Author:
McIntosh, Jess

Title:
Exploring the practicality of wearable gesture recognition

Author:
McIntosh, Jess

Title:
Exploring the practicality of wearable gesture recognition

Exploring the practicality of wearable gesture recognition.

By

JESS MCINTOSH

Department of Computer Science
UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Engineering.

JUNE 2018

Word count: 38139

Abstract

For millenia, gestures have been used by humans as a method of communication. This primitive form of communication is still used by us situationally to either to facilitate, or replace verbal communication. For instance, gestures can indicate spatial information such as direction, or could be used when speaking would disrupt the vicinity.

Thus the field of human-computer interaction has identified that gestures could be an important tool for natural interaction with computers. Since then, a variety of different uses for gestural interaction have been created. It has been envisioned that such an interaction modality could provide a convenient way of interacting with devices near to the person.

It has been suggested that hand gesture control could be useful for wearable devices, in situations where interacting with these devices are difficult with other known modalities. Currently, smart watches and other wrist-worn devices are the dominant form factor for wearable devices, probably due to the watch being a device that is seen as being socially acceptable to wear. Since the watch form factor is currently so dominant, it is intuitive to embed the sensing technology for hand gesture recognition at this location. This is a key problem that is often ignored by most research conducted in hand gesture recognition.

This thesis analyses current wearable techniques for hand gesture recognition, paying particular attention to the practicalities of the techniques which are important for integration with a wrist-worn form factor device. Experimentation is conducted to improve existing techniques, attempting to address these practicality issues, with a focus on restricting the placement and size of the device to conform to the wrist. Further ergonomic and practical issues are uncovered through experimentation in EMG and ultrasonography, leading to a technical innovation to extend the capability of wearable infrared gesture sensing. The initial requirements for the ideal wearable gesture recogniser is revised after insights from each chapter. The final revised requirements and analysis of methods lets us advise that the only reasonably practical method that can be implemented at the moment is infrared. Finally, the analysis of methods also let us give insight on the most promising technologies (such as ultrasonography) and the main problems that hinder their practicality. One of the key problems that we find is that tightness of sensors against the skin are a big practicality concern, which is almost always a factor that is ignored in current research. Finally, we discuss additional robustness and cross-participant issues that remain challenging in all current techniques.

Dedication and acknowledgements

Throughout my time at the Bristol Interaction Group, I've had the pleasure of meeting many faces.

First and foremost, I owe a great deal to my supervisor, Mike Fraser. Although in sometimes difficult circumstances, he managed to find time and effort into guiding me through my PhD. I will always be forever thankful for the help he has given me.

I have a lot to thank for Sriram, my previous supervisor, who first offered me the opportunity to pursue research. Through Sri, I learnt many virtues that a researcher should have - above all to aim high and persevere through difficult challenges.

It goes without saying that I've enjoyed my time at the lab, mostly due to the great company of those surrounding me. The many tea breaks and lunch excursions were an important part of my time here, for not only social reasons but also in fostering creative discussions.

I would like to thank Gareth for providing me a way of procrastinating with table tennis.

In particular, I've had the great pleasure of spending a lot of time here with Asier Marzo. Asier has always been superbly supportive, a good colleague and an even greater friend. His intelligence and research intuition is something that I will always admire and have hopefully learnt from.

I am grateful for the funding from the EPSRC for supporting the research for my PhD.

A special thanks goes to Carol Phillips, Sian Curtis, Margaret Saunders of the BIRCH group within the University Hospitals Bristol NHS Foundation Trust. Without their help and knowledge, the ultrasonography project could not have been accomplished.

Last but most certainly not least, I've been privileged with having the most loving and caring of parents. Without them, I would not have been given this chance. I'm also equally thankful to my long time companion, Ingrid, for sticking by my side through occasionally stressful times in this arduous journey.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ....[redacted]............. DATE: ....28/07/2018...........

Table of Contents

Page

List of Tables

List of Figures

Figure **Page**

Chapter 1

Introduction

1.1 Vision

In recent decades, gesture interaction has been an important topic of research in human-computer interaction (HCI). It has already transformed the way we interact with devices drastically. One notable example of gestures being widely used are in the use of touch gestures in smart phones, e.g. pinch to zoom or swiping.

Many suggest that in the future, hand gestures will also play a key role in computer interaction [101]. These suggestions are based on the same principles as touch interaction on 2D graphical user interfaces, that they are intuitive to us and simple to perform. More of this is explained in the next section.

A number of different applications have been envisioned and studied, ranging from explicit to implicit use. These applications require varying hand tracking accuracy in order to enable. In section 1.3, we discuss the applications and derive from them that mobility is a key requirement for many of them.

This requirement of mobility is what creates a challenge in enabling gesture detection. There have been some research efforts in HCI that improve the mobile gesture recognition. However, as we describe in the forthcoming sections, many of the requirements are still not satisfied. Quite prominently, as a result of making systems mobile, the accuracy is often diminished. There is still a large gap that sits between what is currently possible with stationary equipment for hand tracking (for instance, with depth cameras), versus mobile hand tracking technology. The ultimate aim is to match the accuracy of a stationary device in a wearable form factor device.

In this thesis, we are mostly concerned with the development of better mobile devices for recognising gestures. In particular, we focus on the form factor of a wrist-band type device, as is commonly used by a variety of worn objects such as watches and bracelets. We look to improve specific aspects that are currently lacking in gesture recognition technology. After reviewing past work, we identify and target areas for improvement such as placement and size, accuracy and complexity.

1.2 Gestures

Gestures enable a channel of communication for humans which can be used to convey messages to others. Gestures can be performed using different body parts, but hand gestures are the most commonly used due to the dexterity of human hands. Gestures can be used to convey a variety of meanings.

Here, we use the taxonomy of gestures as described by M.Karam and M.C.Schraefel's work [62], which is tailored towards the use of gestures as an interaction method. These different categories are not exclusive and many of them can be applied to a particular gesture. Here, we describe them.

1.2.1 Deictic

Perhaps the simplest of gestures are pointing to objects and other spatial inferences, which are called deictic gestures. Much work has been done on pointing with the hand to locate a point on a display [13], or more recently in augmented and virtual reality.

1.2.2 Manipulative

By definition, manipulative gestures are those:

"whose intended purpose is to control some entity by applying a tight relationship between the actual movements of the gesturing hand/arm with the entity being manipulated."

For mid-air hand gestures, this is fairly limited as there is no object with we can gesture. In this work, we can imagine various user interface widgets like sliders or a menu being manipulated directly using gestures of the hands. For instance, opening an closing the hand can map directly to the intensity of the light in the room.

1.2.3 Semaphoric

Semaphoric gestures represent a set of gestures that serve as symbols to communicate to the machine. That has also been named as symbolic gestures. More traditional definitions of gestures would likely fit within this category.

Representational gestures may mimic aspects of an action or object, such as climbing. They may also be metaphorical, like using ones hands to mimic pan scales as though weighing two sides of an argument. These gestures convey a more abstract meaning than manipulative or deictic gestures.

Conventional gestures convey culturally specific meaning, for instance a "thumbs up" which represents approval in most cultures, but in some cultures may be considered obscene [8].

1.2.4 Gesticulation

Gestures are also tightly integrated with speech, when somebody gestures whilst talking, it is called gesticulation. Not only does gesticulation help the listener in a communicative manner, but also to the speaker in a cognitive way [17]. For instance, one study found that speakers who gesture are more fluent, producing fewer verbal errors and hesitations, than when not allowed to gesture[48]. There are numerous other indications that gestures are an embodiment of our thought processes, mental imagery and feelings [58, 103].

1.2.5 Language

Linguistically based languages are independent of most other types of gestures. It is also less intuitive and more comparable to speech [62]. Although some elements such as an alphabet made by finger spelling, can be considered semaphoric.

1.2.6 Static and Dynamic Gestures

These are not types of gestures that fit within the taxonomy but are separate descriptions that tell us about whether the hand of the person is moving, or whether it is static. Usually static gestures have a hand pose to convey meaning. There may also be combinations of a static gesture that that moves dynamically. In this thesis, we are mostly concerned with the detection of static gestures (hand poses) as dynamic gestures can be detected with the use of an IMU. However, here we disambiguate dynamic gestures with gestures which change the hand pose continuously, which we call continuous gestures.

1.2.7 Gestures for Interaction with Computers

These suggestions that gestures are instinctive to humans and used as a tool to convey our thoughts, have led many to believe that gestures are an ideal form of interaction with computers [101]. Not only do gestures have a low cognitive load as they are performed instinctively, but they are also easy for us to interpret, and their simplicity and intuitiveness makes them easy to learn. This modality of input could be seen as communicating with the computer in a conversational manner. Gestures for touch-screen interaction in today's technology is ubiquitous. Perhaps the most iconic of them, the "pinch to zoom" gesture. This gives some reassurance that gestural interaction can be successfully applied. But then why is mid-air hand gesture interaction is not yet prevalent? The next sections in this chapter will try to address this question, by firstly analysing where hand gesture interaction could be applied, from which a requirement for a gesture detection device is created. Through analysing the related work, there are many requirements which are not met, which is an indication that the technology is not yet suitable for use today.

1.3 Where Can Hand Gesture Interaction be Applied?

1.3.1 Explicit Gesture Interaction

Wearable devices such as smartwatches have recently become a topic of interest among researchers and industry. Consumer products today mainly use their smartwatches as an easy-to-reach proxy for their smartphone, for viewing notifications and performing simple tasks such as answering calls. Their use remains somewhat limited, partially due to the difficulty in interacting with this device form factor. For instance, the small display size on watch faces makes touch interaction awkward due to the comparatively large size of our fingers (the 'fat finger' problem) [114, 115]. In addition, throughout the day there may be tasks which encumber the persons hand, such as holding an umbrella or a shopping bag. Previous work has identified that such physical encumbrance has negative impacts on touch gesture performance on smartphones [89]. Kerber et al. extended this notion of encumbrance to smartwatch interaction and suggested using hand gestures instead of touch interaction [64]. They conducted studies to elicit gestures from people that use only the hand/arm the device is worn on. Saponas at el. found ways of enabling always available input on smartwatches by designing gestures that can be performed even by the hand that is busy with other objects e.g., holding a mug [106].

While there are other tangible methods for interacting with a smartwatch [111], these might be difficult to perform without looking at the device. There may be tasks where visual attention is critically important at all times (such as driving or running through busy streets), in which case such tangible interaction methods may be dangerously distracting. So gesture control of wearable devices seem to fill some niche use cases where other interaction modalities may be awkward or distracting.

Full engagement with a device can not only be unsafe, but it may also be frowned upon socially. Casual interactions are particularly appropriate in such scenarios. Pohl and Murray-Smith discuss the different scenarios and technologies that may enable this interaction method, some of which is a major topic of this thesis [99].

However, the applications of gesture interaction extends beyond simply controlling smartwatches. The idea of interconnected household devices, controlled either by the user or by intelligent systems has recently peaked the interest of researchers and consumers. Examples of such "smart home" or "home automation" devices include wireless, connected LED lights [4, 5] or cooking appliances [43, 75]. However, the most common way of controlling devices that are on the market today seem to be through apps on the smartphone. Some research has gone into bespoke tangible interfaces [16]. Though neither of these modalities offer the same convenience or instant control that gestures do. Indeed there is work that suggests using gestures for controlling smart home devices for these reasons [95]. Gesture control of devices in the persons surroundings can be extended beyond the house to the car [102], workplace [54] or perhaps in the foreseeable future, interaction with objects/displays in public places [107].

These applications of hand gestures are examples of explicit gesture interaction. That is, the user performs gestures intentionally to achieve a desired action. There are issues with differentiating between gestures that the user performs with intent, to natural movements from the user e.g., gestures during conversation. Although false positive gesture detection is an important topic, it is out of the scope of this thesis. Many other works are dedicated to tackling this problem, attempting to create gesture sets with low false positive rates [67] or activation methods [57]. Although these natural human movements prove to be problematic for explicit gesture detection, it actually provides an entirely new and different application for gesture detection: implicit interaction.

1.3.2 Implicit Gesture Interaction

Implicit gestures are those that are performed with natural human movement, without intent. We can detect activities that we perform on a daily basis for a variety of purposes. Some uses for activity recognition include [76]:

- Health monitoring - The frequency of activities such as eating, drinking or smoking can often have negative health effects. Implicit gesture detection can aid in the detection of such.
- Assisted living - Can assist doctors in diagnosing certain conditions detect deviations from a typical routine or deterioration of a patient's current physical status.
- Context awareness - Detection of an activity such as a meeting at work (possibly inferred from a handshake) could be used as a contextual cue to silence the phone. [108].

1.3.3 Virtual Reality or Augmented Reality

Full hand tracking of every degree of freedom is likely to be beyond what is required for gestural interaction. Hypothetically if such a system did exist, it should in theory also be able to detect all possible hand poses (being used as gestures). Therefore, such systems are still relevant to this thesis.

Both virtual and augmented reality can benefit from accurate hand tracking which enables the user to interact with the virtual world [26]. So far, computer vision methods have produced very accurate tracking algorithms that work well for stationary settings. However, augmented reality may used out of the home or office, as we are going about our day outside of these known environments where cameras may not be set up. Aside from this, cameras can still suffer from occlusion - from objects or other bodies close by - or even from ones own body.

Gloves embedded with sensors to track hand movement, so-called data gloves, are currently a more feasible method. However, this encumbers natural interaction with our hands for daily use, particularly for the mobile augmented reality scenario. Therefore, there is further motivation to find an alternative method that does not obstruct the hand but is still highly accurate.

1.4 Requirements for a Gesture Detection Device

This section formulates a set of requirements for a gesture detection device. These requirements are derived from the aforementioned applications. We consider the obvious requirements, such as the mobility and accuracy of systems. In thesis however, we thoroughly emphasise the placement and size parameter, for reasons explained below. Finally, we discuss the implications of device complexity, which effects immediate feasibility of using the device in the real world.

1.4.1 Mobility

Certainly for interaction with a smartwatch, the device to detect gestures must also be mobile. Most of the implicit gesture recognition applications also rely on recognising gestures on-the-go. These applications point towards a need for a mobile device that is able to detect hand poses for always available input. This is the most important requirement that must be fulfilled.

1.4.2 Placement and Size

The requirement of mobility creates a challenge in how the device is carried by the user. Mounting the device onto the human, a so-called "wearable", aligns well with the vision of ubiquitous computing and seems the most convenient way to enable always-available computers [78]. Still, the location of where the device is to be mounted and it's size are both tied to the ergonomics, practicality and social acceptability of it. The device should be as small as possible to not encumber the person, and if the form factor can match items worn in today's fashion, this will avoid any conspicuity and satisfy social acceptance. Indeed, a focus of this thesis will be motivated by the aesthetics and social acceptability of device placement. As an example of research that backs up this claim, Pateman et al. do an in-the-wild evaluation with different styles and shapes of wearables that were co-designed by the participants and them. A large number of people in this study pointed out that they would like to have the device integrate with existing, socially acceptable objects in order to make them "invisible" [96].

1.4.3 Accuracy and Range of Gestures

The ultimate aim is to achieve a 100% accurate, full hand tracking system as described for the VR/AR applications. In reality there is most likely some margin of error with the detection of hand poses. However, as mentioned this may not be entirely necessary for simpler gestural interaction. The types of gestures that can be performed is also an important factor. In general the more gestures that can be detected, the better. Of course, it may not best to include all possible gestures into the application, but if a bigger pool of gestures to select from gives better freedom for the application designer. This does not just include the number of gestures detected, but also concerns the different range of human hand motions that be detected. This is explained in more detail in section 2.1.

1.4.4 Complexity

The complexity of the device, which is mostly determined by the energy, electronics and computation required, effect the immediate feasibility of the device. For instance, if the device needs to be charged multiple times a day, using present day battery technology, the device is clearly not yet ready. If the data that the sensors collect, require what is currently considered to be a high-end desktop Graphics Processing Unit (GPU), it will take many more years until wearable computing device are able to process this much data. These factors also effect the cost of the device. From a research perspective, this could be considered less important than the other requirements, as energy and electronics are expected to improve over time. This is especially true for computationally expensive tasks.

1.4.5 Baseline Requirements

These four points are the basis for our initial set of requirements for a wearable gesture recognition device:

- Mobility → Wearable.
- Placement & Size → Ergonomics, practicality and social acceptability.
- Accuracy & Range of Gestures → Usefulness.
- Complexity → Immediate feasibility.

1.5 Research Questions and Thesis Contributions

As briefly mentioned at the beginning of the chapter, part of this thesis is mainly driven by the task of getting better accuracy of gesture detection systems worn on the wrist. This satisfies the main criteria we identify in the above requirements and to enable practical usage by ensuring

Technique	Mobility	Placement & Size	Fidelity	Complexity
EMG	Battery concerns	**Variable**	Good	Variable
Ultrasound	Improving but still poor	**Variable**	Very good	Very complex
Infrared	Highly mobile	Thin & wrist	Poor	Very simple
Pressure	Highly mobile	Thin & wrist	Poor	Very simple
EIT	Highly mobile	Thin & wrist	Average	Fairly complex
Optical	Can be mobile	Wrist, but obtrusive	Limited	Fairly complex
Glove	Highly mobile	Obtrusive	Excellent	Simple

Table 1.1: A summary of the analysis of established gesture recognition techniques, with respect to the initial requirements.

a high enough accuracy. We selected the wrist as an ideal location due to the fact that it is a socially acceptable place on the body where people where objects. Technology has already employed wristbands and smartwatches to take advantage of this, currently there is a growing trend towards to the use of these devices. Therefore, it seems appropriate to design hardware that can fit within this form factor. As we will describe in the related work, there is already plenty of research that has done this. However, what has been demonstrated so far is that there is a significant trade-off of accuracy for wearability. Skipping ahead slightly, we show the table that arises from our analysis of the related work (Table 1.1. We will explain this later in the related work, but this is now shown to demonstrate the range of different technologies that have been studied, the variability of them, and the trade-offs between form factors.

From this table, we ask questions about the variability of EMG and Ultrasound device placements and size, and the effect of it on fidelity (accuracy and range of gestures) and complexity. We seek to improve the fidelity always, as current state of the art is still not acceptable for practical usage due to the low accuracy of the systems. To reiterate: the end goal of this line of work is to design a system that is minimally invasive but provides full hand tracking capabilities, similar to what we can currently do with optical or glove based systems.

For clarity, as user-independent (or cross-user) device operation has made very little progress in this field of research, we will not consider this in our analysis of work. All work in this thesis focuses on user-dependent classification of gestures. We consider the implications of this in the discussion chapter.

To begin with, we use the table of analysis to target the variability in the placement of EMG devices. Therefore, the first exploration is to study the assumed trade off between gesture recognition accuracy and placement on the forearm. In particular, we studied the viability of placing EMG devices on the wrist, to conform to our ideal form factor. After two studies, we found that EMG can be used effectively at the wrist, with complementary sensing techniques to further boost the accuracy, such as pressure sensing. However, experientially, we found that gestures needed to be performed forcefully for any reasonably accurate recognition rates. This hinders some applications of the technology, such as implicit gesture detection, where the hand poses may not be performed forcefully on purpose. Additionally, there is the problem of having difficulty in detecting continuous gestures due to the nature of the signal. These are issues which are almost always not mentioned or discussed as being limitations in past work, they are assumed to exist and solutions to them are ignored.

This led us to look back at the different techniques to find better solutions to these limitations of EMG. Ultrasound had recently been used to detect hand poses. Again, in recent research, numerous locations on the forearm had been employed. Similarly, we investigated accuracy changes due to placement variation. We also checked to see if continuous motion tracking is possible. We found that there was not a large difference between placements and accuracy, finding that in fact the wrist placement worked very well. Continuous motion tracking also worked well, enabling more applications. However, from hands-on experience studying the device, we know that there are huge practical issues with the technology. This includes the use of ultrasound coupling gel and massive device complexity. The necessity of acoustic coupling gel spurred us to experiment with hydrogels, which showed early signs of one possible solution. But the issue of device coupling to the skin is not only found in ultrasound. In fact, in many techniques, good coupling is a must for the device to operate effectively. For instance, with EMG, the electrodes must be either coupled using conductive gel, or pressed hard against the skin in the case of dry electrodes. This led us to re-analyse the related work and consider how coupling might affect practical use of each technology.

Finally, after looking at the newly updated analysis of techniques, we discover that infrared techniques are different to most others as skin coupling is easier. We investigate infrared sensing as a gesture recognition technique, and find a way to improve the accuracy of it through a novel sensing architecture. More importantly, we find that it is robust under poor coupling conditions (i.e., when it is not skin-tight). While this seems like the best technology going forward, as it satisfies most of the requirements, the accuracy still leaves much to be desired. For one, there is yet to be a demonstration of the ability to detect continuous motions using infrared.

These three main pieces of work individually progress research in their respective technologies: providing information about the effect of position on accuracy, novel methods to improve accuracy, and an investigation of robustness issues. This accounts for the majority of the contribution of this thesis. With thorough analysis and experience with many of the techniques, we also provide a holistic overview of current state of the art and insights into key limitations or fundamental problems of some of them.

1.6 Thesis Roadmap

Having outlined the envisioned usage and requirements for wearable gesture recognition devices, the next part of this thesis will analyse existing techniques and assess their suitability, with respect to these requirements.

However, before doing so, we will give a primer on relevant anatomy and hand movements in the next chapter. This is vital to the understanding of the methods discussed in depth in the following chapters. Although it is perhaps not so vital to the understanding of the bigger picture of the thesis: which include the implications of the contributions of each piece of work and the strengths and weaknesses of each technique. The primer on anatomy is included merely for completeness of the thesis.

After this, the next three chapters: 4, 5 and 6 are works published during my PhD. They aim to explore practicality while still improving accuracy. After experimentation with each method, other practicality and usability concerns became clearer to me. In the discussions found at the end of these chapters, I will amend the original requirements for the device to include these concerns. Given these new requirements, I re-analyse the "landscape" of work. Each subsequent chapter aims to find a method in past literature that can satisfy the newly revised requirements or gaps in the analysis. A full table of chapters is given here for the reader to get a better sense of what to expect in each chapter.

The following table gives details on the contents of each chapter. For those who are interesting in a brief read of the thesis to gain an understanding of the implications and results of the work, and would rather leave out the details of the individual pieces within, I would recommend that you read: the related work, the beginning and ends of each of the main chapters, (chapters 4, 5 & 6), the discussion and conclusion. For those seeking the final requirements and analysis of methods, which is main outcome of this thesis, I would recommend skipping straight to the discussion and conclusion.

Chapter 2: Primer on Anatomy	A primer on anatomy. Pre-requisite knowledge to gain a full understanding of the inner-workings of each technique. Also helpful in determining the full range of human hand motion made possible by the bio-mechanics of the hand.
Chapter 3: Literature Review	Mostly an in depth literature review of well established methods for gesture recognition. Following detailed explanations for each technique, we evaluate them based on the requirements we formulated in the introduction. We find that there are unanswered questions for certain techniques, in particular the device placement upon the forearm. Based on this, we investigate electromyography and ultrasound in the subsequent sections.

