hahne, Marko Markovic, and Dario Farina. “User adaptation in myoelectric man-machine interfaces”. In: *Scientific reports* 7.1 (2017), p. 4437.

[95] Maria Hakonen, Harri Piitulainen, and Arto Visala. “Current state of digital signal processing in myoelectric interfaces and related applications”. In: *Biomedical Signal Processing and Control* 18 (2015), pp. 334–359.

[96] Yaoyao Hao, Marco Controzzi, Christian Cipriani, Dejan B Popovic, Xin Yang, Weidong Chen, Xiaoxiang Zheng, and Maria Chiara Carrozza. "Controlling hand-assistive devices: utilizing electrooculography as a substitute for vision". In: *IEEE Robotics & Automation Magazine* 20.1 (2013), pp. 40–52.

[97] Levi Hargrove, Yves Losier, Blair Lock, Kevin Englehart, and Bernard Hudgins. "A real-time pattern recognition based myoelectric control usability study implemented in a virtual environment". In: *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. 2007, pp. 4842–4845.

[98] Levi J. Hargrove, Kevin Englehart, and Bernard Hudgins. "A comparison of surface and intramuscular myoelectric signal classification". In: *IEEE transactions on biomedical engineering* 54.5 (2007), pp. 847–853.

[99] Mary M. Hayhoe, Travis McKinney, Kelly Chajka, and Jeff B. Pelz. "Predictive eye movements in natural vision". In: *Experimental brain research* 217.1 (2012), pp. 125–136.

[100] Mark Hays, Luke Osborn, Rohan Ghosh, Mark Iskarous, Christopher Hunt, and Nitish V. Thakor. "Neuromorphic vision and tactile fusion for upper limb prosthesis control". In: *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE. 2019, pp. 981–984.

[101] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[102] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. "Mask r-cnn". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2961–2969.

[103] Jacqueline S. Hebert, Quinn A. Boser, Aïda M. Valevicius, Hiroki Tanikawa, Ewen B. Lavoie, Albert H. Vette, Patrick M. Pilarski, and Craig S. Chapman. "Quantitative Eye Gaze and Movement Differences in Visuomotor Adaptations to Varying Task Demands Among Upper-Extremity Prosthesis Users". In: *JAMA Network Open* 2.9 (Sept. 2019).

[104] Peter Herberts, Christian Almström, Roland Kadefors, and Peter D. Lawrence. "Hand prosthesis control via myoelectric patterns". In: *Acta Orthopaedica Scandinavica* 44.4-5 (1973), pp. 389–409.

[105] Irving Herrera-Luna, Ericka Janet Rechy-Ramirez, Homero Vladimir Rios-Figueroa, and Antonio Marin-Hernandez. "Sensor Fusion Used in Applications for Hand Rehabilitation: A Systematic Review". In: *IEEE Sensors Journal* 19.10 (2019), pp. 3581–3592.

[106] Roy S. Hessels, Diederick C. Niehorster, Chantal Kemner, and Ignace T.C. Hooge. "Noise-robust fixation detection in eye movement data: Identification by two-means clustering (I2MC)". In: *Behavior research methods* 49.5 (2017), pp. 1802–1823.

[107] Corey D. Holland and Oleg V. Komogortsev. "Complex eye movement pattern biometrics: Analyzing fixations and saccades". In: *International conference on biometrics (ICB)*. IEEE. 2013, pp. 1–8.

[108] Kenneth Holmqvist, Marcus Nyström, and Fiona Mulvey. “Eye tracker data quality: What it is and how to measure it”. In: *Eye Tracking Research and Applications Symposium (ETRA)* (2012).

[109] Mark E. Huang, Charles E. Levy, and Joseph B. Webster. “Acquired limb deficiencies. 3. Prosthetic components, prescriptions, and indications”. In: *Archives of physical medicine and rehabilitation* 82.3 (2001), S17–S24.

[110] Bernard Hudgins, Philip Parker, and Robert N. Scott. “A new strategy for multifunction myoelectric control”. In: *IEEE Transactions on Biomedical Engineering* 40.1 (1993), pp. 82–94.

[111] M. Anamul Islam, Kenneth Sundaraj, R. Badlishah Ahmad, and Nizam Uddin Ahamed. “Mechanomyogram for muscle function assessment: a review”. In: *PloS one* 8.3 (2013).