Chapter 4: EMG	In this section, we explore the effect of Electromyography (EMG) sensor placement on classification accuracy. In our initial study, we expected EMG to perform worse when placed at the wrist, however our findings showed otherwise. We hypothesised this was due to motion artefacts in the signal, created by wrist deformations. We tested our hypothesis using a novel sensing system based on the combination of EMG and pressure sensing that enabled us to separate the two signals. We found that wrist deformations picked up by the pressure sensors, are complementary to EMG sensing. Through experimentation with the technology and in literature, we identify that EMG has difficulties capturing hand motions that are continuous, or weakly performed. The content of this chapter is mostly work I have published in the Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16): "EMPress: Practical hand gesture classification with wrist-mounted EMG and pressure sensing" [82].
Chapter 5: Ultrasound	A chapter of two parts. We look at ultrasound to answer one of the questions from our table of analysis - the variation in performance due to placement and orientation. We also investigate it's potential to capture continuous or weakly performed gestures, that may enable control of continuous values, or subtly/implicitly performed gestures. We find that while ultrasound is an ideal technology from a performance perspective, it is massively hindered by feasibility issues but also importantly, issues that require the use of a coupling gel. The content of the former and more prominent part of the chapter is a revised version of work I have published in the Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17): "EchoFlex: Hand Gesture Recognition using Ultrasound Imaging" [81]. The latter part of this chapter that addresses the gel coupling problem of ultrasonography is mostly taken from work that I have published in the Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces (ISS '17): "Improving the Feasibility of Ultrasonic Hand Tracking Wearables" [79].

Chapter 6: Infrared	After reflecting upon the necessity sensor coupling to skin, that occurs in most techniques found in gesture recognition, we find that infrared could be a method that is not be affected by this. Through our experimentation, we find that infrared is more robust to weak coupling to the skin and works without being skin-tight. We also develop a novel sensing approach that increases the performance of infrared sensing, nearing accuracy levels close to EMG. The content of this chapter is mostly a revised version of work I have published in the Proceedings of the 2017 UIST Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17): "SensIR: Detecting Hand Gestures with a Wearable Bracelet using Infrared Transmission and Reflection" [80].
Chapter 7: Discussion	In this chapter, we present the final revision of the requirements for an ideal gesture recognition device. Having gained a holistic view of methods and hands-on experience in some of them, I describe my opinions on the most promising methods for the future. I advise on future research directions for each of these techniques, based on our analysis that points out current weaknesses of each technique. We also identify difficult challenges that lay ahead in the field of gesture recognition, including robustness issues and anatomical uniqueness that hampers practical usage.
Chapter 8: Conclusion	A chapter to briefly concluding all findings in the thesis.

CHAPTER 2

PREREQUISITE KNOWLEDGE: PRIMER ON ANATOMY

2.1 Introduction to Hand Anatomy

The human hand is incredibly dexterous, capable of precisely controlled movements that enable many of the applications that we use on a daily basis. We owe this to the intricate muscular-skeletal design that evolution has crafted. This dexterity is part of what make us as humans unique as a species; in particular we have opposable thumbs that can connect with every other digit.

This section aims to make the reader familiar with the fundamental bio-mechanics of the hand and anatomical terminology. This will provide an easier understanding of the techniques discussed throughout this thesis. This section will begin with an overview of basic anatomical terminology, necessary for the detailed explanation of the hand and forearm that follows. There is an overview of the bio-mechanical hand movements that are possible, with an explanation for how these movements are created. Much of the content here, including images and anatomical definitions, is relevant information taken from "Gray's anatomy for students" [31].

It is worth noting that there are always subtle differences in anatomy between people. Though sometimes these differences can be quite significant. For instance, the palmarus longus (a muscle in the forearm) is absent in about 15% of the population [110]. Not only are there differences within and between ethnic populations, there may also be differences in certain individuals unilaterally rather than bilaterally (one arm different to the other). The implications of this are that a "one size fits all" device is impractical and as explained in more detail later, creates difficulties in the gesture detection across people.

2.2 Basic Anatomical Terminology

The following terms describe the location of a particular anatomical part. The terms are in reference to the left and right of the subject, rather than the observer. In addition, the human in reference is standing, with the arms facing downwards and palms facing forward. These terms are taken from "Gray's anatomy for students" [31]:

- Anterior and posterior. These describe whether the body part is in the front (anterior) or back (posterior) in relation to the body. For example, the palms of our hands are anterior as they are facing forwards.
- Proximal and distal. These describe a position that is further (distal) or closer (proximal) from the trunk of the body. For example, the wrist is proximal to the elbow, and the fingers are proximal to the wrist.
- Superficial and deep. These describe structures that are further from (deep) or closer to (superficial) the surface of the body. These are useful for describing the location of muscles as they are in dense layers within the forearm.
- Medial and lateral. These describe a position that is further from (lateral) or closer to (medial) the mid-line of the body. For instance, the thumb is located laterally to the other fingers. This is also used to describe movements of the hand.
- Superior and inferior. These describe positions above (superior) or below (inferior) another part of the body.

The following terms describe the relevant anatomical tissues needed for explaining the anatomical structure of the hand:

- Bone. The skeletal system consists of many bones which give structure and form to the body. The interface between two (or more) bones are sometimes created as joints to allow different degrees of movement.
- Muscle. Specifically, skeletal muscle, are attached to bones and are capable of strong contractions that are used to move skeletal elements. There are often groups of muscles, opposing each other around a joint. Such grouping of muscles are named antagonistic pairs: when one group contracts, the other relaxes, and vise versa. This is to ensure control of the moving part both forwards and backwards. The nervous system allows the human to control the muscle contractions and thus movement of the body (described in more detail in section 2.4).
- Tendon. The muscles attach to bones with fibrous connective tissue called tendons. As the muscle contracts, the tendon pulls on the bones, towards the muscles origin. Some tendons

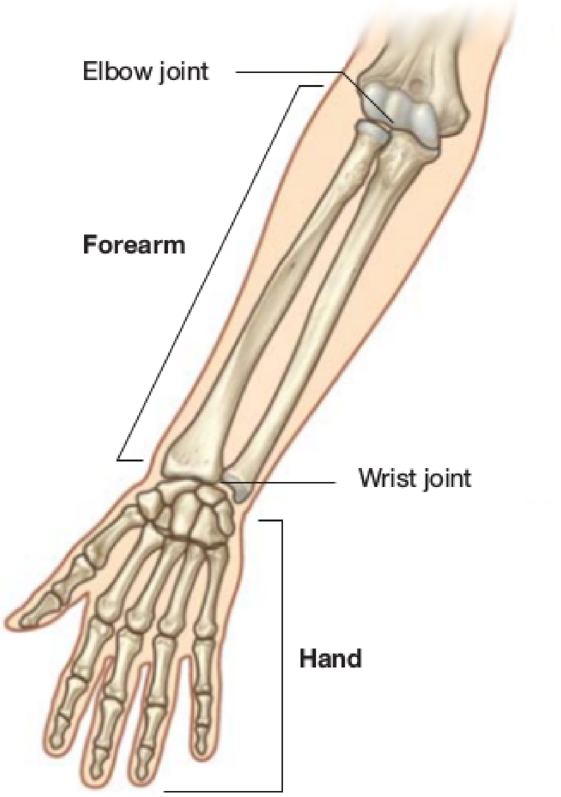

FIGURE 2.1. Basic structural diagram of the hand and forearm. Figure reproduced from [31].

can be seen protruding from the surface of the skin, which is the underpinning of several gesture detection techniques.

- Ligament. Ligaments connect bones together at joints. The elasticity and placement of ligaments restrict the movement of joints.

2.3 Movements & Structure of the Hand and Forearm

The human hand moves due to the way the bones, muscles, tendons and ligaments work together. The bones give structure and rigidity to the hand. The joints between the bones and the muscles used to control them determine the freedom of movement of and within the hand.

The hand consists of the central palm, and the digits (fingers and thumb) that are connected to it (Figure 2.1). The digits themselves have several joints that enable a wide variety of grasps. The hand is connected to the forearm via the wrist. The wrist is comprised of many bones that create joints which allow the hand to rotate freely, sometimes the wrist itself is seen as a joint. The palm does not play a crucial part in the bio-mechanical movement of the hand, because of

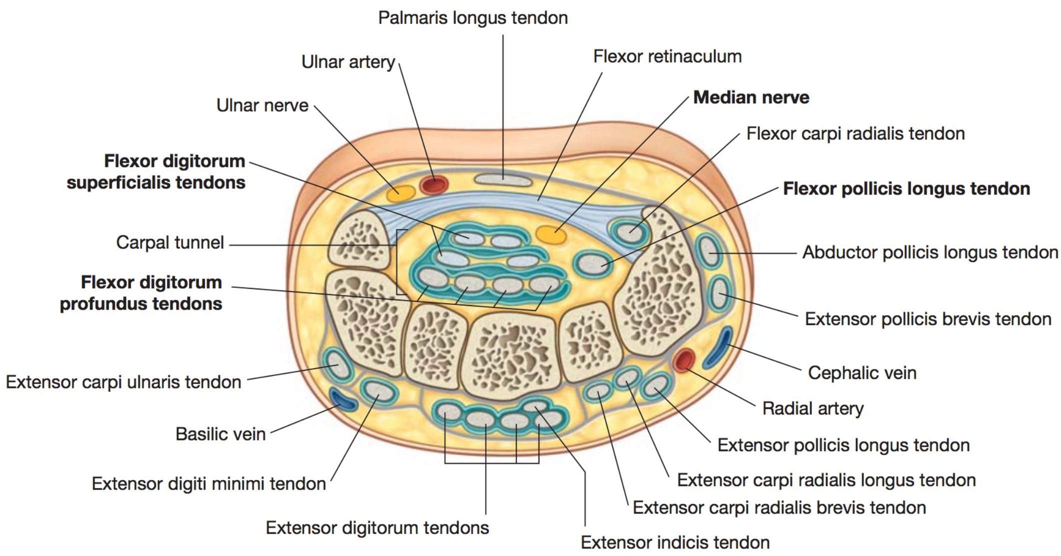

Figure 2.2: A diagram showing a cross section of the wrist. Figure reproduced from [31].

this we will not describe the anatomy of the palm in much detail. The forearm consists of two bones, the radius and the ulna, which can twist to create rotational movement of the hand and wrist. Finally the forearm can move around the elbow joint.

The majority of hand movements are created by muscles in the forearm. These are known as extrinsic hand muscles. Some of the more delicate finger movements and thumb movements are controlled by muscles located within the hand - intrinsic hand muscles. This point is of particular interest to techniques that infer gestures using these extrinsic muscles. The muscles in the forearm transition into tendons which pass through the wrist and into the hand to control the appendages. Each muscle usually has multiple functions, likewise each movement if often requires a group of different muscles working together. It is not vital for the reader to know the exact muscle to movement mapping, and so a less than thorough anatomical overview is presented. However, the full details are provided below, with a diagram of a cross section of the wrist, displaying the tendons (and thus also muscles) in Figure 2.2, is here for self-containment of the thesis.

The following describes the movements of the hand and forearm in further detail, with an explanation for how the anatomy enables such movement (mostly reproduced from [31]).

2.3.1 Pronation and Supination of the Forearm

The muscles in the forearm enable a twisting motion 2.3. The radius and ulna, the two largest bones that support the forearm, are able to cross over in this motion. Pronation is movement that forces the hand to turn with the palm facing downwards. Opposing this motion is supination,

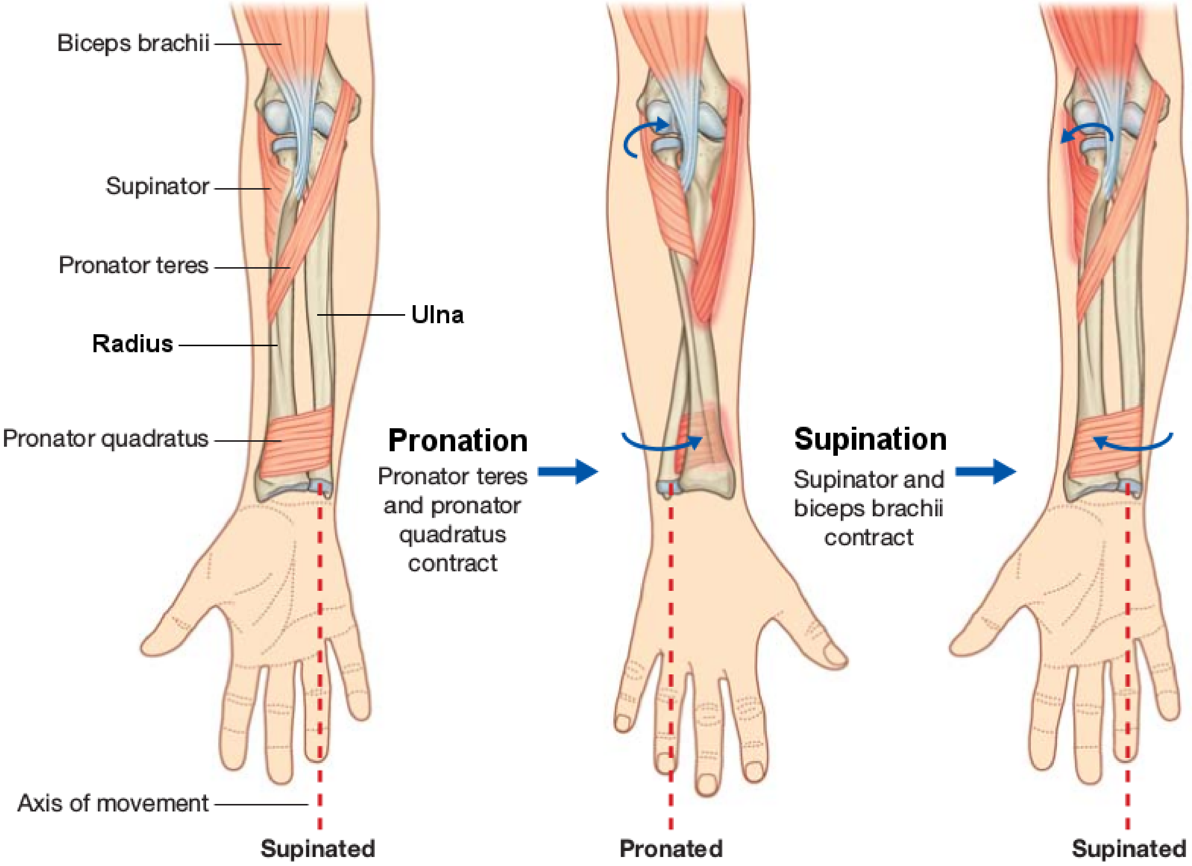

Figure 2.3: Pronation and Supination of the forearm describe a twisting motion. Figure reproduced from [31].

which results in the palm facing upwards. The joints that allow this movement are the proximal radioulnar joint (at the elbow) and the distal radioulnar joint (at the wrist). Muscles responsible for these movements include the supinator, pronator teres and pronator quadratus. Muscles are normally named in accordance to the type of hand movement that they are primarily responsible for.

2.3.2 Flexion, Extension, Adduction and Abduction of the Hand

The wrist joint allows the hand to move forwards or backwards about this joint, as illustrated in Figure 2.4. The forward movement is more accurately known as flexion (to flex), which is described as a bending movement around a joint that decreases the angle between the adjoining bones. The opposite of this is known as extension (to extend), which describes a straightening motion that increases the angle. Sometimes, the joint allows bending both forward and backward. Such is the case with the wrist joint, flexion then refers to bending towards the anterior side of the body (the palm - palmar flexion) and extension refers to bending movements towards the posterior

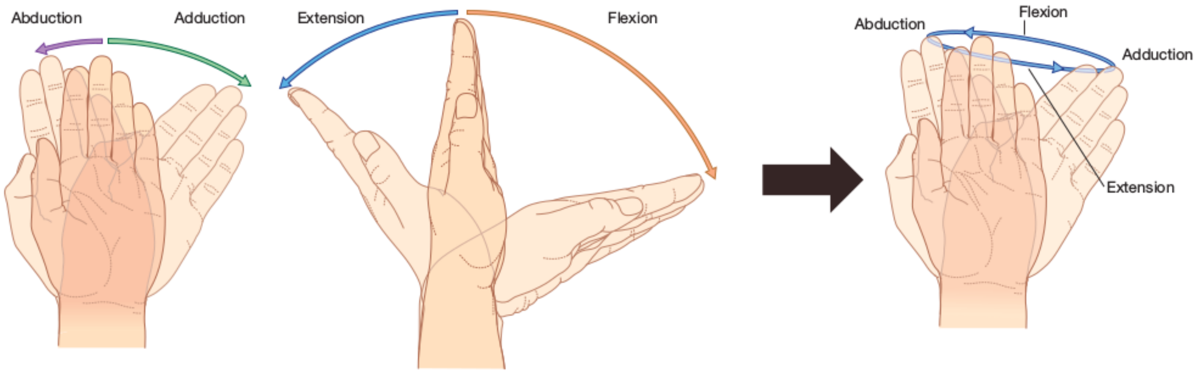

Figure 2.4: Wrist motions of the hand. Figure reproduced from [31].

side. Flexion of the wrist uses these muscles: flexor digitorum superficialis, flexor digitorum profundus, flexor carpi radialis, flexor carpi ulnaris, abductor pollicis longus, flexor pollicis longus. As the names of the muscles suggest, these are used primarily for flexion movements. Similarly, extension of the wrist uses extensor digitorum, extensor carpi radialis longus, extensor indicis, extensor digiti minimi, extensor carpi radialis brevis, extensor pollicis longus, extensor carpi ulnaris.

The joint also enables movement of the hand from side to side, which is used for instance in Western culture, to wave ones hand to signify "hello" or "goodbye". The anatomical terms for these movements are adduction and abduction. Hand abduction around the wrist joint, is movement that pulls the hand towards the side with the thumb, as it is moving away from the body as per the definition of abduction (still considering the neutral anatomical human posture described in Section 2.2). This is sometimes known as radial deviation, due to the pulling movement towards the radius bone. The muscles that perform this movement are: extensor carpi radialis longus, extensor pollicis longus, abductor pollicis longus, flexor pollicis longus, flexor carpi radialis. The opposite movement is called adduction, or ulnar deviation, and uses these muscles: extensor carpi ulnaris, flexor carpi ulnaris, extensor digitorum, extensor digiti minimi.

These four movements enable the hand to rotate freely around the wrist joint with two angular degrees of freedom, as shown in Figure 2.4. Along with pronation and supination of the forearm, this covers all the mechanical movements of the hand that are not within the hand i.e., fingers and thumb. All the muscles listed here belong to the forearm, with tendons that run through the wrist and into the hand.

2.3.3 Finger movement

The digits of the hand are numbered from 1 to 5, from thumb to little finger (figure 2.5). As a simplification, the work in this thesis refers to the digits mostly by their names: thumb, index, middle, ring and pinky/little. Each of the 5 digits of the hand contain several bones, called phalanges. There is one proximal phalanx, one middle phalanx, and one distal phalanx for each

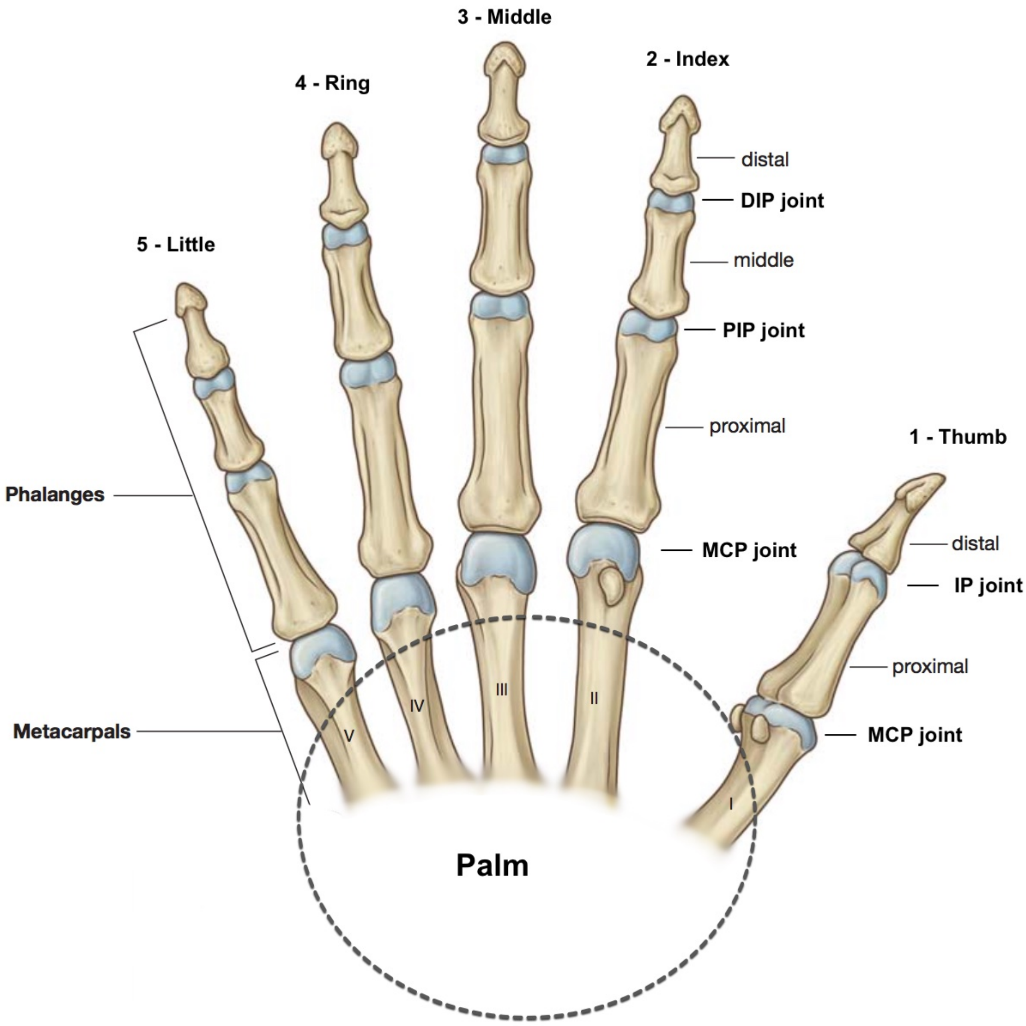

Figure 2.5: Number and names of digits of the hand. The bones and joints of the digits are labelled. Figure edited and reproduced from [31].

digit, except for the thumb which does not include a middle phalanx. The phalanges within each digit are connected together with ligaments. The palm also contains bones, one for each digit, called metacarpals.