[112] Mark Ison and Panagiotis Artemiadis. “The role of muscle synergies in myoelectric control: trends and challenges for simultaneous multifunction control”. In: *Journal of neural engineering* 11.5 (2014).

[113] Noémie Jaquier, Mathilde Connan, Claudio Castellini, and Sylvain Calinon. “Combining electromyography and tactile myography to improve hand and wrist activity detection in prostheses”. In: *Technologies* 5.4 (2017), p. 64.

[114] N. Jiang, I. Vujaklija, H. Rehbaum, B. Graimann, and D. Farina. “Is Accurate Mapping of EMG Signals on Kinematics Needed for Precise Online Myoelectric Control?” In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 22.3 (May 2014), pp. 549–558.

[115] Ning Jiang, Silvia Muceli, Bernhard Graimann, and Dario Farina. “Effect of arm position on the prediction of kinematics from EMG in amputees”. In: *Medical & Biological Engineering & Computing* 51.1 (2013), pp. 143–151.

[116] Roland S. Johansson, Göran Westling, Anders Bäckström, and J. Randall Flanagan. “Eye–hand coordination in object manipulation”. In: *Journal of Neuroscience* 21.17 (2001), pp. 6917–6932.

[117] Keith S. Karn. ““Saccade pickers” vs.“fixation pickers”: the effect of eye tracking instrumentation on research”. In: *Proceedings of the 2000 symposium on Eye tracking research & applications*. ACM. 2000, pp. 87–88.

[118] Shashikala Kattla and Madeleine M. Lowery. “Fatigue related changes in electromyographic coherence between synergistic hand muscles”. In: *Experimental brain research* 202.1 (2010), pp. 89–99.

[119] Rami N Khushaba, Ali Al-Timemy, Sarath Kodagoda, and Kianoush Nazarpour. “Combined influence of forearm orientation and muscular contraction on EMG pattern recognition”. In: *Expert Systems with Applications* 61 (2016), pp. 154–161.

[120] Rami N. Khushaba, Maen Takruri, Jaime Valls Miro, and Sarath Kodagoda. “Towards limb position invariant myoelectric pattern recognition using time-dependent spectral features”. In: *Neural Networks* 55 (2014), pp. 42–58.

[121] Thomas Kinsman, Karen Evans, Glenn Sweeney, Tommy Keane, and Jeff Pelz. "Ego-motion compensation improves fixation detection in wearable eye tracking". In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM. 2012, pp. 221–224.

[122] Đ Klisić, Miloš Kostić, Strahinja Došen, and Dejan B. Popović. "Control of prehension for the transradial prosthesis: natural-like image recognition system". In: *Journal of Automatic Control* 19.1 (2009), pp. 27–31.

[123] Oleg V. Komogortsev, Denise V. Gobert, Sampath Jayarathna, Do Hyong Koh, and Sandeep M. Gowda. "Standardization of automated analyses of oculomotor fixation and saccadic behaviors". In: *IEEE Transactions on Biomedical Engineering* 57.11 (2010), pp. 2635–2645.

[124] Matej Kristan, Ales Leonardis, Jiri Matas, Michael Felsberg, Roman Pflugfelder, Luka Cehovin Zajc, Tomas Vojir, Goutam Bhat, Alan Lukezic, Abdelrahman Eldesokey, et al. "The sixth visual object tracking vot2018 challenge results". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.

[125] Ilja Kuzborskij, Arjan Gijsberts, and Barbara Caputo. "On the Challenge of Classifying 52 Hand Movements from Surface Electromyography". In: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Aug. 2012, pp. 4931–4937.

[126] Michael Land, Neil Mennie, and Jennifer Rusted. "The roles of vision and eye movements in the control of activities of daily living". In: *Perception* 28.11 (1999), pp. 1311–1328.

[127] Michael F. Land. "Eye movements and the control of actions in everyday life". In: *Progress in retinal and eye research* 25.3 (2006), pp. 296–324.

[128] Michael F. Land and Mary Hayhoe. "In what ways do eye movements contribute to everyday activities?" In: *Vision research* 41.25-26 (2001), pp. 3559–3565.

[129] Michael F. Land and David N. Lee. "Where we look when we steer". In: *Nature* 369.6483 (1994), p. 742.

[130] Michael F. Land and Peter McLeod. "From eye movements to actions: how batsmen hit the ball". In: *Nature neuroscience* 3.12 (2000), p. 1340.

[131] Markus Lappe and Klaus-Peter Hoffmann. "Optic flow and eye movements". In: *International review of neurobiology* (2000), pp. 29–50.

[132] Linnéa Larsson, Andrea Schwaller, Kenneth Holmqvist, Marcus Nyström, and Martin Stridh. "Compensation of head movements in mobile eye-tracking data using an inertial measurement unit". In: *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. ACM. 2014, pp. 1161–1167.

[133] Hemin Omer Latif, Nasser Sherkat, and Ahmad Lotfi. "Teleoperation through eye gaze (TeleGaze): a multimodal approach". In: *2009 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE. 2009, pp. 711–716.

[134] Ewen B. Lavoie, Aïda M. Valevicius, Quinn A. Boser, Ognjen Kovic, Albert H. Vette, Patrick M. Pilarski, Jacqueline S. Hebert, and Craig S. Chapman. "Using synchronized eye and motion tracking to determine high-precision eye-movement patterns during object-interaction tasks". In: *Journal of vision* 18.6 (2018), p. 18.

[135] Ronald S. LeFever and Carlo J. De Luca. "A procedure for decomposing the myoelectric signal into its constituent action potentials-part I: technique, theory, and implementation". In: *IEEE transactions on biomedical engineering* 3 (1982), pp. 149–157.

[136] Ian Lenz, Honglak Lee, and Ashutosh Saxena. "Deep learning for detecting robotic grasps". In: *The International Journal of Robotics Research* 34.4-5 (2015), pp. 705–724.

[137] Guanglin Li, Aimee E. Schultz, and Todd A. Kuiken. "Quantifying Pattern Recognition-Based Myoelectric Control of Multifunctional Transradial Prostheses". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 18.2 (2010), pp. 185–192.

[138] Colin M. Light, Paul H. Chappell, and Peter J. Kyberd. "Establishing a standardized clinical assessment tool of pathologic and prosthetic hand function: normative data, reliability, and validity". In: *Archives of physical medicine and rehabilitation* 83.6 (2002), pp. 776–783.

[139] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. "Feature pyramid networks for object detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 2117–2125.

[140] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. "Microsoft coco: Common objects in context". In: *European conference on computer vision*. Springer. 2014, pp. 740–755.

[141] Jianwei Liu, Dingguo Zhang, Xinjun Sheng, and Xiangyang Zhu. "Quantification and solutions of arm movements effect on sEMG pattern recognition". In: *Biomedical Signal Processing and Control* 13 (2014), pp. 189 –197.

[142] B. A. Lock, K. Englehart, and B. Hudgins. "Real-time myoelectric control in a virtual environment to relate usability vs. accuracy". In: *MyoElectric Controls/Powered Prosthetics Symposium (MEC)*. 2005.

[143] Quentin Lohmeyer, Mirko Meboldt, and Sven Matthiesen. "Analyzing Visual Strategies of Novice and Experienced Designers by Eye Tracking Application". In: Sept. 2013.

[144] Jonathan Long, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 3431–3440.

[145] Bartjan Maat, Gerwin Smit, Dick Plettenburg, and Paul Breedveld. "Passive prosthetic hands and tools: A literature review". In: *Prosthetics and orthotics international* 42.1 (2018), pp. 66–74.

[146] D. G. K. Madusanka, L. N. S. Wijayasingha, R. A. R. C. Gopura, Y. W. R Amarasinghe, and G. K. I. Mann. "A review on hybrid myoelectric control systems for upper limb prosthesis". In: *2015 Moratuwa Engineering Research Conference (MERCon)*. IEEE. 2015, pp. 136–141.

[147] Paul D. Marasco, Jacqueline S. Hebert, Jon W. Sensinger, Courtney E. Shell, Jonathon S. Schofield, Zachary C. Thumser, Raviraj Nataraj, Dylan T. Beckler, Michael R. Dawson, Dan H. Blustein, Satinder Gill, Brett D. Mensh, Rafael Granja-Vazquez, Madeline D. Newcomb, Jason P. Carey, and Beth M. Orzell. "Illusory movement perception improves motor control for prosthetic hands". In: *Science Translational Medicine* 10.432 (2018).

[148] Marko Markovic, Strahinja Dosen, Dejan Popovic, Bernhard Graimann, and Dario Farina. "Sensor fusion and computer vision for context-aware control of a multi degree-of-freedom prosthesis". In: *Journal of neural engineering* 12.6 (2015).