Fingers have a large range of movement due to the many joints between the phalanges. The base of each proximal phalanx is connected to the metacarpal bones. The joint at the intersection of these bones is the MetaCarpoPhalangeal (MCP) joint - more commonly known as the knuckle. The MCP joint is the predominant joint allowing the fingers to spread apart from each other (abduction) as seen in figure 2.6. This joint also allows digit flexion. The joints within the digits beyond the MCP are called the interphalangeal joints. The fingers have two, one nearer to the knuckle called the Proximal Interphalangeal Joint (PIP), and the one further to the finger tip

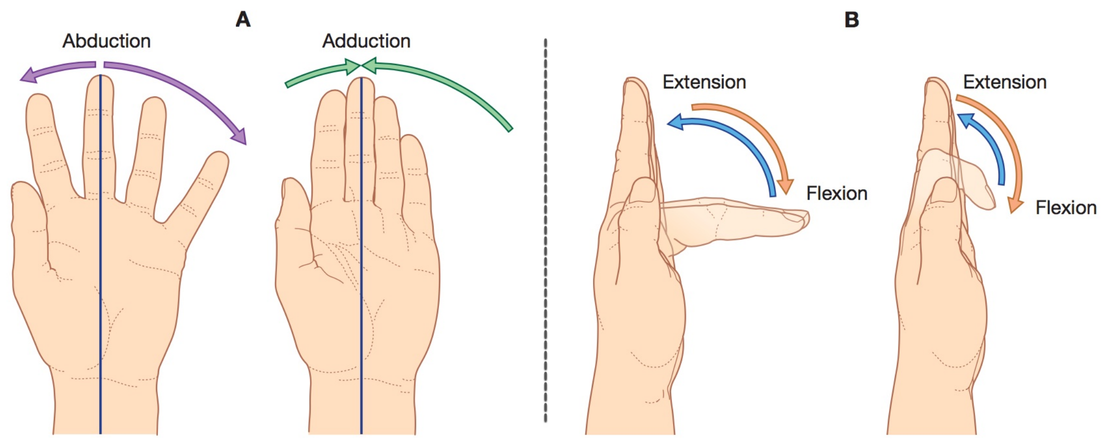

Figure 2.6: A) Finger adduction and abduction. B) Finger flexion and extension at different joints. Left: flexing around MCP joints, Right: flexing around IP joints. Figure reproduced from [31].

called the Distal Interphalangeal Joint (DIP). The Thumb only has one Interphalangeal Joint (IP). The digits can flex and extend with numerous degrees of freedom.

As illustrated in figure 2.6, the fingers can flex at different joints individually. Most of the flexing happens at the MCP and PIP joints, but there is some minor control over the DIP joint. With the addition of adbuction and adduction of the fingers, this means each finger has a total of 4 degrees of freedom (DoF) [36]. The muscles that are primarily used for flexing the fingers are the flexor digitorum profundus that bends the finger around the proximal and distal joints, and the flexor digitorum superficialis that bends the finger at the proximal joints. The flexor digiti minimi brevis assists in flexing the little finger. As for extension of the fingers, the extensor digitorum communis, extensor indicis proprius (for the index) and extensor digiti minimi (for the pinky). These muscles originate from the forearm, with tendons that pass through the wrist and up to the fingers. There are however, intrinsic muscles that are responsible for finer control of the fingers. The lumbricals are intrinsic muscles of the hand, located within the palm, that extend the IP joints and flex the MCP joints of the fingers. The interosseous muscles are those that are attached to the fingers themselves. These muscles assist the lumbricals to flex and extend the fingers, but also adduct (palmar interossei) and abduct (dorsal interossei) the fingers. It is important to note that this particular movement can only be performed by intrinsic hand muscles (interossei).

2.3.4 Thumb movement

The thumb is different, and has an additional degree of freedom compared to the fingers. Although it only has one IP joint between the phalanges that allow flexion/extension and the usual MCP joint that allows flexion/extension and abduction/adduction; the carpometacarpal (CPC) joint between the metacarpal and the base of the wrist joint allows further flexion/extension and

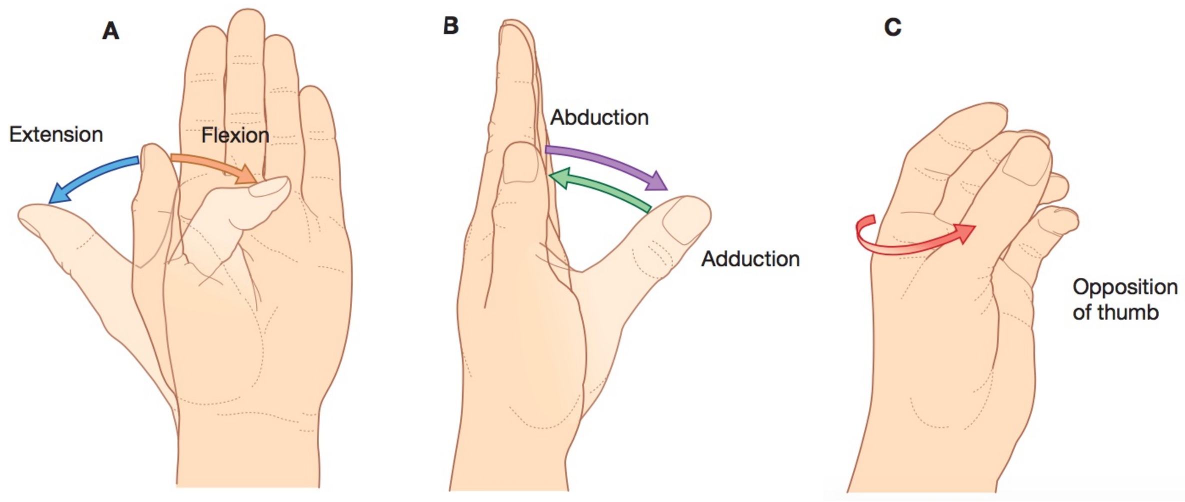

FIGURE 2.7. A) Thumb flexion and extension. B) Thumb adduction and abduction. C) Opposition of the thumb with the another digit, a pinching motion. Figure reproduced from [31].

abduction/adduction of the thumb (additional 2 DoF). The thumb therefore has 5 DoF in total.

Thumb flexion is produced by the flexor pollicis longus and brevis. Thumb extension is produced by the extensor pollicis longus, brevis and abductor pollicis longus. There are some differences between which muscles control flexion and extension at the MCP vs CPC joint. Abduction of the thumb at the CPC joint is mainly controlled by the abductor pollicis longus and brevis; adduction mostly by the adductor; abduction mainly by the adductor pollicis muscle. Some of these muscles form a group called thenar muscles. These are intrinsic hand muscles which are located in the palm of the hand, at the base of the thumb. Similar to the intrinsic finger muscles, these muscles produce finer movements of the thumb relating to abduction, adduction, flexion and opposition of the thumb.

This concludes the motions of the hand and the muscles used to produce them.

2.4 Muscle Innervation

A muscle is comprised of many muscle fibres that are able to contract and relax in order to create movement and forces [30]. The initial action that triggers a chain of events that leads to this contraction begins with a neural impulse from the central nervous system (i.e. the spinal cord and brain). A nerve impulse is able to propagate through to other neurons closer to the intended muscle through relay neurons. This impulse is electrical, caused by the movement of ions (neurotransmitters) in and out of the neuron. When the signal eventually reaches the desired muscle, the motor neuron that innervates (supplied with neurons) several muscle fibres transmits a signal through the neuromuscular junction. In technical terms, the axon of the motor neuron releases acetylcholine (ACh, the neurotransmitter) which binds to the receptors of the motor end-plate of the muscle fibres. Once the acetylcholine binds, a channel in the receptor opens up causing the muscle fibre to depolarise as positively charged ions flow through this channel. Soon after this depolarisation happens, voltage-gated sodium channels are opened, allowing sodium ions to enter causing an action potential to spread along the muscle fibres, initiating contraction. The collection of muscle fibres and the motor neuron that innervates them is called a motor unit. The action potential that is produced by such a unit is called a motor unit action potential (MUAP).

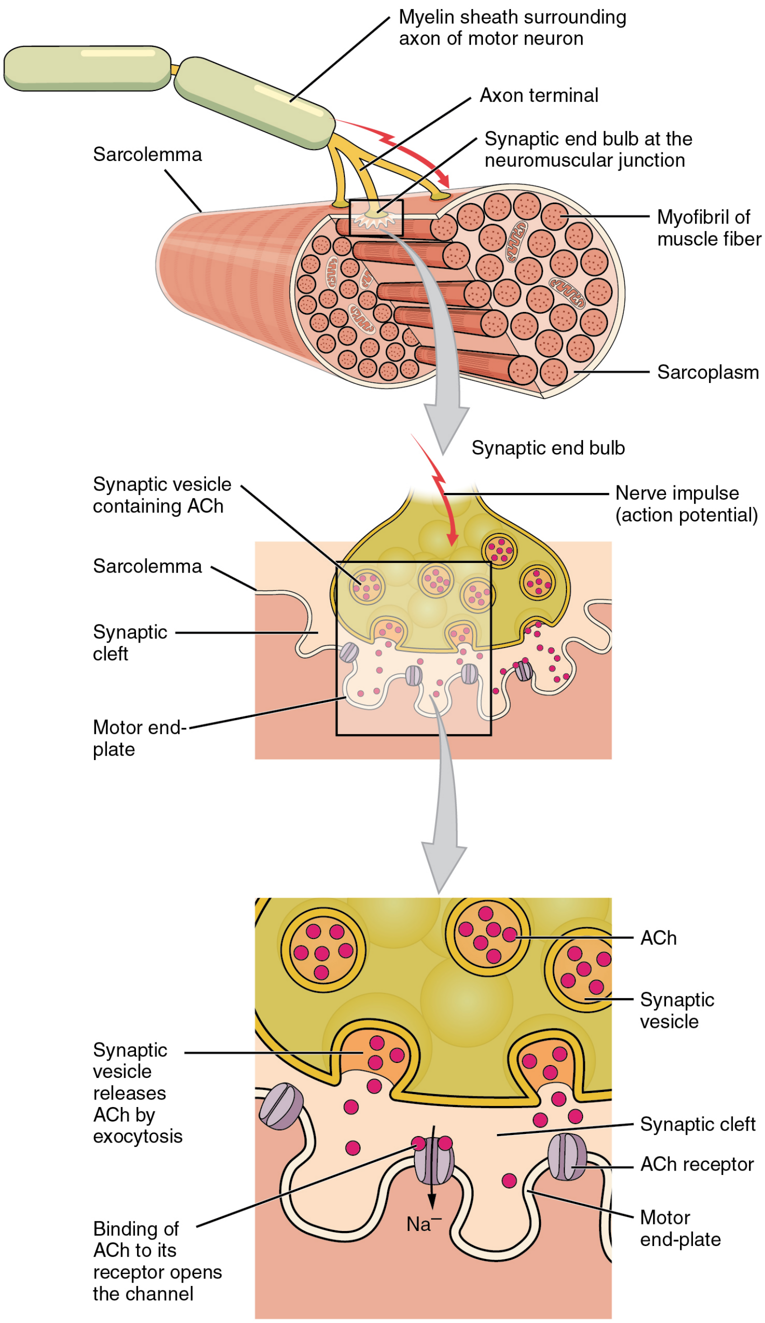

FIGURE 2.8. Motor neuron axon innervation into end plate. Figure reproduced from [49].

CHAPTER 3

LITERATURE REVIEW

3.1 Wearable Gesture Recognition Methods

This section reviews previous work regarding wearable hand gesture recognition techniques. Here is listed the techniques that we will review:

Optical - Methods that use direct optical visuals of the hands in order to infer hand pose.

Glove - Sensors that attach to the appendages to track the flexing of joints in the hand.

EMG - Electromyography is a sensing technique that senses the muscle activity using electrodes, to infer hand movement.

Ultrasound - Ultrasound imaging uses sound waves to image body tissue, including forearm muscles.

Infrared - Using the superficial nature of the tendons in the forearm, this allows sensing of the movement using infrared distance sensors.

Pressure sensing - As above, but with pressure sensors.

EIT - Electrical Impedance Tomography is a technique that uses the impedance of body tissues, which changes between hand poses, as a way of classifying gestures.

Mechanomyography - Movements of the muscles in the hand and forearm produce low frequency vibrations that can be sensed.

IMU - Inertial Measurement Units can detect movements of the hand and facilitates gestures recognition in another dimension.

Gesture detection requires some information of the hand pose. The most obvious way is to do so visually i.e., with cameras. Much work has gone into capturing hand pose using cameras [84, 112], which are now able to very accurately track human hands. However, as we later discuss, vision based tracking solutions are not ideal for mobile hand tacking.

The other obvious way to detect hand pose are gloves with embedded sensors [38]. Although this method can accurately track the hand pose, instrumenting the hand can inhibit fine or expert manipulation e.g., small objects, surgery or simply keeping the physical naturalness of interaction with objects (because the skin may be covered, for example).

Conveniently, the smartwatch provides space to embed sensors in the watch face or strap. In addition to this, the social acceptance of the placement is not an issue since watches have been a desirable location for worn devices for many years. This has spurred research into creating new gesture recognition technology that can fit into this space. The applications of always available gesture recognition are recent, and so development of this technology is still in it's infancy. However, a lot of techniques discussed in the upcoming sections were not designed with this form factor in mind, but I will cover each of these in detail and analyse it's potential as wrist-worn form factor device. The analysis shows that EMG, ultrasound, infrared, EIT and pressure sensing are all methods that have been shown to integrate with wrist-worn form factors or have the possibility of being integrated. Towards the end of this section I will compare these techniques against the original set of requirements listed in the introduction (section 1.4).

3.1.1 Changes in Musculature

Many of the techniques in the literature rely on the basic principle of morphological wrist changes during hand movement. To explain this more clearly, as the hand changes in pose such as a finger being flexed or the hand tilts upwards, the muscles and tendons in the forearm that produce these movements shift around. This shifting can be seen visibly from the surface of the forearm. Proximal to the elbow, the change in shape of the forearm is created by contracted muscles. But distally, close to the wrist, the changes are seen as a result of superficial tendon movement. To illustrate this, figure 3.1 shows the shape of the wrist between two different hand poses. These changes may be seen clearer in upcoming sections which include sonographic images of the anatomy, but interpreting these images is somewhat difficult.

The only caveat to methods that rely on this principle is that it is usually only superficial changes in musculature that can be observed from the surface. This means less information is available to work with when estimating the hand pose.

There are a multitude of methods to detect these changes as shall be discussed in the upcoming sections.

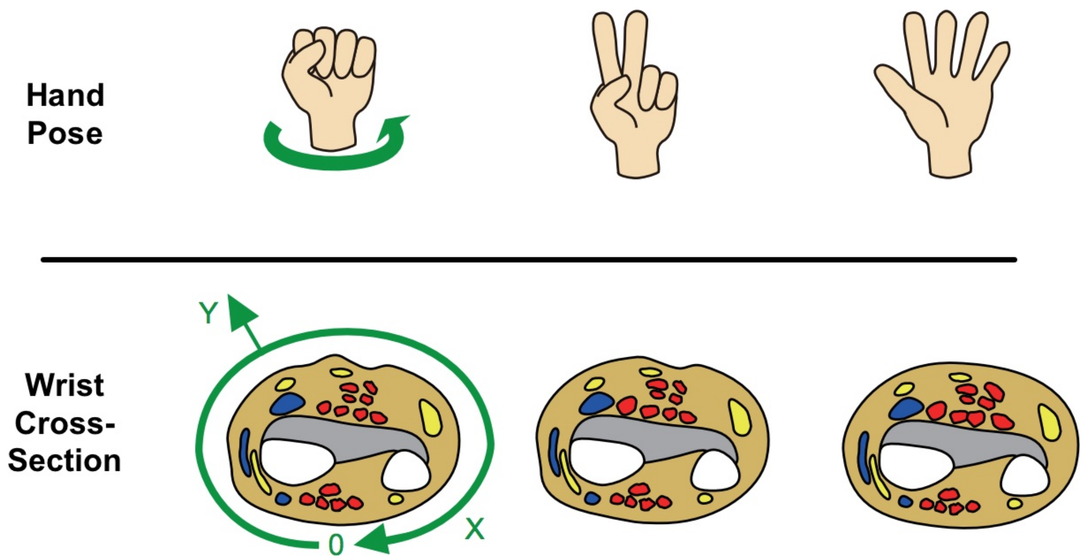

FIGURE 3.1. Morphological changes in the shape of the wrist as a result of changes in hand poses.[1]

3.1.2 Gesture Segmentation

A segmentation or delimiter method, as we define here, is a way of determining when a gesture starts or ends. It may also be used to discriminate between natural body movements and explicitly performed gestures. Delimiter methods are out of the scope of this work, as it is a large topic of research in itself. There are numerous works that present techniques to try to segment gestures to avoid false positive detection, such as using additional sensors as mentioned in Wristwhirl [46] and Pactolus [23]. Conversely, implicit gesture recognition does not require the use of a delimiter.

3.2 Optical (camera)

Optical methods are based on employing a camera to determine 2D or 3D hand position. The camera can be sensitive to the visible spectrum but alternatives exist in the infra-red or below. A review of these methods can be found in work by Rautaray et al. [101].

Figure 3.2 shows possible mounting positions of cameras on the body. As mentioned briefly at the start of the chapter, there are difficulties when using cameras for gesture detection. To begin with, vision based hand tracking is suited for stationary applications where there is space to mount cameras. Cameras attached to the body tend to be conspicuous due to their size. As the camera points outwards, there is no control of what is recorded beyond the wearer's hands. This creates legal and ethical issues using this method [126].

[1]Reproduced with edits from "Hand shape classification with a wrist contour sensor: development of a prototype device" by Fukui et al. [42]

[2]Reproduced from "Effects of camera position and media type on lifelogging images" by K. Wolf et al. [125]

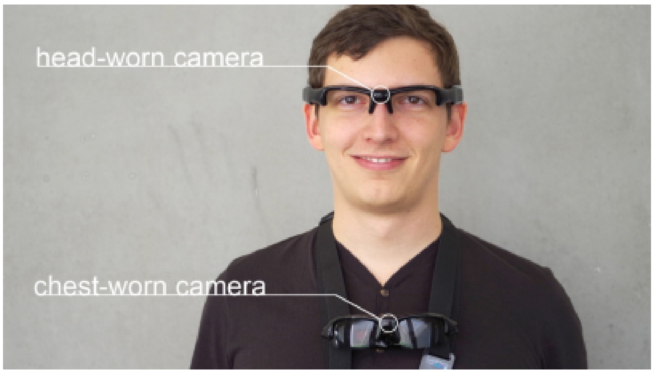

FIGURE 3.2. Some possibilities wearable locations for body-mounted cameras, including the chest and on glasses.[2]

Furthermore, the camera may not always be capturing the user's hands depending on where the camera is mounted. For instance, a camera mounted on glasses can only reliably capture hand pose if the hands are in front of where the user is looking. Even then, the fingers can be easily occluded by the hands or arms of the user. This issue is also described by Bailly et al., where the camera has an inability to capture large and demonstrative gestures [10]. Despite this, there is work that successfully used computer vision in a wrist mounted wearable to track limited movements of the fingers [65]. The system, Digits, used an IR camera attached to the wrist to sense the distance to the fingers. The system does have drawbacks, however.

Firstly, the sensor is worn on the wrist and as with all optical approaches requires direct line of sight to the fingers. In order to achieve this, the camera must protrude far out from the wrist to gain line of sight. This protrusion is significant enough to be considered too large for integration with current wrist-worn form factors. This is a fundamental problem with the technology and not of the specific implementation.

Secondly, the device suffers from occlusions by the hand itself. If the hand is tilted in certain ways, the fingers become occluded. If the thumb is opposed, it can also occlude fingers. Some of these hand movements may be vital to a gesture set.

For these reasons, we cannot consider optical methods to be mobile yet, without some revolutionary technology that can avoid the protrusion of a camera or occlusion problems.

3.3 Glove

Gloves hinder interactions with objects due to the physical encumbrance which either restricts freedom of movement or the additional size. An exemplary scenario where this is an issue is in space exploration. The glove of the space suit creates a barrier that inhibits human sensory perception [109]. Gloves may also create additional fatigue due to restricted freedom or extra strength required to move the fingers [118]. Sorenson et al. attempt to develop an unobtrusive glove that will attempt to "provide a suited crewperson with as close to nude-body hand dexterity as possible".

Despite the unobtrusive nature of gloves, they are extremely accurate for hand tracking, and are sometimes viewed as the de facto method for attaining the ground truth of the hand pose for which other methods can compare against.

Gloves can use a variety of sensors in order to measure finger and wrist movements, from accelerometers [66] to flex sensors [71]. They are also sometimes used in conjuction with computer vision, where the gloves may have markers to aid tracking [122].

3.4 Electromyography

3.4.1 Fundamental Concepts and Example use in Healthcare

As explained in section 2.4, muscle contraction is initiated by a complex series of events that propagates a signal from the motor neuron to an eventual action potential that spreads across the muscle. This action potential can be measured using electrodes to monitor muscle activity. Electromyography (abbreviated as EMG) is the name for this technique. EMG is used for a variety of different purposes, but most applications belong to medicine. It can be used to analyse muscles for abnormalities such as neuromuscular diseases. More relevant to this thesis is the use of EMG for controlling prosthetic devices - in particular prosthetic hands controlled using EMG to monitor residual forearm muscle activity.

The electrodes that are used for measuring the electrical impulses can be either invasive, using needles to probe specific muscles. Otherwise they make contact with the skin above the muscle in a non-invasive manner. This method is called surface EMG (or sEMG) and is the type we are mostly concerned with in this work, since the invasive methods are beyond consideration in the practical design of wearable EMG systems. Surface EMG electrodes need good electrical conductivity. Electrically conductive gel is usually used to achieve high levels of conductivity needed for good signal to noise ratio. Often times adhesive pads are used in conjunction with gel to maintain good mechanical contact.

A simple and typical setup for an EMG measurement of a single muscle is shown in figure 3.3. In this setup, two electrodes are arranged to take bipolar readings of the muscle activity. In this arrangement, the differential of the signals from each electrode is amplified, with respect to a reference electrode which is placed elsewhere on electrically inactive tissue. The other type of setup is monopolar, where there are numerous electrodes and only one reference electrode. Measurements are taken between the reference electrode and each of the other electrodes individually. The advantage of bipolar is that it eliminates the common noise in the electrodes and thus have better signal to noise ratio. An example EMG signal that one can expect to see for large muscle contractions is shown in figure 3.4. The magnitude of the signal corresponds to the intensity of the muscle contraction.

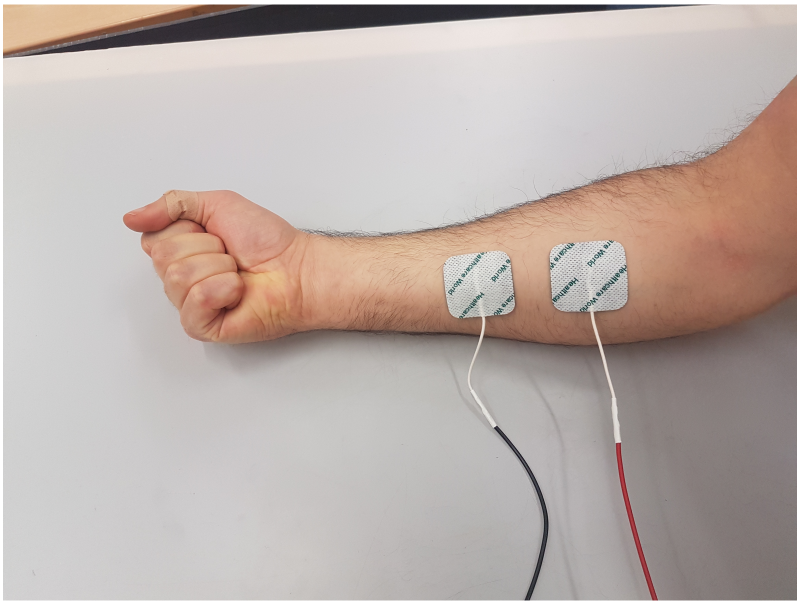

FIGURE 3.3. EMG bipolar setup on the forearm. Reference electrode not shown.

The original signal is minuscule, in the order of a few millivolts [69] and sometimes even as low as a few microvolts. This means that the signal amplification needs to be very high, by a factor of at least 500. This makes the signal quite noisy due to the amplification of noise leaking in from the surrounding environment. Notch filtering to cancel mains noise is generally avoided due to destroying signal information. To amplify the signal by this amount requires a lot of energy. Portable EMG systems have been designed before but require large batteries in order to operate for long periods of time [85].

3.4.2 Gesture Recognition using EMG

Recently, EMG has been proposed as a way to recognise hand gestures using a wearable EMG device worn on the forearm.

From as early as the year 2000, EMG has been proposed as a method for interaction with computers [11, 27]. They were initially proposed as an aid for people with reduced motor functions or other disabilities. However, the applications for gesture control grew and the increased motivation for a gesture detection device spurred further research in other areas, including HCI.

Saponas et al. realised the potential that EMG could provide for enabling always-available input for computers [106]. This would be especially useful in interactions with wearable computers as discussed in section 1. With this vision in mind, they were one of the first to consider the form factor of EMG wearables and designed a small band to fit around the forearm.

However, research aiming to improve EMG has mostly shied away from investigating the size and placement of such devices. Instead, large and dense arrays of EMG electrodes, better

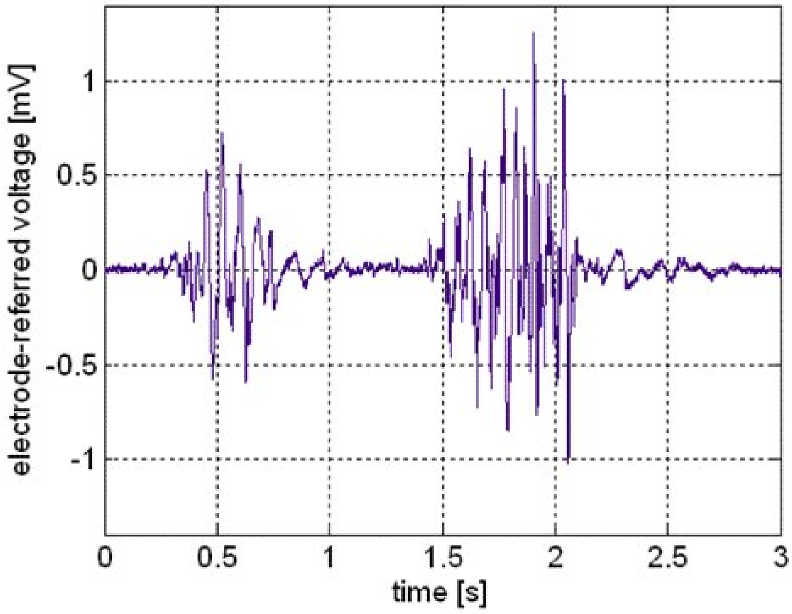

FIGURE 3.4. An example raw EMG signal from two separate contractions of the bicep.

signal acquisition circuits and state-of-the-art machine learning has improved the resolution and accuracy of hand pose detection, at the expense of practicality [7, 45]. However, this research does demonstrate excellent recognition rates for finger and thumb movements [7] and for wrist movements [44]. Large and dense sensor arrays may be more appropriate for integration into clothing than wearable devices [39]. As an example of this, NASA developed a wearable band for astronaut suits made of 17 electrodes and accelerometers [127]. This wearable represents a way of enhancing communications with other astronauts or vehicles.

EMG tends to suffer from a trade-off between accuracy and the surface area used for measurements. Whilst EMG proves to be highly effective for a high number of electrodes, it performs poorly under placement and surface area restrictions, making it awkward to fit into existing wearable device form factors such as watch straps. This is a result of the low number of muscle cells in the wrist compared to the proximal forearm (near the elbow). Additionally, the tendons in the wrist are more difficult to discriminate as they are more tightly packed. Thus there are increased challenges in performing EMG sensing of hand muscle movements by sensing at the wrist.

Despite this, research has shown that EMG placed at the wrist can also contribute information for hand pose [21, 90]. However the electrodes used in those setups also cover other parts of the forearm, so it is yet unknown how effective EMG is used only at the wrist.

The use of EMG sensors in practical wearable scenarios typically also requires calibration in order to account for slipping watch straps and to align sensors with the anatomically optimal detection points. This is an issue primarily for cross-session device placement shifts. While we do not consider the issue of calibration in this paper, shift compensation algorithms [7] and similar methods [32] are recent attempts to mitigate this issue.

High signal quality is difficult to achieve in sEMG. Electrodes are usually affixed to the patient/users skin with adhesives and conductive gel. However, this presents a serious issue as the recurring application of gel and adhesive is neither practical for regular use or comfortable. As a result, 'dry' electrodes (without gel or adhesive) are currently being developed to alleviate this problem. Recent designs for dry electrodes have comparable accuracy to wet electrodes, and are a more practical alternative [88].

Smith at al. demonstrated that EMG can be used to detect continuous angle of the fingers [117] with an NRMSE of 11%. However, the nature of the signal creates particular difficulties regarding continuous angle detection, and so there have been very few attempts to challenge this task using EMG, and is therefore currently more suited to discrete gesture recognition.

3.5 Ultrasound

3.5.1 Fundamental Concepts and Example Usage in Technology

Sound is a mechanical wave that works with the compression and rarefaction (expansion) of the particles in the medium the wave is travelling through. This compression and rarefaction can happen at different frequencies. In air, humans are able to hear frequencies from around 20Hz to 20KHz [104]. Sound above this upper threshold of 20KHz is considered ultrasonic, and is termed ultrasound. The higher the frequency of sound, the longer the wavelength. Thus, ultrasound has a very short wavelength. Sound is usually created by a source that vibrates, which propagates the vibration into the surrounding medium, like a vibrating diaphragm from a speaker that transfers the energy into air. Commonly in applied acoustics, piezoelectric materials are used to generate and receive sound waves. Piezoelectric materials are transducers of energy: they accumulate electric charge in response to mechanical stress, but also in reverse they generate sound waves when an electric field is applied.

Sound waves can be used to detect objects at a distance, by analysing the echoes that return from a short pulse of sound. This can be extended by using multiple emitters and/or receivers to locate the position of objects in 2D or even 3D. This technique is called echolocation and can be seen in nature. The most notable of animals to use echolocation are bats. The basic principle of echolocation is shown in Figure 3.5. Nearly a century ago, we also began utilising this technique to navigate under water (sonar) . Further research in this area paved way for imaging objects in detail using high frequency sound, finding applications in non-destructive testing and medical diagnosis. The effectiveness of this technique spurred more research from industry, creating more complicated imaging algorithms at higher frequencies. Using higher frequencies in ultrasound imaging produces more detailed images. This is because the shorter wavelengths are able to reflect off smaller objects (due to the diffraction limit). Hence, ultrasound is used for imaging,

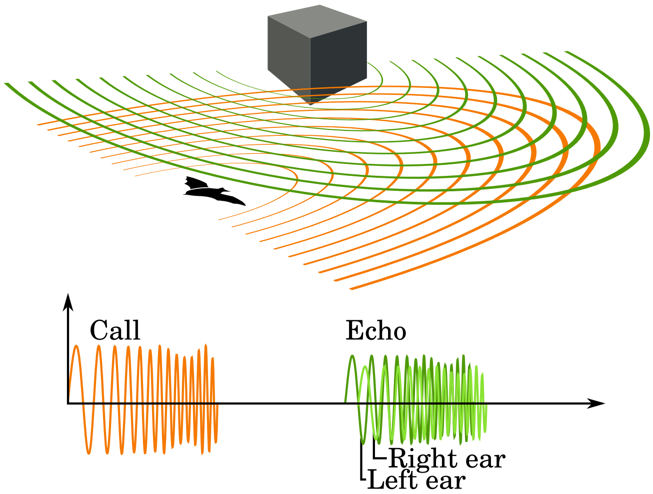

FIGURE 3.5. The working principles of echolocation. By emitting a short pulse of sound and receiving the echo with two ears spaced apart, the bat is able to locate the reflecting object.

FIGURE 3.6. A 3D ultrasound scan of a fetus at 17 weeks.

which is why the technique is named ultrasound imaging (or ultrasonography).

The latest ultrasound imaging machines can now provide extremely detailed images of objects. The most commonly known application for ultrasonography are for scanning the developing fetus during pregnancy. Ultrasound imaging machines are now so advanced, they are capable of creating 3D images (Figure 3.6).

However, the use of medical ultrasound imaging extends far wider, including scanning of

FIGURE 3.7. Medical ultrasound imaging equipment, showing the probe, driving circuity, physical control interface, computer and monitor.

organs such as the heart, or even the eyes. The most relevant use of this technology for our purposes is musculoskeletal imaging, which can image the tendons, muscles, ligaments, nerves and bones. This is commonly used by radiologists in assessing bones, joints and soft tissues for disorders.

The machines are usually comprised of a probe (synonymous with a transducer in this context), where the piezoelectric elements are contained within (Fig 3.7). These elements create focused ultrasonic beams into the desired medium and also receive echoes. The probe has different shapes and sizes depending on the application, but is usually shaped so that it can be held comfortably in one hand by the operator. The analogue signals that are collected by the probe are the sent to a signal conditioning and processing unit to sample the signals. The data is then sent to a computer to process in order to create images. The piezos are also excited by voltages generated from the same unit the probe is connected to.

The signal conditioning and sampling unit is very complicated. This is because the signals are very low in voltage, and therefore need accurate low noise amplification. Ultrasonic signals for

FIGURE 3.8. A flat, linear transducer/probe used for 2D ultrasound imaging.

medical ultrasound are usually beyond 3MHz, which in order to capture requires high precision and fast sampling rate analogue to digital converters. The circuit to drive the piezos in order to create specific beam patterns are also complex, requiring precise, sub-microsecond timing. Modern ultrasonic imaging probes use arrays of hundreds of piezoelectric elements [83], which scales the size and complexity of the driving and sampling circuitry proportional to the number of elements. So traditionally, the driving and sampling circuits are separate to the probe as they often cannot fit within. This has not been a limiting factor for medical ultrasound applications where patients come to the clinician's office for a diagnosis. However, point-of-care ultrasonography has recently been raised as an important application [86]. The necessity for a small and portable ultrasound imaging device has brought improvements to the design of such systems, and with improvements in electronic fabrication technology, it is now possible to integrate these circuits within the probe. Furthermore, the data can be processed using the smartphone as processors in them are more than capable of performing the imaging algorithms.

3.5.2 Using 2D Imaging for Discrete Hand Poses

Since gestures produce movements of the muscles and bones within the wrist (section 2.1, section 3.1.1), musculoskeletal ultrasonography can be used to predict gestures by analysing this movement. There are many examples of prior work that proposes to use this technology for controlling prostheses and other clinical applications. Zheng et al. coined the term "sonomyography" to describe the use of ultrasound to sense muscle activity [133].

Recent work by Akhlaghi et al. has demonstrated that ultrasonography can detect discrete gestures very accurately, with an classification accuracy of 91% for 15 gestures [3], rivalling the accuracy of other techniques such as EMG while using less surface area. They used image processing algorithms to do a pixel-wise comparison of brightness values for a series of images a video. The aggregation of these image differences produces an activity pattern of general movement of musculature for each gesture. They then used these activity patterns with a nearest

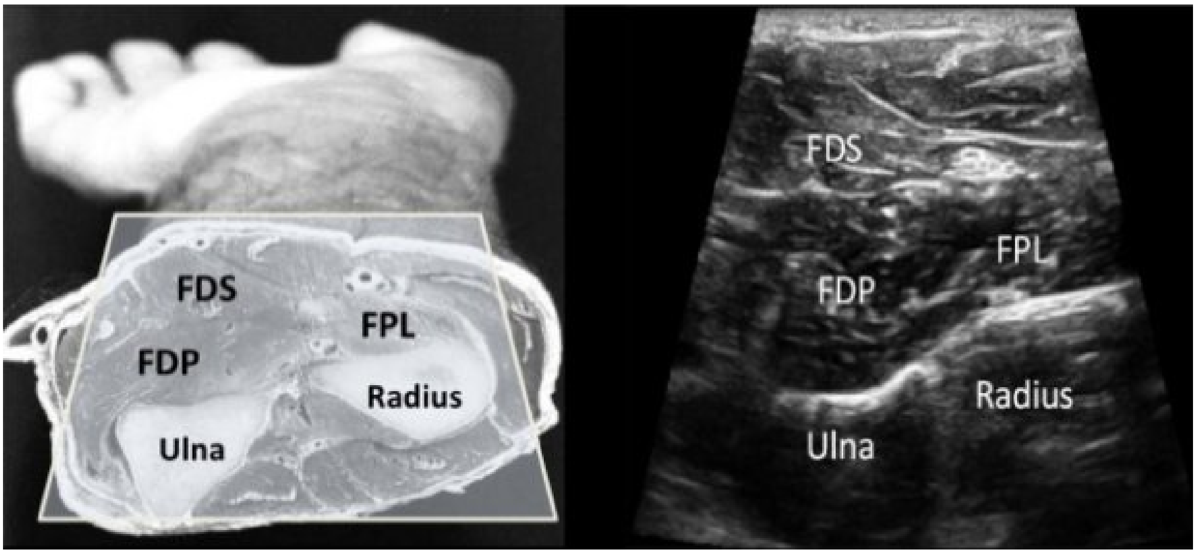

FIGURE 3.9. Cross-section of the forearm muscles and corresponding 2D ultrasound image. The main muscles and their functional compartments are labelled (refer to section 2.1 for anatomical terms). The trapezoidal shape on the left image illustrates the view of the probe, when the probe is placed on top side of the forearm, facing downwards and perpendicularly to the axis of the forearm.[3]

neighbour classifier in order to predict gestures.

An earlier piece of work from the same group focused entirely on differentiating between different digits, and estimating the movement speed of each digit individually [116]. They were motivated to study ultrasonography due to the inherent limitations of surface EMG: A lack of specificity for deep muscles, such as the flexor digitorum profundus. This has implications for identifying fine-grained finger motions that use combinations of the FDP and FDS to flex different joints of the fingers as explained in 2.1. Since deeper sections of the anatomy can be visualised with ultrasonography, it is therefore theoretically possible to identify fine-grained finger movements that flex around different joints. Although they did not try to classify such fine-grained movements, Sikdar et al. should be commended for their identification that ultrasonography could be a potential solution to this problem.

Many variations of work on ultrasonography for inferring hand poses exist. While the concept is roughly the same, different imaging processing and machine learning algorithms are explored. For instance, another group of researchers used an optical flow algorithm to determine the movement of the extensor muscles [113]. Ortenzi et al. conducted a comparative study of features and classification methods on 10 different hand poses [94]. Further to this, they were able to estimate different gripping forces.

[3]Reproduced from "Real-time Classification of Hand Motions using Ultrasound Imaging of Forearm Muscles" by Akhlaghi et al. [3]

3.5.3 Estimating Hand and Digit Angles Continuously

However, ultrasonography is not restricted to recognising only discrete gestures or the strengths of them. The significant advantage of ultrasonography is the capability to determine precise continuous tracking of hand or digit flexion. In simpler terms, ultrasonography visualises the movement of muscles, and this movement correlates linearly to the amount of bending for hands and fingers. This is an aspect of hand tracking that EMG has particular difficulties as is it described in section 3.4 on EMG.

In vivo experiments (performed on living human beings) by Korstanje et al. demonstrated that this technique can be used to continuously track tendons in the wrist [70]. A "speckle" is used to describe the pattern in the ultrasound image that is unique to different anatomical tissues, this is different from properly imaged anatomical landmarks with a clear outline, such as interfaces between tissues or bones. By tracking the speckle that the a particular tendon produced, they were able to track the position within an average error of 0.3 mm (1.6%) using a block matching algorithm.

Castellini et al. estimated continuously changing finger angles by analysing ultrasonic images of the forearm [18] [19], obtaining a normalised root mean square error (NRMSE) of ~2%. This technique was also extended to estimate fingertip forces [47]. They captured ultrasound images at the wrist and divided them into regions. Then, the features from each region were used in a linear regressor to establish a relationship between the features and the angle or force of each finger. In more recent work, it was shown that it is possible to recognise 10 different hand poses and grasps with 80% accuracy, and also 3 levels of force for 4 different grasps (a total of 12 "gestures") with 60% accuracy [94].

3.5.4 Using Non-Imaging Methods

Due to the complexity of ultrasound imaging machines, it may be favourable to use a simpler system. The device could have fewer piezoelectric elements, leading to a lower resolution that may be sufficient. Such a system would use less power and might integrate easier with wearable devices due to simple circuitry and decreased costs.

Another way to mitigate complexity issues would be to use a simpler scanning mechanism. An A-scan is ultrasound terminology for one or several one-dimensional scans, rather than the 2D cross-sectional scans talked about in the previous section. Chen et al. demonstrated that A-scans are in fact capable of tracking the opening and closing of the hand continuously, with an RMS tracking error of 12.9% [22]. They differentiated this technique with their previous work using 2D ultrasonography (sonomyography [133]), by naming this method 1-D sonomyography. In fact, this work used only a single element piezoelectric transducer but was still effective. Indeed, work by [55] expanded on this work by using 8, 5MHz transducers, each pulsing and receiving individually. Their study on 10 amputees showed 0.852±0.04 (r^2 = 72%) mean correlation with the training set for all 5 digit flexions/extensions.

Comparisons between EMG and 1D sonomyography were favourable for the latter in detecting continuous wrist flexion [51] and hand grasping force [52]. For the moment, 1D sonomyography is an appropriate method to detect a reduced set of gestures using a simpler system. It provides limited information and lacks the possibilities offered by modern phased arrays of ultrasound transducers for more detailed imaging.

Mujibiya et al. presented a device that is capable of detecting arm grasps (grasping ones own arm with the other hand) and other on-body touch interactions, using a band of ultrasonic transducers around the forearm and fingers [87]. This device has the advantage of working at much lower frequencies and with fewer transducers than imaging devices, but the drawback is that it can only detect interactions when one hand is touching the arm of the other. However, detecting hand movements involving only one hand is a requirement for certain situations [64].

A technique called vector doppler imaging is an advanced technique that can be used simultaneously with regular 2D imaging. This provides information about the movement of different fluids or tissues in the frame of the image. Especially relevant to our goal is tissue doppler imaging, which can provide precise muscle or tendon velocities at different points in the image [37]. This is a more direct way of measuring muscle movement and alleviates the need to process the images to estimate the movement. The hardware required to do this is not a lot more than what is already contained in regular 2D imaging apparatus. Therefore, the system suffers from the same size and complexity issues mentioned in section 3.5.1.

3.6 Infrared

3.6.1 Fundamental Concepts and Example Usage in Technology

Infrared (IR) is a kind of light that is mostly invisible to naked eye. It refers to light with a specific range of wavelengths that are longer than that of visible light (7̃00nm), but shorter than radio waves (1̃mm). A more technically accurate term is infrared radiation, as it is in fact electromagnetic radiation. However, the term IR light is used loosely because of the similar properties to visible/actual light at similar wavelengths. IR with wavelengths near to visible portion of the electromagnetic spectrum is termed near-infrared (NIR). NIR has wavelengths from around 750nm to 1400nm [15]. A typical application of NIR is for imaging in the dark: sources of infrared illuminate the scene and the reflections can be captured using a camera that is sensitive to NIR. It can also be used to roughly estimate distances between objects: the amount of reflected infrared radiation from an object in close proximity to an emitter can vaguely infer the distance to the reflecting object. This principle is illustrated in Figure 3.10. A pair consisting of an emitter and receiver of light adjacent to each other is known as a photoreflector. Usually for photoreflectors, the emitter is a Light Emitting Diode (LED) and the receiver is a photodiode. As the distance between an object and a photoreflector increases, the amount of reflected light received by the photoreflector decreases due to the inverse square law for light. This law states

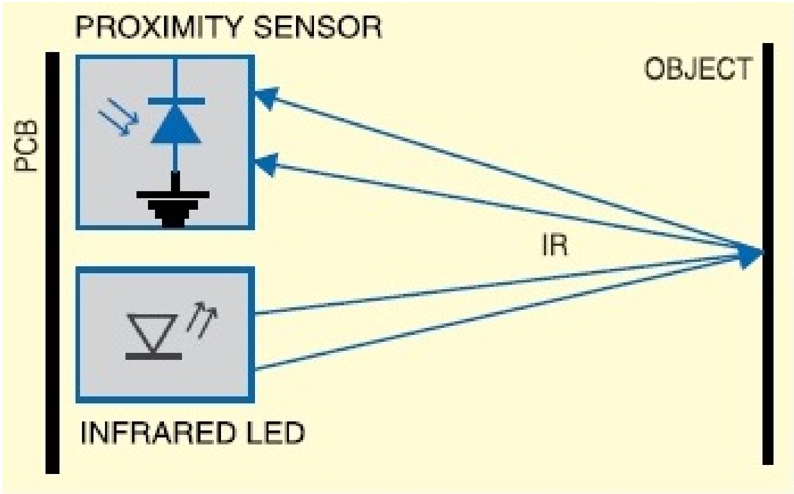

FIGURE 3.10. Proximity sensing using an IR emitter and receiver.

that the intensity of light is inversely proportional to the square of the distance from the source. This technique is commonly employed in mobile phones to detect the proximity to the users head. This is useful for when the user picks up the phone to call somebody, the sensor detects this and turns off the display to save battery and avoid undesired touch input.

3.6.2 Gesture Recognition using Infrared

The deformations of the wrist shape during gesture performances (as mentioned in section 3.1.1) can be measured using this IR proximity sensing technique. A prime example of this is work from Fukui et al. in 2011 [42]. Figure 3.11 shows the design of their prototype device that can accurately sense the deformations of the wrists contour. The device consists of 150 IR photoreflectors, which each sense distance to the skin. As the wrist deforms differently between gestures, these distance measurements also change. The deformations are fairly consistent each time the gesture is performed. This makes it possible to use machine learning techniques on this data in order to classify gestures. Indeed, the authors in this example used a K-nearest neighbour classifier to distinguish between 8 hand gestures with 70% accuracy. More recently, Muhammed et al. used the same approach for rehabilitation and prosthesis control purposes [53]. They were able to detect both finger flexion and wrist motions. They used the term "Optomyography" to describe this method.