[149] Marko Markovic, Strahinja Dosen, Christian Cipriani, Dejan Popovic, and Dario Farina. "Stereovision and augmented reality for closed-loop control of grasping in hand prostheses". In: *Journal of neural engineering* 11.4 (2014).

[150] Marko Markovic, Meike A. Schweisfurth, Leonard F. Engels, Tashina Bentz, Daniela Wüstefeld, Dario Farina, and Strahinja Dosen. "The clinical relevance of advanced artificial feedback in the control of a multi-functional myoelectric prosthesis". In: *Journal of neuroengineering and rehabilitation* 15.1 (2018), p. 28.

[151] Susana Martinez-Conde, Stephen L. Macknik, and David H. Hubel. "The role of fixational eye movements in visual perception". In: *Nature reviews neuroscience* 5.3 (2004), p. 229.

[152] Francisco Massa and Ross Girshick. *maskrcnn-benchmark: Fast, modular reference implementation of Instance Segmentation and Object Detection algorithms in PyTorch*. `https://github.com/facebookresearch/maskrcnn-benchmark`. 2018.

[153] Matthew R. Masters, Ryan J. Smith, Alcimar B. Soares, and Nitish V. Thakor. "Towards better understanding and reducing the effect of limb position on myoelectric upper-limb prostheses". In: *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. 2014, pp. 2577–2580.

[154] Enzo Mastinu, Johan Ahlberg, Eva Lendaro, Liselotte Hermansson, Bo Håkansson, and Max Ortiz-Catalan. "An alternative myoelectric pattern recognition approach for the control of hand prostheses: A case study of use in daily life by a dysmelia subject". In: *IEEE journal of translational engineering in health and medicine* 6 (2018), pp. 1–12.

[155] Leonard Matin and Douglas Grant Pearce. "Three-dimensional recording of rotational eye movements by a new contact-lens technique." In: *Biomedical sciences instrumentation* 2 (1964), pp. 79–95.

[156] Jess McIntosh, Asier Marzo, Mike Fraser, and Carol Phillips. "Echoflex: Hand gesture recognition using ultrasound imaging". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM. 2017, pp. 1923–1934.

[157] David P. McMullen, Guy Hotson, Kapil D. Katyal, Brock A. Wester, Matthew S. Fifer, Timothy G. McGee, Andrew Harris, Matthew S. Johannes, R. Jacob Vogelstein, Alan D. Ravitz, William S. Anderson, Nitish V. Thakor, and Nathan E. Crone. "Demonstration of a semi-autonomous hybrid brain–machine interface using human intracranial EEG, eye tracking, and computer vision to control a robotic upper limb prosthetic". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 22.4 (2013), pp. 784–796.

[158] James Mercer. "Functions of positive and negative type, and their connection the theory of integral equations". In: *Philosophical transactions of the royal society of London. Series A, containing papers of a mathematical or physical character* 209.441-458 (1909), pp. 415–446.

[159] Roberto Merletti. "Standards for Reporting EMG data". In: *Journal of Electromyography and Kinesiology* 9.1 (1999).

[160] Roberto Merletti, Philip A. Parker, and Philip J. Parker. *Electromyography: physiology, engineering, and non-invasive applications*. Vol. 11. John Wiley & Sons, 2004.

[161] Silvestro Micera, Jacopo Carpaneto, and Stanisa Raspopovic. "Control of hand prostheses using peripheral information". In: *IEEE reviews in biomedical engineering* 3 (2010), pp. 48–68.

[162] Ajay Mishra, Yiannis Aloimonos, Loong-Fah Cheong, and Ashraf Kassim. "Active Visual Segmentation". In: *IEEE transactions on pattern analysis and machine intelligence* 34 (Apr. 2012), pp. 639–653.

[163] Thomas M. Mitchell. *Machine Learning*. 1st ed. New York, NY, USA: McGraw-Hill, Inc., 1997.

[164] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of machine learning*. MIT press, 2018.

[165] P. Morasso, E. Bizzi, and J. Dichgans. "Adjustment of saccade characteristics during head movements". In: *Experimental Brain Research* 16.5 (1973), pp. 492–500.

[166] Carlos H. Morimoto and Marcio R. M. Mimica. "Eye gaze tracking techniques for interactive applications". In: *Computer vision and image understanding* 98.1 (2005), pp. 4–24.