A current issue of this particular style of device are that they have been designed so that there is a significantly large gap between the sensors and the skin. This means that integrating this sensor technology in a smartwatch wrist-band is difficult, if not obviously impractical. Current state of the art using this technique has also shown much worse classification accuracy for a

[4]Reproduced from "Hand shape classification with a wrist contour sensor: development of a prototype device" by Fukui et al. [42]

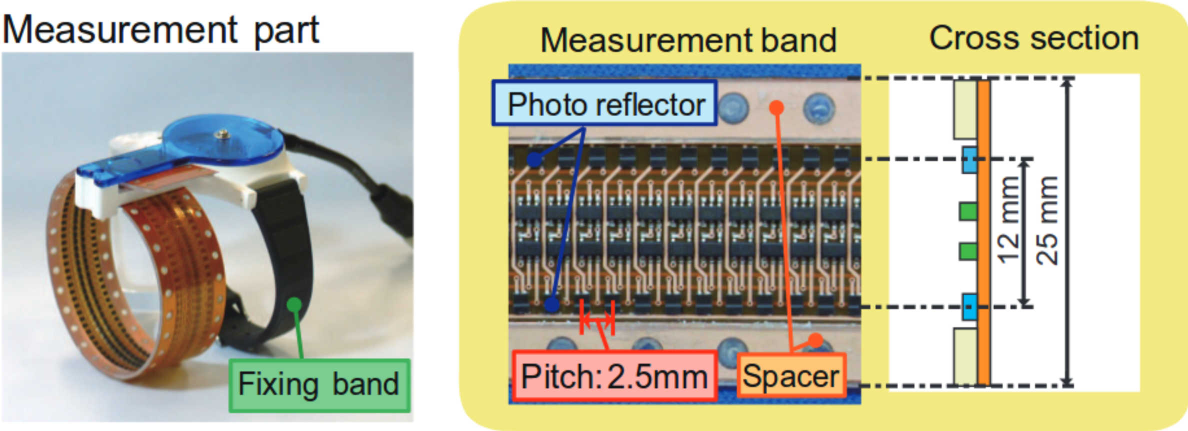

FIGURE 3.11. Fukui et al. created a wrist band consisting of 150 photoreflectors to detect wrist deformations. [4]

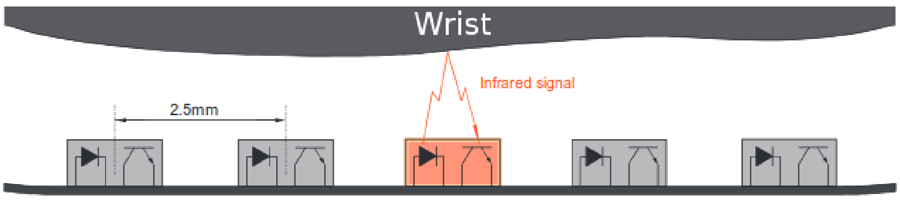

FIGURE 3.12. The photoreflectors measure distance to the skin, which changes as the wrist deformed during gestures.

smaller range of gestures (70% for 8 hand gestures), compared to other techniques such as EMG (27 gestures with 90% accuracy [7]) or Ultrasonography (91% for 15 gestures [116]). However, this technique does have the advantage of having much simpler hardware which makes the device less expensive.

Using the same infrared proximity sensors, Gong et al. took a different approach that allowed them to accurately sense hand abduction/adduction and flexion/extension [46] (i.e. the entire range of rotation of the hand around the wrist joint). In their approach, they instead aligned the sensors so that they were parallel to the forearm, faced towards the hand. The distance measurements could then be used to infer the rotational angle of the hand. They demonstrated it's use for discrete gestures and 2-DoF continuous input. Interestingly, these two different approaches are not mutually exclusive. They may be used to complement each other, although this has not yet been investigated.

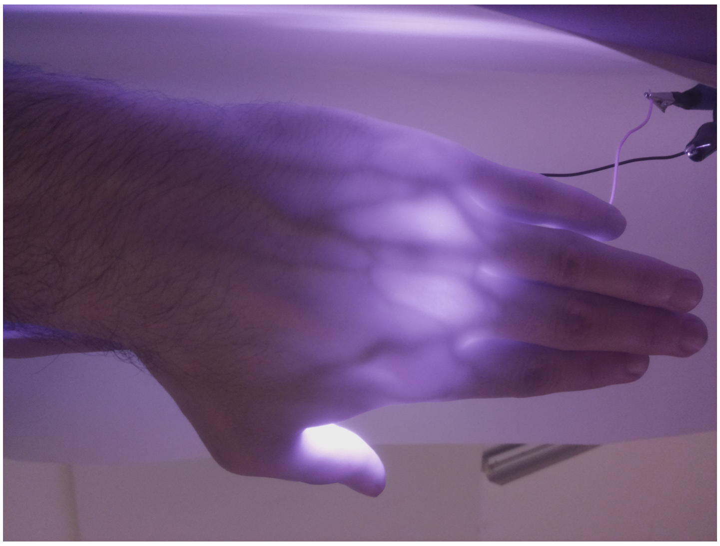

FIGURE 3.13. Image captured by an IR sensitive camera. A source of IR behind passes diffusely through the hand, highlighting the veins.

3.6.3 Transmission of Infrared Through Flesh and Diffuse Optical Tomography

We have so far discussed the use infrared to sense distance using reflections. But as a matter of fact, some wavelengths of near-infrared penetrate through human flesh. This effect can be seen with a simple setup: an infrared camera and IR LEDs placed behind a hand (Figure 3.13). IR interacts with anatomical parts differently. For instance skin, bone and muscle tissue is largely transparent to NIR, but blood has a different absorption coefficient which mostly absorbs the IR. This is why in the photo, the veins can be seen, as the haemoglobin in the blood absorbs IR. Furthermore, haemoglobin that is oxygen bound has a different absorption rate of IR to haemoglobin that is not carrying oxygen. This is in fact the basic principle of pulse oximetry (a method monitoring oxygen saturation). Functional NIR Spectroscopy (fNIRS) also uses this principle, albeit in a much more complicated manner, in order to assess brain activity. These methods are convenient as they are external non-invasive equipment. Additionally, IR is non-ionising radiation and is therefore safe to use.

Diffuse optical imaging (DOI) is an advanced method of imaging that utilises these optical properties of NIR to create scans of the human anatomy. These scans can provide information about the anatomy, or functional information as fNIRS provides. Of relevance to research in this thesis are anatomical scans produced using diffuse optical imaging, such as work by Zhao et al. [132]. Figure 3.14 shows a scan of a humans lower left leg. In this scan, the bones can be clearly distinguished from the surrounding muscle tissue. This method can be used tomographically to produce 3D scans. To my knowledge, there is no previous research that suggests using DOI for interaction with computers or prostheses.

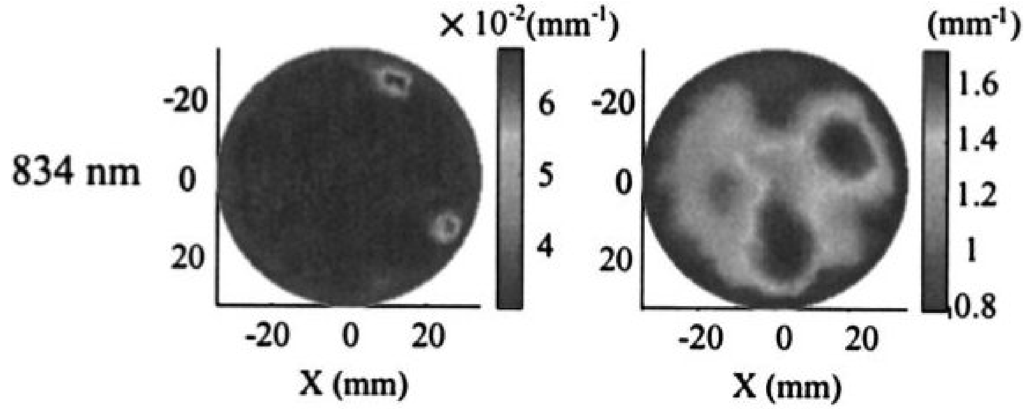

FIGURE 3.14. A scan of a human lower leg, using time-resolved NIR diffuse optical imaging.[5]

3.7 Pressure Sensing

Deformation of the wrist shape (section3.1.1 can also be sensed with pressure sensors. Dementyev et al. presented WristFlex in 2014 which uses 15 pressure sensors around the wrist and a Support Vector Machine (SVM) to classify five pinch gestures with an average accuracy of 80% [29].

Dementyev et al. demonstrated the energy efficiency of the system, measuring an average sensor power consumption of 60.7 µW. Like the infrared technique, this device is designed to be worn around the wrist where it is effective at picking up small tendon movements.

The actual sensors that were used in this work were force sensitive resistors (FSR). That is, a force that is exerted upon the FSR changes the resistance of it. They consist of a thin conductive polymer, which can be screen printed. It's thickness and screen printing option enables easy integration with wearables such as watch straps.

3.8 Electrical Impedance Tomography

[*Throughout the course of my PhD, Electrical Impedance Tomography (or EIT) was published [130]. For this reason, it is not mentioned in my early work, predominantly EMG*]

Electrical impedance tomography is a well known medical imaging technique that has recently been applied to hand gesture recognition. The device that Zhang et al. showed in [130] used 8 electrodes, in tight contact with the skin. By measuring the electrical impedance between each pair, they were able to reconstruct a very vague image of the musculature within. They went on to improve the resolution of the device in [131]. In a similar manner to the wrist deformation, hand gestures create changes in the musculature of the forearm. This allows them to recognise

[5] Reproduced from "Time-resolved diffuse optical tomographic imaging for the provision of both anatomical and functional information about biological tissue" by Zhao et al. [132]

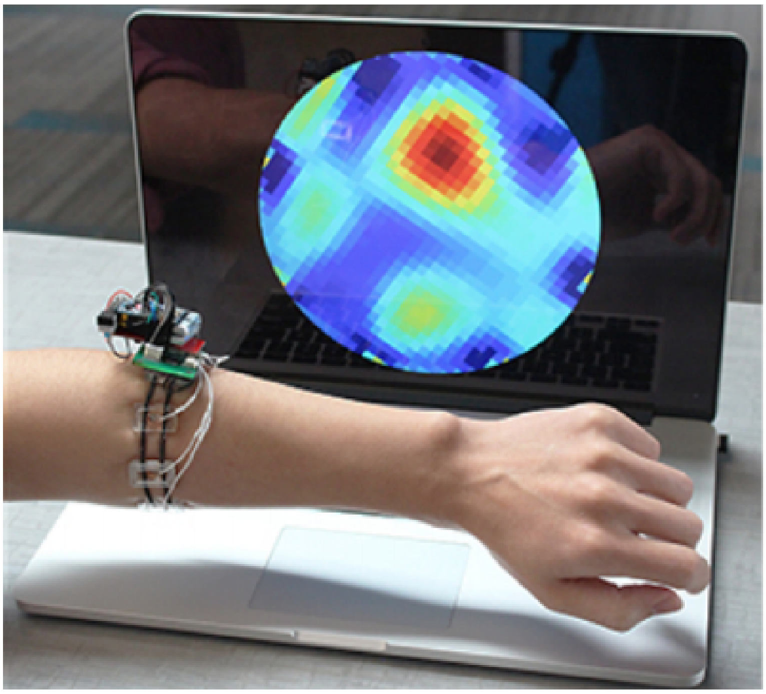

FIGURE 3.15. Zhang et al. created a wrist band consisting of electrodes and measuring the impedance's between them to detect wrist deformations. [6]

11 discrete hand gestures with a high accuracy (94.3%), however continuous tracking has not been tested yet.

3.9 Mechanomyography

Mechanomyography (abbreviated to MMG) is a technique which records vibrations of the muscle fibres after contraction, which oscillate at their resonant frequencies and is characteristic of a specific muscle's activity [93]. Although mechanical vibrations of this kind do produce sound, it is important to note that it is not an imaging technology like ultrasonography. Mechanomyography supports prosthesis and switch control since the placement of the sensors does not have to be precise and the change in the skin impedance due to sweating does not affect the performance [60]. Although this technique uses sound, it is not an imaging technique and has been under-explored within the HCI community (with the exception of [128]) despite its simplicity in hardware and robustness to sensor placement or skin condition. Currently, studies thus far do not show high classification rates for a range of gestures yet. Because this technology has not yet matured, this will not be considered as an established method for analysis in the next section.

[6]Reproduced from "Tomo: Wearable, Low-Cost, Electrical Impedance Tomography for Hand Gesture Recognition " by Zhang et al. [130]

3.10 Inertial Measurement Units

Inertial Measurement Units (IMU's) are units that contain accelerometers, gyroscopes and magnetometers. It is possible to accurately track the position and orientation of a device using these sensors. IMU's can provide data to enhance gesture classification [127]. However, data collected from a wrist-mounted wearable can't be used to detect rotations about the wrist joint, because the hand can rotate separately to the forearm. Many gestures change meaning significantly based only on localised wrist rotation, for example the difference between pointing gestures in different directions, or the difference between a 'thumbs up' and a 'thumbs down'. Researchers have addressed this issue by placing accelerometers on the hand in rings or other hand-mounted wearables [50], but this requires additional hardware over a single wrist-worn wearable. Wrist-mounted IMUs can detect arm movement in a wider coordinate system, and can help to detect supination and pronation in the forearm but it is still difficult to differentiate between rotation of the whole arm and forearm, as the locally measured movements are very similar.

Although IMU's can be used to facilitate the position and orientation of a device, it has very recently been shown that it can in fact detect gestures. With a high-frequency accelerometer it was possible to sense gestures of high kinetic energy (i.e. pinching, flick, snap, wave up) [74]. Also, using the sensors integrated in a commercial smartwatch it was possible to detect 5 gestures with 87% accuracy [123]. This is an inexpensive and easily available alternative, but the gesture set and accuracy these works have demonstrated are very limited, and requires high consistency in the way they are performed. Due to the infancy and somewhat limited use of IMU's in hand gesture detection, we will be leaving this technology out of the analysis too. However, IMU's can facilitate the detection of gestures that relate to the world spatially. Because of this, most of the work in this thesis is geared towards detection of gestures that are iconic or representational. This requires detection of the hand pose, which IMU's struggle to detect based on past work.

3.11 Analysis of Gesture Recognition Techniques

Having gathered an overview for the main established methods for hand tracking / gesture detection, we will now attempt to analyse these with respect to the baseline requirements listed in section 1.4: Mobility, placement & size, accuracy & range of gestures and complexity.

3.11.1 EMG

Mobility Wearable EMG devices have been created in the past [72, 85]. The circuitry can be complicated and the amplification requires significant amounts of power. This scales in proportion to the number of EMG channels used. The "Myo", developed by Thalmic labs, is a wearable EMG device worn around the proximal forearm. This device uses 8 channels and can reportedly operate for the duration of a whole day [73].

Placement & Size The placement and number of EMG electrodes can be varied. Performance increases with number of electrodes, at the cost of practicality due to a larger surface area. Placement at the wrist is suspected to perform worse due to less muscle mass. Previous work has shown EMG to work in a small band [106].

Accuracy & Range of Gestures The range and accuracy is similarly expected to increase proportionally to number of electrodes. Few electrodes can only recognise a few gestures but arrays of hundreds are able to detect gesture sets that represent a large range of human hand motions (still discretised) [7].

Complexity The acquisition circuitry can be complex due the low amplitude bio-signals that need amplifying without introducing noise. The complexity also increases with the number of EMG channels.

3.11.2 Ultrasound

Mobility Portable ultrasound imaging systems have been developed, but they are still too large for integration with wearable form factors. This is partly due to battery requirements and circuit board space.

Placement & Size The amount of surface area that the probe of the ultrasonic imaging device occupies is actually small. The sensor is usually designed to be long and thin, which can match the form factor of a wrist-band quite well. The placement of the probe can vary a lot. It is not yet known which placement is best, but Castellini et al. have found it to be effective at the wrist [18]. The trade-off of accuracy vs location is still unknown.

Accuracy & Range of Gestures Previous work has demonstrated superb range and accuracy [3].

Complexity This is the most complicated of all systems. As with EMG, complexity scales with the number of piezoelectric elements. However, there are several differences. Firstly it requires a high number of elements to attain an image at all. Secondly, the ultrasonic signals are much more difficult to detect than electromyographic signals, requiring complicated receiving circuitry. Unlike EMG, beam-formed ultrasonic signals must also be produced with the array which adds another layer of complexity.

3.11.3 Infrared

Mobility This technique is highly mobile due to low power and small form factor.

Placement & Size Devices created using infrared have had a particular focus on wearability and the form factor typically tends to be wrist-band shaped. It is therefore an ideal technology from a portability perspective. Have been tested primarily on the wrist due to visible changes of the tendons.

Accuracy & Range of Gestures The range and accuracy tends be poorer than other methods. This is due to mostly observing changes from the surface of the forearm, where deeper muscle movements go largely unnoticed. The signals are also weaker and harder to correlate with muscle movement.

Complexity Past designs for these systems have been very simple, only measuring reflected light intensity.

3.11.4 Pressure

Mobility Similar to infrared devices, this is a highly mobile technique. Power requirements are potentially lower than infrared.

Placement & Size As with infrared devices, they are usually designed with mobility in mind and are wrist-band shaped [29].

Accuracy & Range of Gestures The range and accuracy is again poor for pressure, with them both being an indirect method for measuring muscle movement through changes in wrist shape.

Complexity Device complexity is similarly simple, only requiring measurements of resistance.

3.11.5 EIT

Mobility The second iteration of the device Zhang et al. produced showed that it was possible to integrate the electronics into a small size [131]. The power consumption of the device is also low. These factors make the device very portable.

Placement & Size The device has been tested at the wrist. As above, the size of the device is very small.

Accuracy & Range of Gestures Their prototype demonstrated good accuracy for a range of hand gestures, somewhere between infrared and EMG.

Complexity The algorithm to reconstruct images is complex, but on the other hand, the hardware is relatively simple.

3.11.6 Optical

Mobility Optical systems tend not to be designed with mobility in mind. However, some wearable form factors have been designed (on wrist and glasses).

Placement & Size As explained in section 3.2, the protrusion of the camera makes it conspicuous as a wrist-worn device.

Accuracy & Range of Gestures This is one of the best methods for real-time full hand tracking. The only drawbacks (ignoring ethical and legal issues) being occlusion from other objects or ones self. This is especially true for wrist-worn systems where the palm of the hand usually occludes view of the fingers. Because of this, the range of gestures is diminished.

Complexity Depending on the hardware used, the system can be very simple. The algorithms can also vary from very complicated to very simple. Although in order to gain high fidelity without markers on the hands, it is necessary to use sophisticated algorithms. For practicality, markers on the hands should not be considered, and therefore complex algorithms need to be employed in a wrist worn form factor.

3.11.7 Glove

Mobility Simplistic sensing mechanisms allow for a small computer to be embedded within the glove. It is therefore adequately mobile.

Placement & Size This is standard, but the thickness can vary, which affects the encumbrance of the glove. They are almost certainly obtrusive given that it can hinder everyday interactions with other objects.

Accuracy & Range of Gestures The range and accuracy of these methods are perhaps on par with optical methods. The advantage of this method is that it is not affected by occlusion.

Complexity There a multitude of different sensing technologies used in glove form factor hand trackers. Most of them tend to be simplistic, for example using flex sensors. These require low sampling rate analogue signal acquisition units which are trivial.

3.12 Summary and Further Analysis

A summary of the analysis is shown in table 3.1. For the sake of brevity, fidelity encompasses both accuracy and range of gestures. There is a clear overlap between the mobility and complexity requirements. This is because they have similar factors such as circuit size and complexity, which increases the battery power required. As stated previously, some of these factors are likely to

Technique	Mobility	Placement & Size	Fidelity	Complexity
EMG	Battery concerns	**Variable**	Good	Variable
Ultrasound	Improving but still poor	**Variable**	Very good	Very complex
Infrared	Highly mobile	Thin & wrist	Poor	Very simple
Pressure	Highly mobile	Thin & wrist	Poor	Very simple
EIT	Highly mobile	Thin & wrist	Average	Fairly complex
Optical	Can be mobile	Wrist, but obtrusive	Limited*	Fairly complex
Glove	Highly mobile	Obtrusive	Excellent	Simple

Table 3.1: A summary of the analysis of established gesture recognition techniques, with respect to the initial requirements. **The range of gestures are poor at the wrist.*

improve in future. Consequently, a reduction in the complexity of the device should also facilitate device mobility. The same logic should apply to finding ways to manufacture integrated chips to reduce size and power requirements of devices.

The optical and glove methods are both inherently obtrusive methods that do not conform to the ideal form factor we initially set out to investigate. For these reasons, we will be largely ignoring them for further exploration, but acknowledging the supreme accuracy for full hand tracking that these devices enable. The ultimate goal of this thesis is to explore methods for tracking the hand accurately in a wrist-worn form factor, with a resolution that matches those of optical and glove devices. Work in this thesis achieves steps towards this goal in two ways: by improving the effectiveness of existing hand trackers at the wrist, but also finding innovative improvements to those sensing methods on a hardware level, all while still considering the practicality.

The infrared and pressure methods are near identical in terms of how well they fit the requirements, with subtle differences mentioned before this summary. Electrical impedance tomography (EIT) is also similar to these techniques but has demonstrated better range and accuracy of gestures. As it stands, these three technologies are the only ones that can currently be considered to satisfy the primary concerns of mobility and placement. It is unfortunate that these methods do not show the same level of tracking fidelity that EMG or ultrasound demonstrate.

EMG and ultrasound methods found in literature vary in placement and size, significantly. The fidelity of which varies with respect to the placement and size, but is generally much better than infrared or pressure. It is uncertain whether using EMG at the wrist is a possibility, or even if it still provides a decent gesture recognition accuracy. Ultrasound on the other hand is also variable, but has been tested on the wrist. However, the loss in quality compared to different probe orientations or locations proximal are untested. The problem of complexity and mobility of ultrasound imaging systems are also not addressed in previous literature. There is very little interest in the industry to create wearable ultrasonographic imaging apparatus (namely, the healthcare industry). Therefore, to my knowledge, there is no exploration of research that looks at the feasibility of ultrasonic imaging devices and possible reductions of the complexity of them at the cost of fidelity.

These are missing gaps that prevent us from writing a more concrete table of methods vs requirements. To fill these gaps in literature are the initial steps of my work. Starting with chapter 4, the feasibility of EMG at the wrist and the unknown fidelity achieved at this position are investigated. Although it was not intentional, the battery concerns of EMG are facilitated by the integration of simpler, more power efficient pressure sensors in the design of our prototype. This happened as a result of a pilot study which suggested that motion artefacts were beneficial to the classification of hand gestures, which pressure sensors could measure.

CHAPTER 4

ELECTROMYOGRAPHY

4.1 Introduction

This chapter explores variation in sensor placement of EMG devices to fill a gap in the literature: to find out the presumed trade-off in fidelity as the placement and size varies. Practical wearable gesture tracking requires that sensors align with existing ergonomic device forms. Thus, of particular interest to us, is whether an EMG device can be integrated with wrist-worn form factors.

This chapter begins with an initial study to investigate the performance of EMG at the wrist. The results from this study exhibited surprisingly high EMG classification accuracy in the wrist condition. This high accuracy led us to suspect that variable wrist pressure on the EMG electrodes was modulating and enhancing our EMG finger gesture classification rate. We theorised that these electrode motion artefacts introduced additional features into the collected data which improved classification accuracy. From the unexpected results of the pilot study, we hypothesised that pressure data on the wrist strap could provide our classifier with features that enhance the EMG results.

To test our hypothesis, we built the EMPress system, which uses EMG and pressure sensors, drawing on previous work on low-power gesture input with Force Sensitive Resistors (FSR)[29]. With explicit and separate pressure sensing in the gesture detection process, we aimed to isolate and quantify this effect in a second study. The prototype we developed for our second study uses padded wet electrodes to moderate pressure effects on the EMG signal, and a cross-arm reference electrode to determine an upper bound of EMG performance.

The end result of our final study showed that combining EMG and pressure data sensed only at the wrist can support accurate classification of hand gestures. The EMPress technique senses both finger movements and rotations around the wrist and forearm, covering a wide range of gestures, with an overall 10-fold cross validation classification accuracy of 96%. We show that EMG is especially suited to sensing finger movements, that pressure is suited to sensing wrist and forearm rotations, and their combination is significantly more accurate for a range of gestures than either technique alone. The technique is well suited to existing wearable device form factors like smart watches that are already mounted on the wrist.

Although it was not our primary objective, the battery concerns of EMG are facilitated by the use of simpler, more power efficient pressure sensors. This happened to be a byproduct of an experimental sensor fusion prototype used for investigating motion artefacts.

Our key contributions are:

- A novel design combining EMG and pressure data using machine learning to accurately detect and classify hand gestures.
- Two studies which identify and then quantify the performance of combining these sensors in the prototype EMPress system.
- Experimental evidence that these sensors are strongly complementary, emphasising EMG for detecting finger movements and FSR for detecting wrist movements.

In the following sections we explain the relevant forearm anatomy which supports our wrist-based EMG approach. We then describe the different types of sensors which we use across the two studies in this paper.

The content of this chapter is a revised version of work I have published in the Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16): "EMPress: Practical hand gesture classification with wrist-mounted EMG and pressure sensing" [82].

4.2 Background

4.2.1 Electromyography

This section gives a brief overview of the EMG technique, and the relevant anatomy. A more in-depth description of EMG and the anatomy can be found in chapter 2. However, what can be found in this section that is not shown in the earlier chapter, are the groupings of muscles and tendons. These groupings are illustrated in the diagram via colour coding. These groups are used later in the chapter to clearly show the link between the grouping of gestures based on the anatomy.

When a muscle is contracted, an electrical potential difference is created by the electrically or neurologically activated muscle cells. Surface Electromyography (EMG) measures the difference with electrodes on the skin close to the muscles of interest, which can infer muscular activity. EMG data can be used to determine which muscles are active and even the amount of force they produce. Well-placed sensors are key to identifying patterns of EMG signals, which relate to

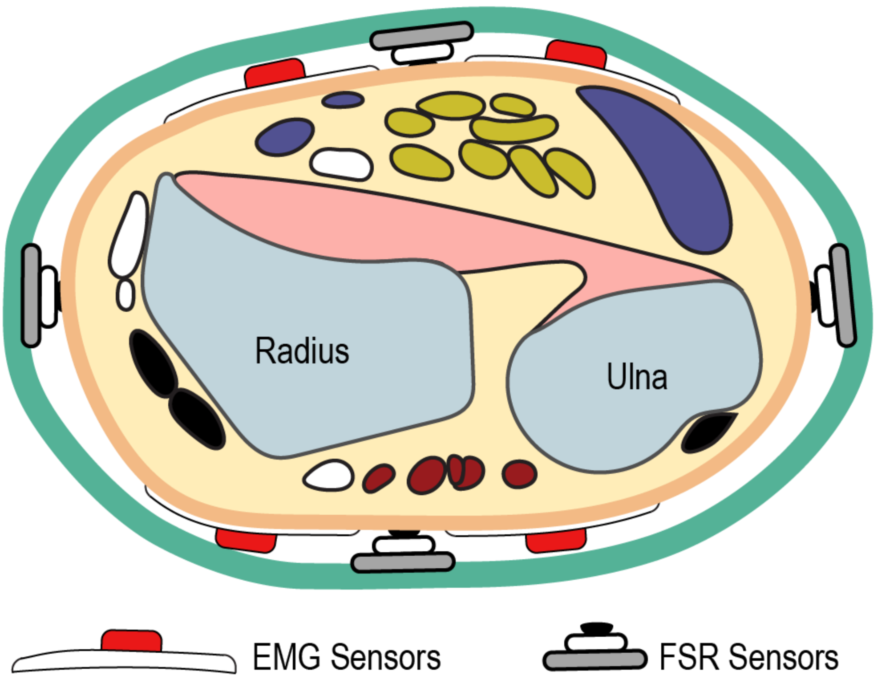

Figure 4.1: Cross section of the forearm at the distal ends of the radius and ulna. Our final experimental prototype is seen around the wrist. The anterior of the arm is at the top. The muscles and tendons are coloured to represent groups with similar functionality.

specific movements of muscles. However, there are comparatively few muscle cells in the wrist compared to the proximal upper forearm (near the elbow), and the distal tendons (near the wrist) are more difficult to discriminate as they are more tightly packed. Thus there are increased challenges in performing EMG sensing of hand muscle movements by sensing at the wrist. As a result, current off-the-shelf solutions such as the Myo armband typically capture EMG signals from muscles in the upper forearm [72].

Figure 4.1 shows a cross section of the wrist, the position for a typical wearable strap. There are three main groups of muscles which are responsible for the flexing and extending of the fingers. The flexor digitorum superficialis and flexor digitorum profundus (Fig 4.1, top) flex the fingers. The extensor digitorum communis, extensor indicis proprius, extensor medii proprius (Fig 4.1, bottom) extend the fingers, and aid a little to extend the hand. These muscles control all fingers except the thumb, the muscles responsible for this are known as the pollicis muscle group (Fig 4.1, left). As these muscles flex both the whole hand and the fingers it can be challenging to differentiate between gestures using EMG, for example flexion of the wrist and flexion of all fingers. The flexor carpi radialis, and the flexor carpi ulnaris (Fig 4.1, top left/right) control hand flexion at the wrist, and control hand abduction and adduction, respectively. Similarly, the extensor carpi ulnaris/radialis (Fig 4.1, bottom left/right) extend the hand at the wrist joint, and are also capable of abduction/adduction of the hand.

Commonly, bipolar electrodes are used to measure the electrical potential generated by the muscles in EMG systems. Normally, three electrodes are attached to the skin, two within close proximity of one another and another reference electrode to an area with less muscle activity. The signal can then be acquired by measuring the output of a differential amplifier, using the bipolar electrodes as input, removing any common noise that is measured by the reference electrode. Electrodes can be 'wet' (mounted onto the skin with adhesive and conductive gel) or 'dry' (without gel or adhesive). Recent designs for dry electrodes have comparable accuracy to wet electrodes, and are a more practical alternative [88].

Prototype EMG sensors can be designed in high density arrays worn on the forearm. These designs demonstrate excellent recognition rates for finger movements [7] and even for wrist movements [44, 127]. However, these sensor arrays require a myriad of electrodes spread across a significant proportion of the arm's surface and may be more appropriate for integration into clothing than wearable devices. Finger muscle movement can also be captured using targeted wrist-mounted EMG sensors. For robustness these are commonly used in conjunction with additional EMG sensors mounted on the hand and/or distal forearm eg [21], making the overall configuration impractical for integration into a single wrist-mounted wearable. The use of EMG sensors in practical wearable scenarios typically also requires calibration in order to account for slipping watch straps and to align sensors with the anatomically optimal detection points. While we do not consider the issue of calibration in this paper, shift compensation algorithms [7] could further improve the results of our work.

4.2.2 Hand Gesture Sensing Techniques

Given the challenges of accurate EMG sensing, a number of other techniques have been applied to wearable gesture recognition. In the related work chapter, we gave an extensive list of the most prominent techniques to date.

We've selected to talk about pressure sensing for gesture recognition here, as it's highly reveleant to this chapter. Recent work has explored whether Force-Sensitive Resistors (FSRs) can provide a useful additional channel of data from a wrist-mounted wearable. FSRs have two copper traces that sandwich a special type of conductive polymer in between, which decreases in resistance as force is applied to it. By measuring the resistance of the resistor (e.g. by using a voltage divider), the amount of force applied can be inferred. Superficial tendons will move as hand gestures are performed, and because of their proximity to the surface of the skin, the movement can also recorded by FSRs to classify certain gestures. WristFlex [29] used an array of FSRs in a wearable wrist strap to detect finger pinch gestures based on the subtle tendon movements in the wrist. The authors found a high classification accuracy with low power consumption.

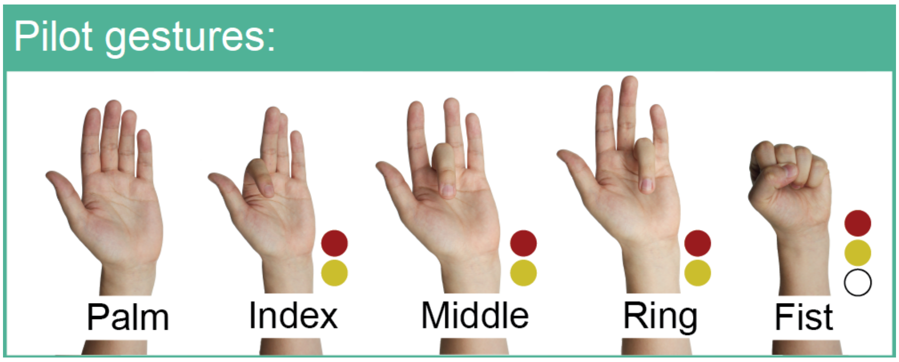

Figure 4.2: Gestures used for the pilot study.

4.2.3 Gestures

The types of gestures that we wish to classify are shown in Figure 4.3. The coloured dots next to each gesture indicate the corresponding coloured muscles in Figure 4.1 that are predominantly responsible for producing the gesture, starting and finishing with the Palm hand posture. Most gestures are categorised by their muscle groups.

While there is no standardised hand gesture taxonomy, most gesture techniques described above evaluate against variations of general movements which draw on finger and wrist rotations. For our pilot study, we try to classify a set of gestures which contain several finger gestures, a fist and a palm gesture. Finger gestures and Wrist gestures are significantly different due to the different muscle groups which are required, illustrated by the colour variations between the gesture classes. For our follow-up study, we grouped and supplemented these gestures to explore a more challenging set of 15 gestures in three classes:

Finger Gestures These are gestures which only involve movements of the fingers. Anatomically, any movement which only involve rotations of the phalanges around their phalangeal joints will be classed as a finger gesture. In this class, we want to test whether individual fingers can be discriminated, and therefore include single finger flexions.

Wrist Gestures Hand movements which rotate the whole hand around the wrist joint are classified as wrist gestures. Although supination and pronation of the forearm occurs because of rotations at both the wrist and elbow joint, we include them as wrist gestures because they rotate the whole hand.

Other Gestures This set of gestures are not naturally focused only on single fingers or the wrist, consisting of whole-hand gestures that use multiple fingers simultaneously.

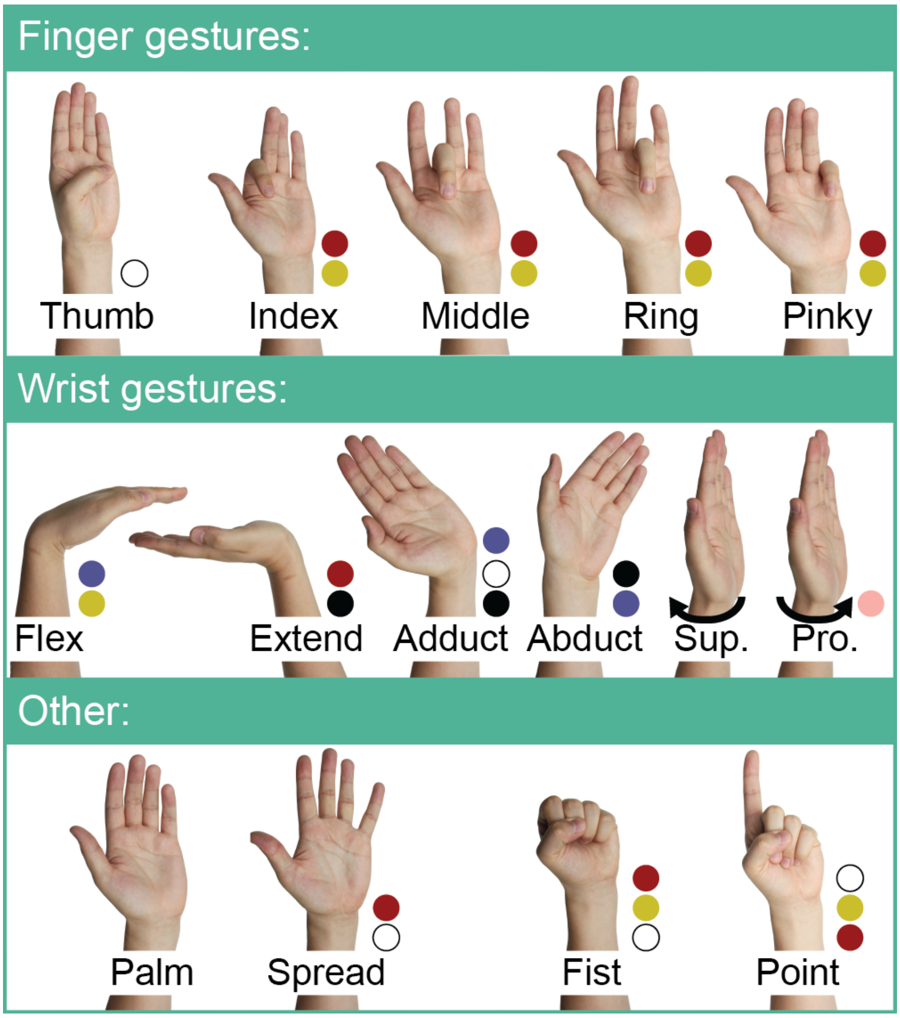

Figure 4.3: The entire set of gestures that were used for the latter study. The coloured dots represent the muscle groups shown in Figure 4.1 which are predominantly used during the gesture.

4.3 Pilot Study

4.3.1 Motivation

The aim of our pilot study was to compare the effectiveness of hand gesture recognition using EMG when using sensors located on the wrist in comparison to another device located on the proximal upper forearm. We were especially interested in how the relocation of sensors would affect the classification rates of finger gestures. We did not include wrist gestures in our pilot, as we expected to rely on existing studies that have measured wrist gesture performance with EMG [1]. The gesture set that we tested for included the following from Figure 4.2.3: Palm, Index, Middle, Ring and Fist. The palm gesture represents a relaxed state with no muscle activity.

4.3.2 Hardware

Hand-gesture recognition using electromyography is more challenging to implement around the wrist than if the sensors were located at the proximal end of the forearm, for the reasons explained previously, including smaller surface area, reduced muscle mass and the muscles being closer together.

Commonly in EMG, each muscle is individually measured. However, due to the constraints set by the surface area of the wrist and considering practically, we decided to use just two pairs of bipolar electrodes. The positioning of these electrodes is crucial, as they both need to have good coverage over different areas of the flexor digitorum muscles. The number of muscles in the forearm which we wish to detect are far greater than the number of data channels the device provides. The underlying principle for detecting which muscles are active in this situation is as follows.

Several muscles are measured by a single electrode. The sensitivity of an electrode to any particular muscle is proportional to it's distance away from the electrode's conductive centre. Since the values that are collected are roughly consistent each time a gesture is performed, a machine learning classifier can predict the gesture. The muscles of interest here are the flexor digitorum superficialis and profundus, and therefore we placed our 2 pairs of sensors orthogonal to these muscles, on the anterior of the forearm, so that each pair of sensors are more receptive to certain fingers. Figures 4.4 & 4.5 show differences in signals when flexing different fingers, for two bipolar electrodes placed horizontally across the wrist, as per the placements in Figure 4.1. The graphs show that each sensor is indeed more receptive to a particular muscle, as confirmed by the differences in amplitude.

The reference electrode was placed to the side and on the posterior of the forearm, close to the ulna. Here there are fewer muscles and thus less muscle activity to be picked up by the electrode, making it the ideal for a wrist-mounted device to place a ground electrode. The electrodes are held in place with an elastic strap.

For this prototype, we created our own electrodes using conductive metal pins. It was necessary to apply gel to the electrodes for good electrical contact with the skin.

The electrodes are connected to circuit boards that apply a differential amplification, rectification and smoothing of the signals from the bipolar electrodes [120]. Each of the four boards are connected to an Arduino Uno. The Arduino is programmed to read analogue data from the sensors, which are then sent to the computer via serial communication for data collection and processing. The data is collected at approximately 60Hz.

4.3.3 Software

The data samples recorded for each channel from the device is of the form of a 1D time series. We extracted several time-domain features from each sensor in a given recorded sample. We used a support vector machine to classify the gesture data that we recorded. We chose to use

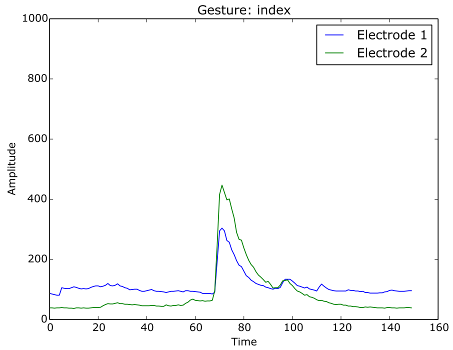

Figure 4.4: Index flexion EMG signals

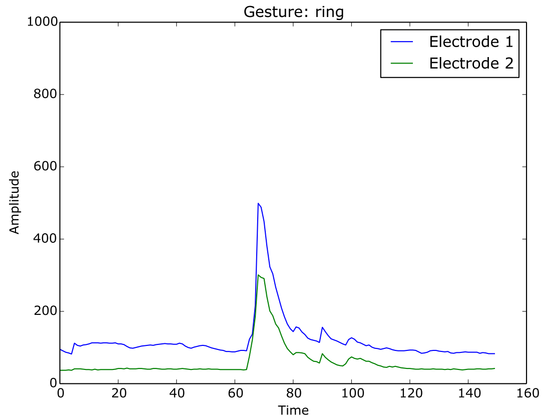

Figure 4.5: Ring flexion EMG signals

Figure 4.6: EMG signals from flexing a) the index finger and b) the ring finger

an SVM over alternative classifiers because we knew our feature space would be small, and the kernel that the SVM uses will increase the dimensionality of the feature space. We also found that previous work on hand gesture recognition using EMG demonstrate that SVMs yield good performance [6, 127]. We chose to use Libsvm, a simple yet efficient library for support vector machines [20]. After extracting the features, we normalised the values of each feature vector, since an SVM is not scale invariant. We used C-Support vector classification with a *radial basis function* kernel due to it's better classification performance over linear kernels [63], although a linear kernel would be faster if interactive feedback were required from a real-time wearable system with limited processing capabilities. The C-SVC classifier implemented by Libsvm uses a "one-versus-one" method for multi-class classification.

For the data collected with each device, we used the k-fold cross validation technique in order to find suitable parameters for the SVM classifier. Using these parameters to train the SVM on the training data, we could then use the classifier to predict each testing instance, and compare this to the training label to find out if the outcome was correct. The percentage accuracy of the classifier for each participant then equates to the number of correctly classified instances divided by the total number of instances.

4.3.4 Procedure

We used a within-subjects experimental design, with the independent variable being the placement of the electrodes, and dependent variable being the accuracy of the SVM classifier. We kept the feature set and the supervised learning implementation the same in each iteration of the study.

12 participants (9 male, 3 female) participated in the study, all participants were healthy with no known muscle impairment. The study consisted of a 30 minute session, during which the

	Predict Palm	Predict Ring	Predict Middle	Predict Index	Predict Fist
Actual Palm	10	0	0	1	1
Actual Ring	0	6	0	5	1
Actual Middle	0	0	11	1	0
Actual Index	0	0	2	10	0
Actual Fist	0	0	0	0	12

Table 4.1: Confusion matrices for the predicted gestures of the wrist-worn device.

	Predict Palm	Predict Ring	Predict Middle	Predict Index	Predict Fist
Actual Palm	8	2	0	2	0
Actual Ring	0	3	0	2	7
Actual Middle	0	2	6	0	4
Actual Index	0	1	0	4	7
Actual Fist	1	0	0	0	11

Table 4.2: Confusion matrices for the predicted gestures of the forearm-worn device.

participant would perform gestures while wearing one device worn at the wrist, and then again with the second device on the upper forearm. During each of the 5 gestures, the device would record the data from each bipolar differential channel. The gesture would be repeated multiple times over the course of the session.

4.3.5 Results

The results of the pilot study are shown in Tables 4.1 and 4.2, as confusion matrices of each predicted gesture against their truths.

A Shapiro-Wilk test showed that the classification sample was not normally distributed ($W=0.731$, $p <0.05$), so a non-parametric Wilcoxon Signed Ranks Test was applied to compare the devices' classification accuracy. This test indicated that the classifier was significantly more accurate for the data collected from the device worn on the wrist (mean 82% correctly classified) than for the device located on the upper forearm (mean 53%), $Z = -2.359$, $p <0.01$.

4.3.6 Discussion

Our pilot study demonstrated a significant increase in classification accuracy when the electrodes are worn on the wrist compared to the proximal forearm location. While we hoped that the new placement might prove comparable to the usual upper forearm placement for finger gesture detection, we were surprised that the difference was so significantly in favour of the wrist, which we believed should be more anatomically difficult to discriminate the EMG signals.

This result led us to explore other potential reasons for the increased performance of the classifier in the wrist condition. Looking at the differences in the system design, one key area was the home-made electrodes used under the wrist band. During some sessions, we found that the

Figure 4.7: An Image of the prototype worn around the wrist.

resting amplitude of the EMG sensor data varied within the same session. One reason for this could be a change in electrical contact with the skin. This change can be caused by a displacement of the electrodes, due to movements of the forearm. Upon further experimentation with the device, we came to believe that the increased performance could be attributed to changes in pressure on the EMG electrodes from the elastic wrist strap. The pressure was modulating the EMG signal in ways which we suspected may be providing additional gestural features to the classifier than pure EMG signals alone.

The hypothesis that wrist pressure was providing predictable features by modulating our EMG data was rather surprising. Existing studies in the literature directly using wrist pressure to detect hand gestures had selected specific movements to ensure classification accuracy, such as the use of a finger pinch in the WristFlex study [29]. Nonetheless, we hypothesised the wrist pressure was the most likely variable in the higher classification rate. In order to test this hypothesis, we conducted a second user study in which we isolated the pressure and EMG data collected only on the wrist, in order to quantify the effect of the pressure changes in the wrist on the gesture classification results.

4.4 Main Study

This section describes the design process of the device for our main study. The following section describes the study and results.

4.4.1 Hardware

The changes in pressure applied to each electrode from the strap are present for a number of reasons. The key explanation is that muscles/tendons in the forearm become displaced upon contraction. Stretching of the skin can also affect the pressure between the sensor and strap. Factors of the strap design such as the elasticity, can significantly change how the movement

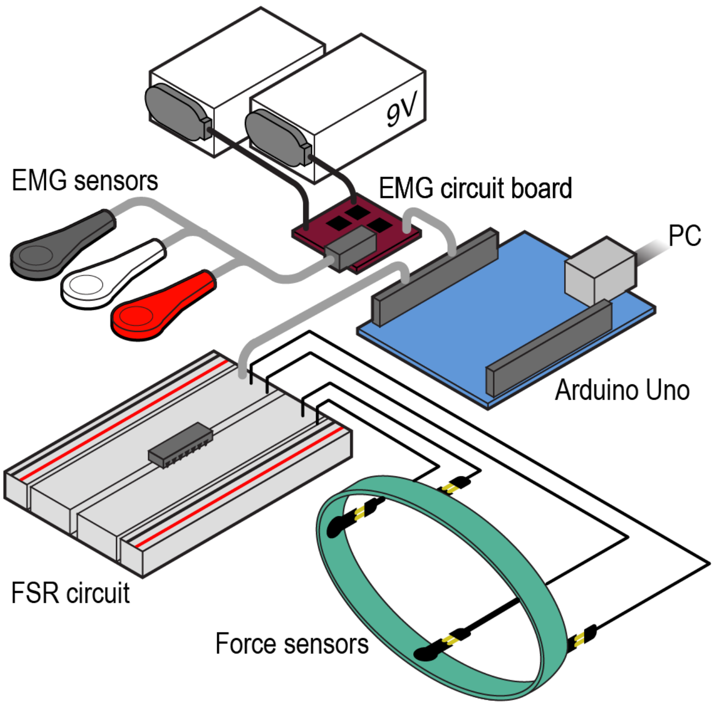

FIGURE 4.8. Diagram of the hardware components used in the prototype.

of the skin affects the pressure, as we found out during trial and error of different strap types. Eventually we decided that a simple elastic band would suffice.