[167] Carlos Hitoshi Morimoto, Dave Koons, Arnon Amir, and Myron Flickner. "Pupil detection and tracking using multiple light sources". In: *Image and vision computing* 18.4 (2000), pp. 331–335.

[168] O. H. Mowrer, Theodore C. Ruch, and N. E. Miller. "The corneo-retinal potential difference as the basis of the galvanometric method of recording eye movements". In: *American Journal of Physiology-Legacy Content* 114.2 (1935), pp. 423–428.

[169] Maria Niedernhuber, Damiano G. Barone, and Bigna Lenggenhager. "Prostheses as extensions of the body: Progress and challenges". In: *Neuroscience & Biobehavioral Reviews* 92 (2018), pp. 1–6.

[170] Christian Nissler, Mathilde Connan, Markus Nowak, and Claudio Castellini. "Online tactile myography for simultaneous and proportional hand and wrist myocontrol". In: *Proceedings of the Myoelectric Control Symposium (MEC), Fredericton, NB, Canada*. 2017, pp. 15–18.

[171] Domen Novak and Robert Riener. "Enhancing patient freedom in rehabilitation robotics using gaze-based intention detection". In: *2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR)*. IEEE. 2013, pp. 1–6.

[172] Marcus Nyström and Kenneth Holmqvist. "An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data". In: *Behavior research methods* 42.1 (2010), pp. 188–204.

[173] Anneli Olsen. "The Tobii I-VT fixation filter". In: *Tobii Technology* (2012).

[174] Max Jair Ortiz-Catalan, Rickard Brånemark, and Bo Håkansson. "BioPatRec: A modular research platform for the control of artificial limbs based on pattern recognition algorithms". In: *Source Code for Biology and Medicine* 8.1 (Jan. 2013).

[175] Oskar Palinko, Francesco Rea, Giulio Sandini, and Alessandra Sciutti. "Robot reading human gaze: Why eye tracking is better than head tracking for human-robot collaboration". In: *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2016, pp. 5048–5054.

[176] Johnny Vic Vincent Parr, Samuel J. Vine, Mark R. Wilson, Neil R. Harrison, and Greg Wood. "Visual attention, EEG alpha power and T7-Fz connectivity are implicated in prosthetic hand control and can be optimized through gaze training". In: *Journal of neuroengineering and rehabilitation* 16.1 (2019), p. 52.

[177] JVV Parr, Samuel J Vine, NR Harrison, and Greg Wood. "Examining the spatiotemporal disruption to gaze when using a myoelectric prosthetic hand". In: *Journal of motor behavior* 50.4 (2018), pp. 416–425.

[178] Gauravkumar K Patel, Markus Nowak, and Claudio Castellini. "Exploiting knowledge composition to improve real-life hand prosthetic control". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25.7 (2017), pp. 967–975.

[179] Aftab E. Patla and Joan N. Vickers. "How far ahead do we look when required to step on specific locations in the travel path during locomotion?" In: *Experimental brain research* 148.1 (2003), pp. 133–138.

[180] Bart Peerdeman, Daphne Boere, Heidi Witteveen, Rianne Huis in 't Veld, Hermie Hermens, Stefano Stramigioli, Hans Rietman, Peter Veltink, and Sarthak Misra. "Myoelectric forearm prostheses: State of the art from a user-centered perspective". In: *Journal of Rehabilitation Research and Development* 48.6 (2011), pp. 719–738.

[181] Jeff Pelz, Mary Hayhoe, and Russ Loeber. "The coordination of eye, head, and hand movements in a natural task". In: *Experimental brain research* 139.3 (2001), pp. 266–277.

[182] L. Peng, Z. Hou, Y. Chen, W. Wang, L. Tong, and P. Li. "Combined use of sEMG and accelerometer in hand motion classification considering forearm rotation". In: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. July 2013, pp. 4227–4230.

[183] Sam L. Phillips and William Craelius. "Residual kinetic imaging: a versatile interface for prosthetic control". In: *Robotica* 23.3 (2005), pp. 277–282.

[184] Angkoon Phinyomark, Pornchai Phukpattaranont, and Chusak Limsakul. "Feature reduction and selection for EMG signal classification". In: *Expert systems with applications* 39.8 (2012), pp. 7420–7431.

[185] Pedro O. Pinheiro, Tsung-Yi Lin, Ronan Collobert, and Piotr Dollár. "Learning to refine object segments". In: *European Conference on Computer Vision*. Springer. 2016, pp. 75–91.

[186] Alex Poole and Linden J. Ball. "Eye tracking in HCI and usability research". In: *Encyclopedia of human computer interaction*. IGI Global, 2006, pp. 211–219.

[187] Roy M. Pritchard. "Stabilized images on the retina". In: *Scientific American* 204.6 (1961), pp. 72–79.

[188] A. Radmand, Erik Scheme, and K. Englehart. "A characterization of the effect of limb position on EMG features to guide the development of effective prosthetic control schemes". In: *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. 2014, pp. 662–667.

[189] Ashkan Radmand, Erik Scheme, and Kevin Englehart. "High-density force myography: A possible alternative for upper-limb prosthetic control." In: *Journal of Rehabilitation Research & Development* 53.4 (2016).

[190] Estelle Raffin, Jeremie Mattout, Karen T Reilly, and Pascal Giraux. "Disentangling motor execution from motor imagery with the phantom limb". In: *Brain* 135.2 (2012), pp. 582–595.

[191] Estelle Raffin, Pascal Giraux, and Karen T Reilly. "The moving phantom: motor execution or motor imagery?" In: *Cortex* 48.6 (2012), pp. 746–757.

[192] Eyal M. Reingold. "Eye tracking research and technology: Towards objective measurement of data quality". In: *Visual cognition* 22.3-4 (2014), pp. 635–652.

[193] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. "Faster r-cnn: Towards real-time object detection with region proposal networks". In: *Advances in neural information processing systems*. 2015, pp. 91–99.

[194] Linda Resnik, Marissa R. Meucci, Shana Lieberman-Klinger, Christopher Fantini, Debra L. Kelty, Roxanne Disla, and Nicole Sasson. "Advanced upper limb prosthetic devices: implications for upper limb prosthetic rehabilitation". In: *Archives of physical medicine and rehabilitation* 93.4 (2012), pp. 710–717.

[195] Linda Resnik, He Helen Huang, Anna Winslow, Dustin L. Crouch, Fan Zhang, and Nancy Wolk. "Evaluation of EMG pattern recognition for upper limb prosthesis control: a case study in comparison with direct myoelectric control". In: *Journal of neuroengineering and rehabilitation* 15.1 (2018), p. 23.

[196] Giacomo Rizzolatti and Giuseppe Luppino. "The cortical motor system". In: *Neuron* 31.6 (2001), pp. 889–901.

[197] David A. Robinson. "A method of measuring eye movemnent using a scieral search coil in a magnetic field". In: *IEEE Transactions on bio-medical electronics* 10.4 (1963), pp. 137–145.

[198] Aidan D. Roche, Hubertus Rehbaum, Dario Farina, and Oskar C. Aszmann. "Prosthetic myoelectric control strategies: a clinical perspective". In: *Current Surgery Reports* 2.3 (2014), p. 44.

[199] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. "Learning Internal Representations by Error Propagation". In: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*. Cambridge, MA: MIT Press, 1986, pp. 318–362.

[200] Dario D. Salvucci and Joseph H. Goldberg. "Identifying Fixations and Saccades in Eye-tracking Protocols". In: *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*. ACM, 2000, pp. 71–78.

[201] Dario D. Salvucci and John R. Anderson. "Tracing eye movement protocols with cognitive process models". In: (1998).

[202] Akanksha Saran, Srinjoy Majumdar, Elaine Schaertl Shor, Andrea Thomaz, and Scott Niekum. "Human gaze following for human-robot interaction". In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2018, pp. 8615–8621.

[203] Alessandra Sciutti, Ambra Bisio, Francesco Nori, Giorgio Metta, Luciano Fadiga, and Giulio Sandini. "Robots can be perceived as goal-oriented agents". In: *Interaction Studies* 14.3 (2013), pp. 329–350.

[204] Fredrik C. P. Sebelius, Birgitta N. Rosen, and Göran N. Lundborg. "Refined myoelectric control in below-elbow amputees using artificial neural networks and a data glove". In: *Journal of Hand Surgery* 30.4 (2005), pp. 780–789.

[205] Shai Shalev-Shwartz and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

[206] John Shawe-Taylor and Nello Cristianini. *Kernel Methods for Pattern Analysis*. New York, NY, USA: Cambridge University Press, 2004.

[207] David Sierra González and Claudio Castellini. "A realistic implementation of ultrasound imaging as a human-machine interface for upper-limb amputees". In: *Frontiers in neurorobotics* 7 (2013), p. 17.