In order to measure the pressure between the strap and the wrist, we chose to use the same Force Sensitive Resistors that are used in the WristFlex prototype [29]. The FSR400 component [35] provides good performance in their study and our requirements are similar: inexpensive, small and highly sensitive.

In order to isolate the pressure data from the EMG readings, we chose to use a more robust EMG sensor (SKINTACT Ag/AgCl aqua-wet electrode, ref: FS-TF). 'Wet' ECG electrodes that are commonly used for medical purposes have a design which mitigates the effect of pressure, due to the use of highly porous foam beneath the sensor. The electrodes are connected to circuit boards via shielded cables with snap connectors. The circuit boards apply differential amplification, rectification and smoothing of the signals from the bipolar electrodes [120]. For each pair of bipolar electrodes this provides a single channel of EMG data. Each of the four signal processing boards are connected to an Arduino Uno.

The prototype for our second study extends the capabilities of the device used previously. We kept the same configuration of 2 bipolar EMG electrodes located on the anterior side of the forearm, because of the reasonable accuracy for single finger gesture recognition that our first study showed is attainable.

The increased gesture set which we aim to identify in this study includes gestures other than finger gestures. Some gestures require the use of the muscles located in the posterior compartment of the forearm (Figure 4.1, bottom), as indicated by the coloured dots in Figure 4.3. Chiefly, these are the Extend, Adduction, Abduction, Spread and Point gestures. Therefore to detect the activity of these muscles, we decided to place two additional bipolar sensors close to the extensor muscles on the posterior of the forearm. This has the added effect of detecting muscle activity from the extensor digitorums, thus facilitating the classification of Finger gestures. The muscles that are active upon wrist Flex, Extend, Adduct and Abduct are the flexor/extensor **carpi** radialis/ulnaris.

The new EMG sensors we have chosen are unnecessarily large (excessive amounts of adhesive) and could be engineered at a fraction of their size with no loss of functionality. For our purposes, this limited the amount of space left to allocate to FSR sensors. We therefore chose to put the reference electrode on the other hand. In an ideal situation where the sensors are smaller, it should be possible to have the reference electrode placed close to the ulna as in our pilot study, without overlapping with any of the other sensors.

The force sensors we have used are much smaller than the EMG sensors. This allows us to easily place them on the sides, close to the tendons of the muscles which control thumb movement and hand abduction/adduction. We were also able to fit two more FSRs in between the two EMG sensors on the anterior and posterior of the wrist (Figure 4.1. These latter two are conveniently placed onto the tendons of interest for finger gestures: tendons of the flexor and extensor digitorum muscles. The sensors are spaced around an elastic band which is worn around the wrist. Our design supports adjustment of the sensors' placement on the band, to account for differently sized wrists. The elastic is required so that there is slight pressure exerted onto every force sensor, so that when the shape of the wrist changes, there is a change in pressure. In the absence of such an external force, the sensor would simply move with the surface of the skin, and there would be minimal change in signal. The force sensors are connected to a small circuit that consists of a digital multiplexer and a voltage divider which work in tandem. A resistor of $180k\Omega$ gave us values within a suitable range. The output signal from the circuit is then connected to the Arduino's analogue pin.

The approximate positions of all the sensors are shown in Figure 4.1 and Figure 4.7. In total there are eight electrodes worn on the wrist, and there are also reference electrodes for each pair of bipolar electrodes attached to the other arm. Our aim is to be able to detect finger and wrist gestures with as few electrodes as possible while still maintaining a reasonable accuracy, as increasing the number of electrodes makes a device more impractical due to the size, power, computational load and cost.

The diagram in Figure 4.8 shows how each of the hardware components are connected. In the figure, only one of four sets of EMG components are shown.

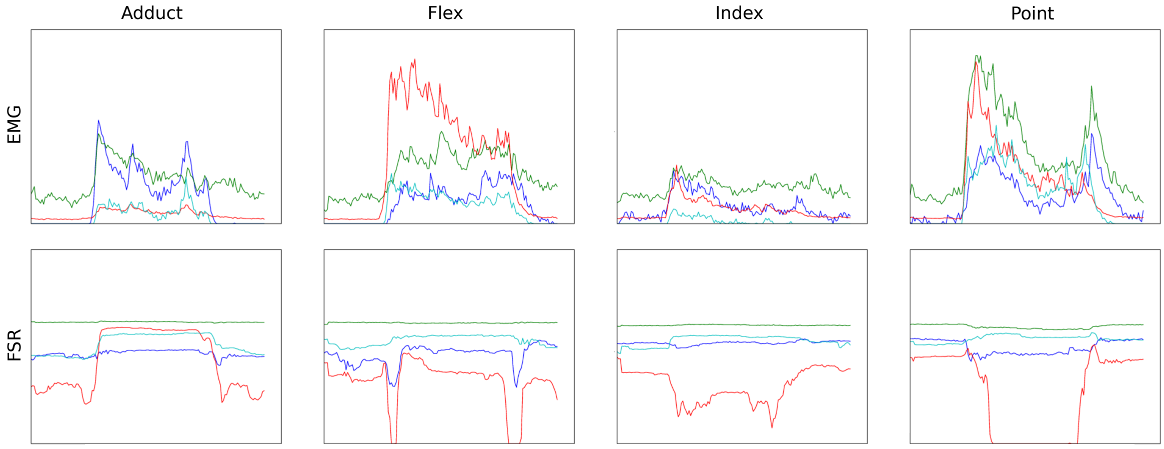

Figure 4.9: Graphs of electromyographic and pressure signals for select gestures. The horizontal axes show time, and the vertical axes show amplitude.

4.4.2 Software

The data collection process that we use remains similar to that of the pilot study. Instead of receiving 2 data channels from the Arduino, we now have 4 FSR and 4 EMG channels. Features for EMG could theoretically be extracted from both the frequency and time domain. However, due to the limitations of our hardware, we could not collect sensor data at a high enough frequency for there to be any significant information in the frequency domain. Useful frequency information would have to sample data at a rate several orders of magnitude higher than our current rate of 60Hz [1, 98]. We extracted the following features for each 1D time signal (2.4s):

- Root mean square (RMS) - This feature is correlated with the signal energy, and thus muscle activity in the case of EMG. This has been found to be a good feature for machine learning with EMG data [7]. Significant changes in signal energy also occur in the FSR data, as seen in the sample data shown in Figure 4.9. The muscles in the forearm have different sizes, and the larger they are, the higher the amplitude of the received EMG signals. Large muscles, such as the flexor carpi ulnaris, can influence several sensors due to their size. These properties facilitate the process of classification, as the muscles which could have produced the electric potential difference can be inferred.

- Standard deviation (SD) - Also correlated with the muscle activity, although this feature is invariant to amplitude offsets.

- Peak amplitude - Measures the maximum value of the sensor data. This is chosen to take into account the shape of the signal, as two signals with the same RMS could look entirely different.

We tested the accuracy of our classifier by using a leave-one-out 10-fold cross validation on our data sets. For each iteration, we use a stratified shuffle split on the training portion, and then used a grid search algorithm for selecting the hyper parameters, C and γ. Once suitable parameters were found, we then trained the SVM and tested the classification on the test fold. Using this method, we remove any bias from the parameter selection that would have otherwise occurred if the parameters were instead chosen using a grid search on the entire data set.

The time taken to read the data file, compute the features and classify the instance using the SVM, is less than 5ms on a desktop class Core i7 Intel processor , across all gestures and sensors. This is 3 times less than the sampling rate, which suggests the feasibility of real time classification using our system entirely possible, as further optimisations would reduce the time even more.

Our software suite also included an application that would show the participant a sequence of videos of someone performing certain gestures that they would then mimic. While the participant is performing the gesture, a second Python script would record 2.4 seconds of sensor data that was sent to the computer's serial port from the Arduino. It is then stored as comma separated value files for later analysis.

4.4.3 Participants

A total of 12 participants (different from the first study) took part in our experiment. There were 3 female, and 9 male participants. The average age across all participants is 33 years, with a standard deviation of 9.1 years. The circumferences of their wrists averaged 16cm, with a standard deviation of 1.3cm.

4.4.4 Procedure

An experiment began with the participant being asked to wear the device around their left wrist. We first adhered the electrodes for the electromyography onto the participants arm. The band with the force sensors was then placed around the wrist, making sure that each force sensor had good contact with the skin, with slight pressure on each of them to ensure the output signal was in a suitable range.

We then instructed them to mimic the gestures that were shown in a video clip on a screen in front of them. They were told to keep the timing of their hand movements in sync with the one in the video clip, as closely as possible. Each video clip is 4 seconds, and the gesture in every clip starts and ends at the same point in time. This is to ensure that the gestures are performed consistently throughout. Each gesture video clip is shown 6 consecutive times, the first clip is to let the participant acknowledge that a new gesture has started, and so that they can practice it once. The data is not recorded during this period, only for the 5 subsequent gestures. The participants perform 15 different hand gestures in this manner, these gestures are all those shown in Figure 4.3. This process is repeated once more, to give a total of 10 data points per

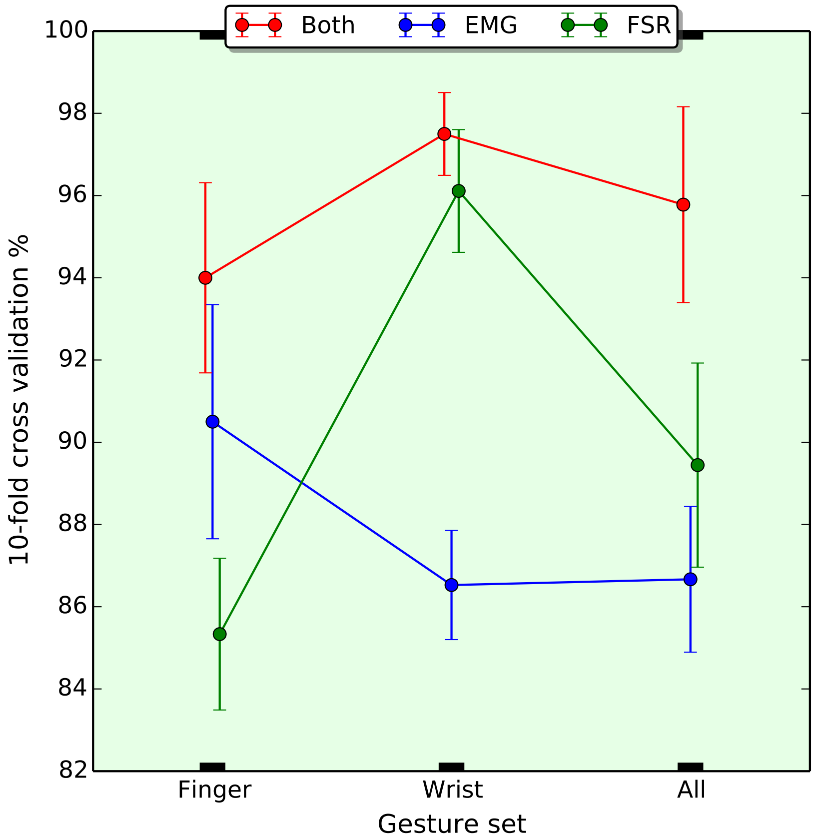

Figure 4.10: Cross validation accuracies for each test condition, with standard error bars.

gesture. For each participant's data set, we tested 10-fold cross validation accuracies using a separate SVM to train each individual's gesture data.

In our study, we used three sensor conditions (EMG, FSR, Both) and three gesture sets (Fingers only, Wrist only, All), where the 'All' category combines the first two gesture classes with the 'Other' gestures shown in Figure 4.3. Since we vary both of these variables simultaneously, data is collected across a total of 9 different experimental conditions.

To re-iterate our hypothesis: we expected that incorporating pressure data increases the classification rate of gestures compared with using EMG alone on the wrist.

4.4.5 Results

Figure 4.10 shows the mean results for each experimental condition. For the full set of 15 gestures, average 10-fold cross validation classification rate for both EMG and FSR data across all participants is 95.8%. This gives us a classification rate for our overall device with respect to our gesture set. We also compared the Sensor and Gesture conditions to identify main effect and to identify any interaction between them.

A Shapiro-Wilk test showed that the classification sample was normally distributed for EMG ($W=0.944912$, $p < 0.05$) and FSR ($W=0.973691$, $p <> 0.05$). A two-way repeated measures ANOVA indicated that the main effect of Gesture was significant, $F(2,10) = 5.67$, $p < 0.001$. Post hoc analysis with Bonferroni correction accounting for multiple comparisons showed that Wrist Gestures were classified significantly better than All Gestures ($p < 0.05$), but not significantly different to Finger Gestures (although approaching significance, $p=0.06$). The overall difference between Finger Gestures and All Gestures was also not significant.

The main effect of Sensor was also significant, $F(2,10) = 17.70$, $p < 0.001$. Post hoc analysis, again with Bonferroni correction, indicated that Both sensors were significantly better than either EMG ($p < 0.001$) or FSR ($p < 0.001$), but that these did not significantly differ from one another.

The interaction effect between Gesture and Sensor was also significant $F(4,8) = 11.69$, $p < 0.001$. The distributions of Gesture and Sensor classification rates suggest that this effect was caused by EMG data correctly classifying more Finger Gestures (90.5%) than Wrist Gestures (85.3%), while FSR data correctly classified more Wrist Gestures (96.1%) than Finger Gestures (86.5%). The EMG and FSR techniques in combination showed significant complementarity when classifying All Gestures, each technique alone providing fewer correct classifications (EMG 86.6%, FSR 89.4%) while together they classified All Gestures correctly to an accuracy of 95.8%.

The confusion matrix in Figure 4.11 shows the performance of the classifier for the case with Both sensor types and All gestures. There is no obvious misclassification for any particular gesture, although the main confusion appears in the finger gestures. The ring gesture has the highest number of false positives, wrongly classifying the middle, pinky, and palm gestures. Similarly, the SVM classified several index and pinky gestures as middle finger gestures, and fist as point gestures. This error could be attributed to a few possibilities. The most likely of them however, is that there is an insufficient number of sensors for the flexor/extensor muscles that control the fingers. The Finger gestures being the most incorrectly classified gestures can also be confirmed by the fact that the average cross validation result is worse than that of Wrist gestures, and also All gestures. The thumb gesture appears to have fewer false positives and false negatives, and this is likely due to do the fact that the thumb gesture uses muscles which are different from those the other finger gestures use.

We found the standard deviation to be much larger for Finger gestures (3.72%) than Wrist gestures (1.67%) and All gestures (1.98%), when all sensors are used. When the EMG sensors are only used, the deviation is 6.22% for the Finger gesture set. This is a much larger variance than Wrist gesture classification using only FSR (2.05%). One possible reason for this could be the variance in hand dexterity of participants, due to the somewhat difficult control of individual finger flexing; the flexor digitorum superficialis is principally responsible for this.

In theory, larger wrist sizes should increase the recognition rates of the classifier, since the muscles are more interspersed, and the larger muscle mass should increase the difference in elec-

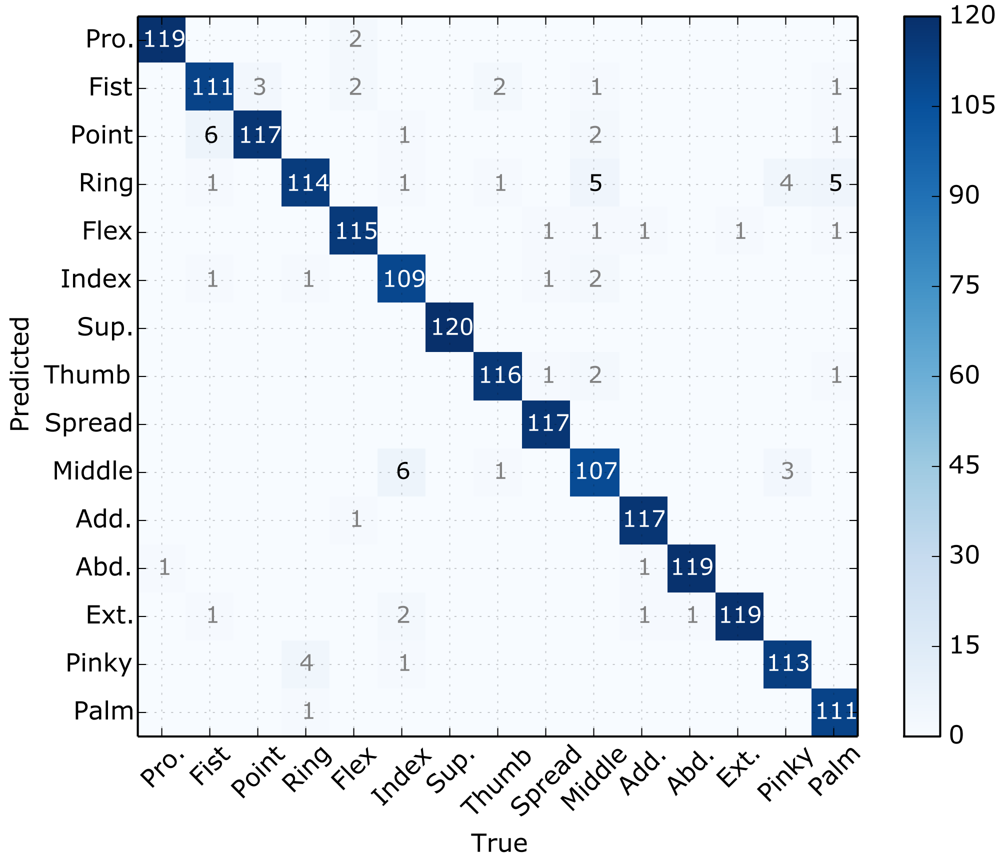

Figure 4.11: Confusion matrix across all participants, using both sensors and all gestures.

trical potential. Though our data does not suggest such a correlation between the circumference of the subjects wrists and the classification accuracies, our sample has too limited variance to determine this.

4.5 Discussion

Our principal result is the accurate classification, just below 96%, demonstrated by the combined EMG and FSR data across our gesture set. This indicates that the EMPress technique is viable for hand gesture recognition, and that this approach is significantly better than either sensor type on its own.

It is not possible to directly compare EMPress to other published gesture sensing methods, because gesture sets, number and type of sensors used, and anatomical position vary widely across the literature. Our overall classification rate for Finger gestures is 94.0%, while the rate for Wrist gestures is 97.5%. Confusion matrices show that the weakness in our system is generally due to the mis-classifications of adjacent fingers within the Finger gesture set. We

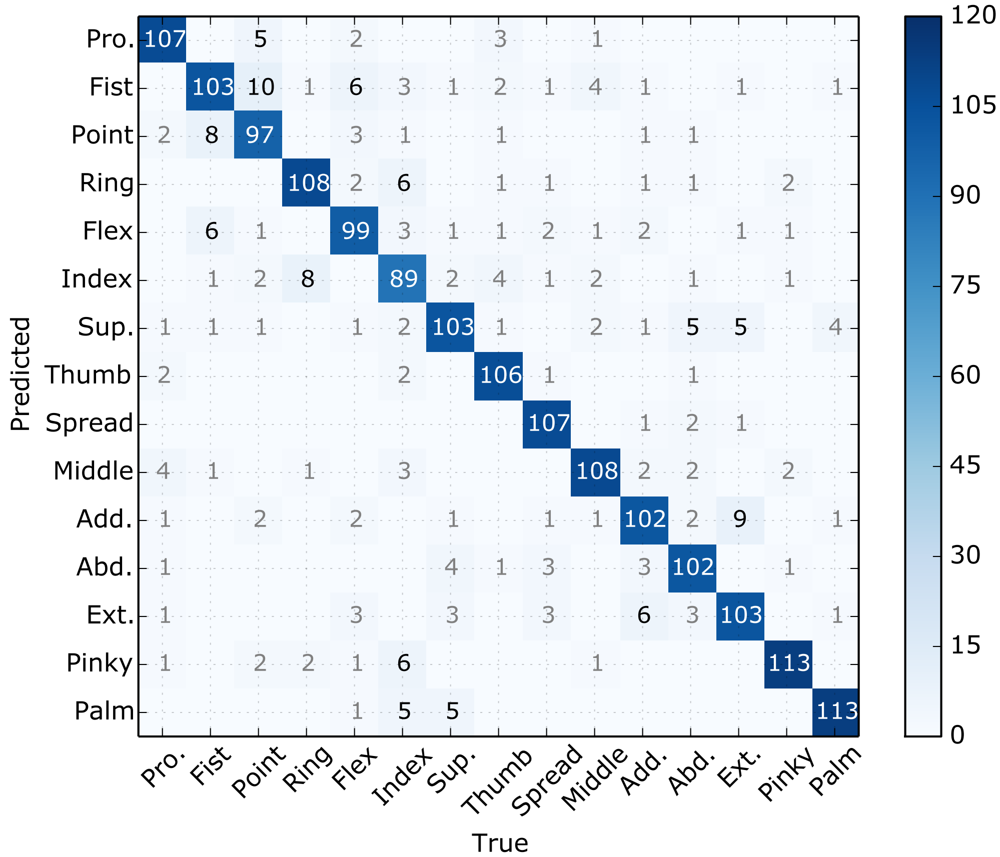

Figure 4.12: Confusion matrix across all participants, using only EMG sensors and all gestures.

suspect that this is partly due to the low electric potential generated by the few muscle cells that are present in the wrist. Sensing EMG spikes predictably at the wrist against low signal-noise ratios is the most challenging aspect of the EMPress technique. Nonetheless, our data shows that even with these challenges that EMG outperforms FSR when classifying Finger gestures, and its inclusion therefore remains an important component of our technique. Uniquely, we have also demonstrated that this high level of classification accuracy is possible without resorting to distributed arrays of EMG sensors across the forearm, instead localising the sensors to a wrist-mounted device. Following [7], even if constrained to "spare" real estate on the existing wrist strap, we expect that further EMG sensors should increase the classification rate.