[208] J. Silva, T. Chau, and A. Goldenberg. "MMG-based multisensor data fusion for prosthesis control". In: *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Vol. 3. IEEE. 2003, pp. 2909–2912.

[209] J. Silva, T. Chau, S. Naumann, and W. Heim. "Systematic characterisation of silicon-embedded accelerometers for mechanomyography". In: *Medical and Biological Engineering and Computing* 41.3 (2003), pp. 290–295.

[210] A. M. Simon, L. J. Hargrove, B. A. Lock, and T. A. Kuiken. "Target achievement control test: Evaluating real-time myoelectric pattern-recognition control of multifunctional upper-limb prostheses". In: *Journal of Rehabilitation Research and Development* 48.6 (2011), pp. 619–628.

[211] Ann M. Simon, Kristi L. Turner, Laura A. Miller, Levi J. Hargrove, and Todd A. Kuiken. "Pattern recognition and direct control home use of a multiarticulating hand prosthesis". In: *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)*. IEEE. 2019, pp. 386–391.

[212] Jeroen B. J. Smeets, Mary M. Hayhoe, and Dana H. Ballard. "Goal-directed arm movements change eye-head coordination". In: *Experimental brain research* 109.3 (1996), pp. 434–440.

[213] Lauren H. Smith, Levi J. Hargrove, Blair A. Lock, and Todd A. Kuiken. "Determining the optimal window length for pattern recognition-based myoelectric control: balancing the competing effects of classification error and controller delay". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 19.2 (2010), pp. 186–192.

[214] Richard S. Snell and Michael A. Lemp. *Clinical anatomy of the eye*. John Wiley & Sons, 2013.

[215] Mohammad M. D. Sobuh, Laurence P. J. Kenney, Adam J. Galpin, Sibylle B. Thies, Jane McLaughlin, Jai Kulkarni, and Peter Kyberd. "Visuomotor behaviours when using a myoelectric prosthesis". In: *Journal of neuroengineering and rehabilitation* 11.1 (2014), p. 72.

[216] Antonietta Stango, Francesco Negro, and Dario Farina. "Spatial correlation of high density EMG signals provides features robust to electrode number and shift in pattern recognition for myocontrol". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 23.2 (2014), pp. 189–198.

[217] Ilaria Strazzulla, Markus Nowak, Marco Controzzi, Christian Cipriani, and Claudio Castellini. "Online bimanual manipulation using surface electromyography and incremental learning". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25.3 (2016), pp. 227–234.

[218] Benjamin W Tatler. "Eye movements from laboratory to life". In: *Current trends in eye tracking research*. Springer, 2014, pp. 17–35.

[219] Benjamin W Tatler, Mary M Hayhoe, Michael F Land, and Dana H Ballard. "Eye guidance in natural vision: Reinterpreting salience". In: *Journal of vision* 11.5 (2011), pp. 5–5.

[220] Luke T. Taverne, Matteo Cognolato, Tobias Bützer, Roger Gassert, and Otmar Hilliges. "Video-based Prediction of Hand-grasp Preshaping with Application to Prosthesis Control". In: *2019 International Conference on Robotics and Automation (ICRA)*. IEEE. 2019, pp. 4975–4982.

[221] E. Llewellyn Thomas and Eugene Stasiak. "Eye Movements and Body Images". In: *Canadian Psychiatric Association Journal* 9.4 (1964), pp. 336–344.

[222] Tobii AB. *Tobii Pro Glasses 2 API, Developer's Guide*. Version 1.23. Tobii AB. 2017. October, 2017.

[223] Tatiana Tommasi, Francesco Orabona, and Barbara Caputo. "Discriminative cue integration for medical image annotation". In: *Pattern Recognition Letters* 29.15 (2008), pp. 1996–2002.

[224] Yi-Hsuan Tsai, Ming-Hsuan Yang, and Michael J Black. "Video segmentation via object flow". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 3899–3908.

[225] Marco D. Tundo, Edward Lemaire, and Natalie Baddour. "Correcting Smartphone orientation for accelerometer-based analysis". In: *2013 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. IEEE. 2013, pp. 58–62.

[226] Jack Valmadre, Luca Bertinetto, Joao F. Henriques, Ran Tao, Andrea Vedaldi, Arnold W. M. Smeulders, Philip H. S. Torr, and Efstratios Gavves. "Long-term tracking in the wild: A benchmark". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, pp. 670–685.

[227] Drew Van Der Riet, Riaan Stopforth, Glen Bright, and Olaf Diegel. "An overview and comparison of upper limb prosthetics". In: *2013 Africon*. IEEE. 2013, pp. 1–8.

[228] Dimitris Voudouris, Jeroen B. J. Smeets, Katja Fiehler, and Eli Brenner. "Gaze when reaching to grasp a glass". In: *Journal of vision* 18.8 (2018), pp. 16–16.

[229] Ivan Vujaklija, Aidan D. Roche, Timothy Hasenoehrl, Agnes Sturma, Sebastian Amsuess, Dario Farina, and Oskar C. Aszmann. "Translating Research on Myoelectric Control into Clinics-Are the Performance Assessment Methods Adequate?" In: *Frontiers in neurorobotics* 11 (2017), p. 7.

[230] Qiang Wang, Li Zhang, Luca Bertinetto, Weiming Hu, and Philip H. S. Torr. "Fast online object tracking and segmentation: A unifying approach". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2019.

[231] Longyin Wen, Dawei Du, Zhen Lei, Stan Z. Li, and Ming-Hsuan Yang. "Jots: Joint online tracking and segmentation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015, pp. 2226–2234.

[232] Samuel Wilson and Ravi Vaidyanathan. "Upper-limb prosthetic control using wearable multichannel mechanomyography". In: *2017 International Conference on Rehabilitation Robotics (ICORR)*. IEEE. 2017, pp. 1293–1298.

[233] Michael Wininger, Nam-Hun Kim, and William Craelius. "Pressure signature of forearm as predictor of grip force." In: *Journal of Rehabilitation Research & Development* 45.6 (2008).

[234] D. A. Winter, Andrew J. Fuglevand, and S. E. Archer. "Crosstalk in surface electromyography: theoretical and practical estimates". In: *Journal of Electromyography and Kinesiology* 4.1 (1994), pp. 15–26.

[235] Roy W. Wirta, Donald R. Taylor, and F. Ray Finley. "Pattern-recognition arm prosthesis: a historical perspective-a final report". In: *Bulletin of Prosthetics Research* 10.30 (1978), pp. 8–35.

[236] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang. "Online object tracking: A benchmark". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2013, pp. 2411–2418.

[237] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. "Aggregated residual transformations for deep neural networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1492–1500.

[238] Michele Xiloyannis, Constantinos Gavriel, Andreas A. C. Thomik, and A. Aldo Faisal. "Dynamic forward prediction for prosthetic hand control by integration of EMG, MMG and kinematic signals". In: *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE. 2015, pp. 611–614.

[239] Linjie Yang, Yanran Wang, Xuehan Xiong, Jianchao Yang, and Aggelos K. Katsaggelos. "Efficient video object segmentation via network modulation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 6499–6507.

[240] Alfred L. Yarbus. *Eye movements and vision*. Springer, 2013.

[241] Aaron J. Young, Levi J. Hargrove, and Todd A. Kuiken. "The effects of electrode size and orientation on the sensitivity of myoelectric pattern recognition systems to electrode shift". In: *IEEE Transactions on Biomedical Engineering* 58.9 (2011), pp. 2537–2544.

[242] Laurence R. Young and David Sheena. "Survey of eye movement recording methods". In: *Behavior research methods & instrumentation* 7.5 (1975), pp. 397–429.

[243] Mahyar Zardoshti-Kermani, Bruce C. Wheeler, Kambiz Badie, and Reza M. Hashemi. "EMG feature evaluation for movement control of upper extremity prostheses". In: *IEEE Transactions on Rehabilitation Engineering* 3.4 (1995), pp. 324–333.

[244] Micera Zecca, Silvestro Micera, Maria C. Carrozza, and Paolo Dario. "Control of multifunctional prosthetic hands by processing the electromyographic signal". In: *Critical Reviews in Biomedical Engineering* 30.4-6 (2002).

[245] Yong Zeng, Zhengyi Yang, Wei Cao, and Chunming Xia. "Hand-motion patterns recognition based on mechanomyographic signal analysis". In: *2009 International Conference on Future BioMedical Information Engineering (FBIE)*. IEEE. 2009, pp. 21–24.

[246] Guangquan Zhou and Yong-ping Zheng. “Human motion analysis with ultrasound and sonomyography”. In: *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. 2012, pp. 6479–6482.