While we expected pressure sensing to significantly supplement EMG gesture classification accuracy, our study demonstrates that pressure not only contributes to EMG classification of gestures, but can actually parallel or even beat EMG performance in complementary ways. The study data reveals that the overall FSR-only classification success rate is 89.4%, showing a strong predictability of wrist pressure over a wide range of gestures. While we expected pressure sensing to perform well under conditions of wrist movement (96.1%), perhaps most surprising is that

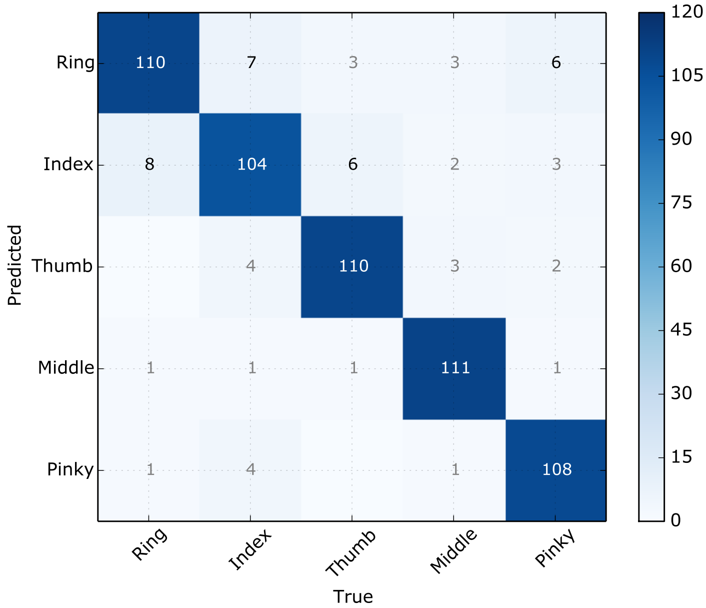

Figure 4.13: Confusion matrix across all participants, using purely EMG sensors for classifying finger gestures.

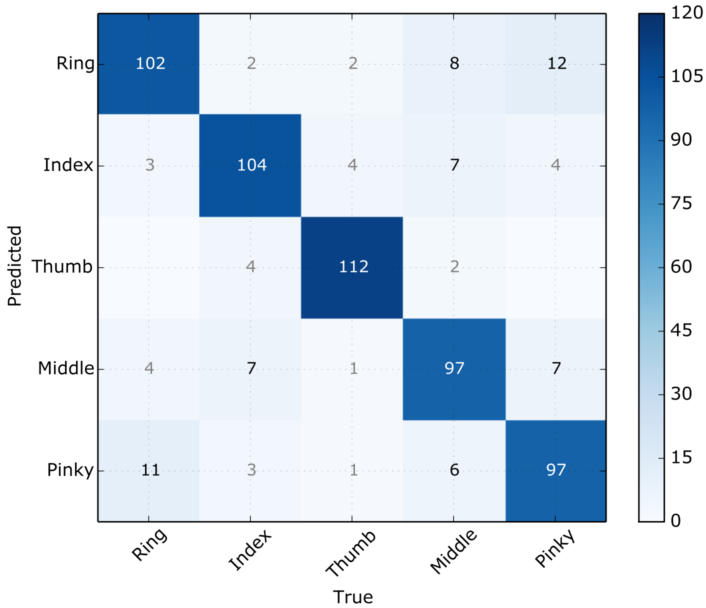

Figure 4.14: Confusion matrix across all participants, using purely FSR sensors for classifying finger gestures.

pressure sensing alone was able to detect finger gestures to a classification accuracy of 85.3%. While WristFlex [29] already demonstrated a high predictability for a particular anatomically targeted pinch gesture using FSR sensing at the wrist, we believe the EMPress technique is the first time such an approach has been shown to be potentially strong for a varied range of gestures without special anatomical targeting of the sensors, and particularly for combined wrist and finger movement classification.

Finally, our study shows that the EMG and FSR techniques are strongly complementary, with EMG significantly better for detecting finger gestures and FSR significantly better for Wrist gestures. We therefore propose that, for wrist-based wearables, the techniques are used in conjunction so that a high recognition rate can be achieved while locating sensors only at the wrist. Note that our device consists of wet electrodes with cross-arm references, this suggests that our findings apply under optimal EMG conditions. It is not necessarily the case that this complementarity will yield true for a system with dry electrodes with a same-arm reference electrode placed over the ulna, and further work will be required to verify this level of performance against dry and locally grounded EMG such as our pilot study used.

4.6 Future Work

We anticipate exciting future work in the design of the electrodes deployed in EMPress-enabled wearables.

Firstly, it may be possible to integrate the different sensing types. Having separated EMG and pressure sensing in order to separate their contribution to the overall effect, it is now possible to consider engineering a design which re-integrates EMG and pressure sensing. Designs would be possible which directly exploit the pressure-modulated EMG signal we observed, alongside a comprehensive exploration of optimal feature sets of the signal for classification. It would also be possible to retain the software design for our second study but integrate hardware within the EMG electrode itself, supporting pressure sensing while greatly increasing the space available on the wearable strap for a larger array of sensors around the wrist.

Moreover, with this approach the FSR component could then directly measure the pressure exerted onto the EMG sensor. In our final prototype, even with foam underneath the electrode, there is still a slight change in electrical contact as the foam is compressed. Dry electrode sensors such as those used in the Myo will also suffer from pressure changes due to the movement of the wrist strap. The use of conductive foam [9] may mitigate but not remove the effect of pressure. However, an integrated EMPress sensor would be able to measure the pressure exerted onto the EMG contact from the band. This would allow a predictive model of the modulating effect of pressure on the EMG signal to underpin an estimate of the 'true' EMG signal.

Our gesture set allows us to differentiate between Finger gestures and Wrist gestures. However, we have not tried to classify nuanced combinations of both Finger and Wrist gestures, nor have we attempted to combine multiple Finger gestures simultaneously. Detecting multiple Finger gestures with a reasonable accuracy will probably demand higher density of EMG sensors placed precisely above muscles of interest as described above. Since the sensors seem to be suited to recognising separate components of gestures, calibration algorithms to map sensor-anatomy offsets [7] will be important. There is also the potential to biologically tailor the device so that additional pressure sensing can be targeted towards regions where there are limited surface EMG signals owing to wrist anatomy.

Finally, given existing wearables typically already include an IMU, our technique could potentially draw on acceleration data to detect many more gestures and/or with further increases in accuracy. In particular, gestures which require movements that stem from the elbow, arm, or whole body would be well-supported by such data, and we may observe some further improvement to Wrist gesture classification accuracy as well.

4.7 Conclusion

Our initial study planned to compare the effect of EMG sensor placement on the forearm. We found that placement of EMG around the wrist has comparable accuracy to when the sensors are located on the upper forearm. This unexpected outcome led us to believe that pressure exerted from the wrist-band which held the EMG sensors to the skin, modulated the signal in a semi-predictable manner.

Our main study confirmed our hypothesis that including the pressure around the wrist does indeed increase the classification rate for different kinds of gestures involving both finger and wrist movements. Furthermore, we found significant complementarity between the two types of sensors: the pressure sensing surpasses EMG for classifying Wrist gestures, and the reverse is true for Finger gestures.

We believe this technique is sufficiently accurate and ergonomically practical to have significant potential for underpinning a new generation of wearable gesture sensing technology.

4.8 Reflection and adjustments to the requirements

Although we have demonstrated that EMG works at the wrist and found improvements for this, there were problems with the technique that became clearer after experimenting with it. These issues are seldom described in EMG literature, and therefore can be misleading or at the very least uninformative of the types of interactions that EMG can currently enable.

The first problem is the necessity to perform gestures forcefully. This is required since the amplitude of the signals generated by the muscles are very weak. This problem is exacerbated by constraining the placement of electrodes to the wrist, where the muscle mass is much smaller and therefore the signals diminish further, leading to poorer signal to noise ratio. In fact, during the study, the participants were asked to produce gestures very clearly, with intent. Performing gestures in this way is not ideal as it may lead to muscle fatigue after prolonged use, which can be expected in some usages. Gestures performed weakly is also likely to yield greater classification errors. This problem makes the technology unsuitable for subtle gestures that are typically detected for the application of contextual awareness and other implicitly detected activities.

The second problem of EMG is also one that stems from the nature of the electromyographic signals. The signals are generated by the muscles and there are features of the signals that vary depending on the length of the contraction (in time). Faster contractions produce a higher amplitude signal that makes it easier to measure and identify. However, slower, steadier contractions are more difficult to pick up due to the power of the signal being spread out over time. This is the type of contraction that needs to be detected for control of continuous variables using gestures.

Therefore, EMG is a technological solution to a subset of application usages. It does however, lack the precision and fidelity needed for subtly or implicitly performed gestures and gestures used for controlling continuous variables. The requirements for a gesture recognition device can thus vary depending on the intended application. We've used 'fidelity' as a term that encompasses accuracy and range of gestures in the table of analysis (table 3.1 in section 3.12). We will now extend our definition of fidelity to include additional precision for continuous or weakly performed gestures.

Since the requirements of an application may or may not require such a level of fidelity, we include update our baseline requirements to include a situational requirement. Updating the initial baseline requirements to include (from section 1.4.5), changes marked in bold:

- Mobility → Wearable.
- Placement & Size → Ergonomics, practicality and social acceptability.
- Accuracy & Range of Gestures & Continuous and Weak → Usefulness.
- **Continuous control or subtly performed gestures.**
- Complexity → Immediate feasibility.

It is assumed here that a technology that is able to detect the continuous states of hand poses (as opposed to discrete gestures), is also able to detect gestures performed weakly. This assumption is made because changes in continuous states of the hand pose is essentially a a weakly performed gesture. For this reason, continuous control and subtly performed gestures are put into the same requirement. In a way, this also resembles a level of precision, which is why we still consider it a level of fidelity.

To sum up the the chapter, we will also update the table of analysis with our new findings:

The analysis of the EMG technique has been revised in the above table, with experiments to confirm it's functionality remains with good fidelity at the wrist location. It is also fairly effective for a wide range of gestures with fairly simple, low-cost equipment. The size of the equipment is

Technique	Mobility	Placement & Size	Fidelity	Complexity
EMG	Battery concerns	**Wrist**	Good*	**Simple**
Ultrasound	Improving but still poor	Variable	Very good	Very complex
Infrared	Highly mobile	Thin & wrist	Poor	Very simple
Pressure	Highly mobile	Thin & wrist	Poor	Very simple
EIT	Highly mobile	Thin & wrist	Average	Fairly complex
Optical	Can be mobile	Wrist, but obtrusive	Limited	Fairly complex
Glove	Highly mobile	Obtrusive	Excellent	Simple

Table 4.3: An updated table of analysis for gesture recognition techniques, with respect to the requirements.

still somewhat of an issue, with electrode-spacing a variable to investigate in the future (this is why it is not written 'Thin and wrist'). The asterisk on the fidelity (*) is there as a reminder of the fact that EMG has difficulties with detecting continuous or subtly performed gestures. Additionally, the fidelity also scales with the number of electrodes. Reading through literature not in EMG, but in ultrasound imaging confirms this phenomenon. An example of work in this area that identifies this problem, work by Zheng. et al on ultrasound imaging for detecting muscle movement [133] uses this limitation of EMG as a motivation to pursue higher tracking fidelity with ultrasound imaging. However, as analysed in the related work, ultrasound has serious practicality issues that need to be addressed. Thus, in order to search for a technology that can satisfy this separate requirement of continuous control / subtle gestures, we now investigate ultrasound imaging as a potential technology for integration with wearables.

Chapter 5

Ultrasonography

5.1 Introduction

Ultrasound imaging, or ultrasonography (US is also a typical abbreviation of ultrasound, which in turn can be a name given to the ultrasound imaging technique), has long been available as a medical technique for safe anatomical inspection, for example to see developing fetuses or heart disease. It's uses extend beyond clinical uses, with previous research utilising ultrasonography as a control mechanism for devices. This has been motivated by uses in: controlling mechanical hands for either prostheses, teleoperation or exo-skeleton control. Despite this work being found outside of the HCI community, there is an obvious connection to it's purpose in controlling computers. Therefore, here we explore the possibilities that ultrasound may enable in future gesture recognition devices, for HCI purposes.

Ultrasound imaging provides a direct visualisation of the muscles in real time, enabling precise estimations of body movement. It is unlikely that this technique provides more accuracy than systems which directly measure moving body parts, such as cameras or data gloves, but US has potential benefits over these existing wearable hand tracking techniques. Directly imaging the muscles is not subject to occlusion, unlike imaging the body parts externally which can be obscured by other body parts or be out of view depending on the mounting of the camera. In contrast to wearing instrumentation on the hand, imaging the muscles leaves the hands free to perform actions unencumbered by devices or sensors. This could be beneficial for sensitive or expert manipulation as in the case of fragile objects, surgical operations, or simply to keep the physical naturalness of a handshake.

This differs from most other techniques such as EMG, or methods that rely on deformation of the skin surface, as these instead indirectly measure signals.

Moreover, the signals from EMG are usually sampled over a window of time during which the gestures are performed. In contrast to this, Ultrasonography is able to give an estimation of the hand pose at any single point in time (relevant works cited in section 3.5.3 "Estimating Hand and Digit Angles Continuously").

In fact, the ability to give an estimation for any single point in time would enable the applications that require higher fidelity (as mentioned at the end of the previous chapter): weakly or subtly performed gestures or continuous control.

However, with this technology being relatively new, there are numerous questions surrounding it's use as a wearable gesture recognition device. To begin with, there is the question of device placement. Previous work using ultrasonography for hand gesture recognition has done so with a variety of device placements and orientations (as listed in the related work and table of analysis). Location plays a fundamental role in ergonomics and performance since the anatomical features differ among positions. We therefore compare the performance of different forearm mounting positions for a wearable ultrasonographic device. We analyse the performance of this in two different ways: discrete gesture recognition and continuous finger flexion. In addition, we also provide a simple but effective method for compensating for cross-session sensor misalignment.

Not only is the technology fairly recent, but also rapidly improving. One of the obvious challenges is integrating this technology into wearable devices, due to it's size. In the latter parts of this chapter, we show that although the technology is gradually decreasing in size, integration of ultrasonography devices into wearable devices is currently not feasible. We go further to explain why exactly this is the case, with a preliminary exploration into alternative device designs to reduce the size and complexity. Finally, we highlight one of the key practicality issues of ultrasonography that is often disregarded or ignored as problem: the use of gel between the device and the skin in order to allow the sound waves to pass through. We try to investigate alternative dry coupling mediums to mitigate this issue.

The content of the former and more prominent part of the chapter is a revised version of work I have published in the Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17): "EchoFlex: Hand Gesture Recognition using Ultrasound Imaging"[81].

Sample data is provided in the supplementary material found in the ACM digital library, under "Source Materials": `https://doi.org/10.1145/3126594.3126604`. The SonographySamples directory contains many video clips of all the gestures in each of the locations, for one participant.

The latter part of this chapter that addresses the gel coupling problem of ultrasonography is mostly taken from work that I have published in the Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces (ISS '17): "Improving the Feasibility of Ultrasonic Hand Tracking Wearables" [79].

5.2 Related Work

In this section we focus on describing previous work on ultrasound to support interactive applications. As with the previous chapter, we will forgo describing other sensing technologies due to them being described in the literature review, in chapter 3. The relevant ultrasound related work is re-iterated here for ease of reading, but a more thorough review can be found in section 3.5, within the literature review.

5.2.1 Ultrasound (without 2D imaging)

Sonomyography is the use of ultrasound imaging to create 1-dimensional scans (A-scans). An early study by Hodges et al. [56] proved that muscle contraction of the tibia, biceps and abdomen can be measured with an ultrasound sensor. Later, Zheng et al. [133] showed that it is also possible to infer the wrist angle with an mean error of 7.2%. They defined Sonomyography as measuring the dimensional change of muscles with ultrasound sensors.

Sonomyography has the advantage of using simple hardware because it only needs one transducer element per A-scan; current ultrasonic probes have more than 128 elements and are becoming the standard for ultrasound imaging. Sonomyography can also be used to continuously detect the opening and closing of the hand with an RMS of 12.8% [22] and recently to detect 6 common gestures with an accuracy of 74% [55]. The latter study used a wearable device and a bracelet with five transducers. Comparisons between EMG and sonomyography were favourable for the latter in detecting continuous wrist flexion [51] and hand grasping force [52].

Sonomyography is an appropriate method to detect a reduced set of gestures, but provides limited information and lacks the possibilities offered by modern phased arrays of ultrasound transducers for more detailed imaging.

Mujibiya et al. presented a device that is capable of detecting arm grasps and other on-body touch interactions, using a band of ultrasonic transducers around the forearm and fingers [87]. This device has the advantage of working at much lower frequencies and with fewer transducers than imaging devices, but the drawback is that it can only detect interactions when one hand is touching the arm of the other. Detecting hand movements involving only one hand is a requirement for certain situations [64].

5.2.2 Ultrasound Imaging

Ultrasound imaging has been used previously to infer the position of the foot by tracking the muscle displacement [77]. An image of the muscles and tendons was taken and the insertions of the tendons into the muscle were used as markers to track. A comparison between the estimated position and the ground truth (metallic plate inserted in the tendon) showed errors of less than 10 micrometers or the equivalent of 0.7 degrees on the ankle rotation. Markers were tracked using cross-correlation. Later, this technique was used to track the tendon in the wrist with an error of 80 micrometers [70].

Castellini et al. estimated continuously changing finger angles by analysing ultrasonic images of the forearm [18, 19], obtaining an NRMSE of approximately 2%. This technique was also extended to determine the force that the fingers were exerting [47]. They captured ultrasound images at the wrist and divided them into regions. Then, the features from each region were used in a linear regressor to establish a relationship between the features and the angle or force of each finger. In more recent work, it was shown that it is possible to recognise 4 different grasps with 80% accuracy, and also 3 levels of strength with 60% accuracy [94].

Sikdar et al. focused their work on recognising discrete hand gestures by imaging the muscles mid-forearm. They divided the image and calculated the average brightness change per region to create different activity patterns for each gesture. A Nearest Neighbour search was able to classify input gestures correctly with 98% accuracy for the flexion of 4 fingers [116] and 91% for 15 gestures [3]. Another group of researchers used an optical flow algorithm to determine the movement of the extensor muscles [113], but they were only able to qualitatively detect different finger flexions.

The results from Castellini and Sikdar demonstrate the capabilities of modern ultrasound imaging in gesture detection, but did not explore probe location or apply recent machine learning techniques. Considering that the anatomy can vary a lot between different locations, there is likely to be some variation in performance. Cross-session accuracies have also not yet been studied, though it is vital for practical use. Given the importance of location to interactive wearables, further research is required to know whether probe location significantly effects the recognition accuracy. We shall also examine the effects of cross-session performance for each of these different locations.

5.3 Ultrasound Imaging Principles

Sound is a mechanical wave that travels by sequential compression and expansion (rarefaction) of the medium. The frequency describes the number of times the molecules expand and contract per second whereas the amplitude refers to the amount of compression.

Sound travels at a speed that depends on the temperature, pressure and the medium (e.g. 340m/s in air and 1500m/s in water under normal conditions). The product of the density of the medium and the speed of sound through it is called the acoustic impedance. When a sound wave passes from one medium to another with a different acoustic impedance, some of the energy is reflected at the boundary. The proportion that gets reflected is proportional to the mismatch in impedance between the two media. Our forearms are made up of many complex tissues with different acoustic impedances. The reflections from these tissues allow us to see the boundaries between them.

The most basic idea of ultrasonic imaging is to emit a short wave (pulse) and reading back the different reflections from the tissues. The pulse will get partially reflected at different depths of the body, the delay in arrival will be proportional to the distance whereas the intensity will indicate the type of tissue since different tissues have characteristic acoustic impedance.

An ultrasonic transducer or probe is a device which houses one or several piezoelectric elements. The elements can transform an electric pulse into a mechanical pulse to generate a wave. The elements can also transduce the mechanical energy from the reflected pulse into electric signals. These phenomena are referred to as the inverse and forward piezoelectric effect and support pulsing and reading with the same probe by quickly switching the electronic components. Modern ultrasound imaging devices use algorithms to form images, applying beamforming and harmonic imaging to increase resolution.

Ultrasonography typically uses sound waves with frequencies in the order of MHz. The higher the frequency, the better the resolution that can be attained, however the attenuation is higher and therefore imaging depth is reduced. In this paper we image musculoskeletal tissue using transducers with a central frequency ranging from 8 to 12MHz for good resolution and enough image depth.

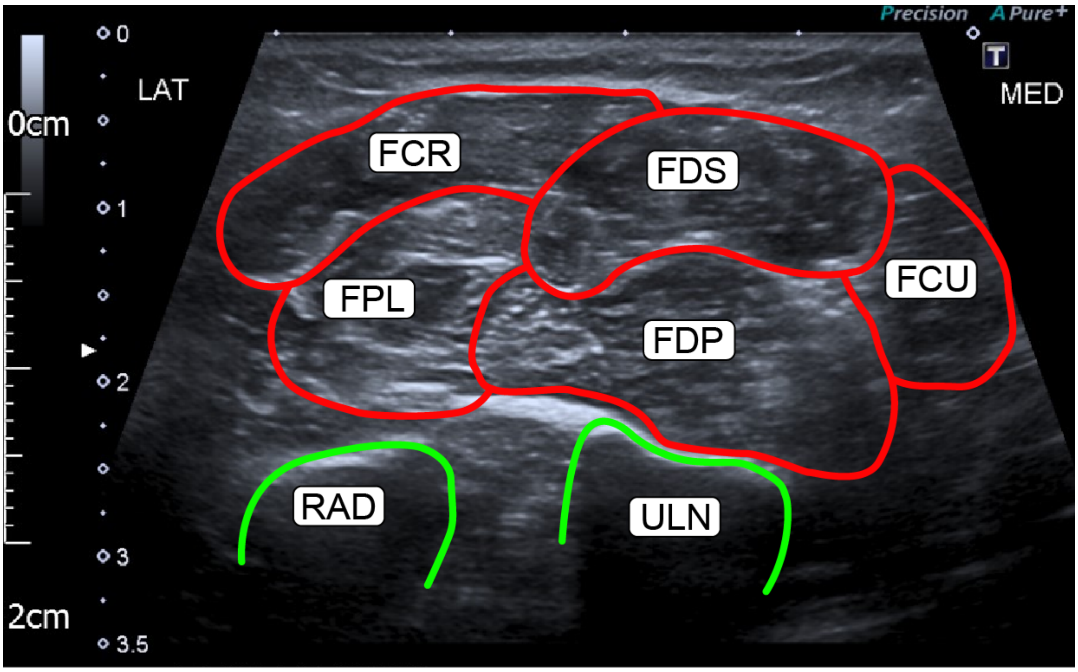

Figure 5.1: A transversal scan of the mid-forearm, anterior aspect. Labelled in red are the various muscles, and the bones in green.

5.4 Hand anatomy

The muscles which control the hand can be categorised into extrinsic muscles which originate in the upper/mid forearm, and intrinsic muscles which originate within the hand itself. The extrinsic muscles control wrist motions and some digit movements. The intrinsic muscles are used for finer motor control of the hand, for example pinch gestures and finger adduction/abduction. The extrinsic muscles can be further categorised into flexor and extensor muscles which bend or straighten the digits respectively and, in combination, move the wrist from side to side.

In this paper we exclusively image the extrinsic muscles at positions from the wrist to the forearm. These locations provide a position for the probe that will not interfere with hand movements. However, it will not be possible to capture hand movements that use intrinsic muscles, and this is a tracking limitation for all systems that measure muscle movement at the forearm level.

The muscles that are principally responsible for the flexion and extension of the fingers are the digitorum muscles (Fig. 5.1, FDS, FDP). These also assist with flexing and extending at the wrist, however the main muscles for the wrist are the carpi radialis/ulnaris (Fig. 5.1, FCR, FCU). In addition, the carpi radialis and ulnaris assist with adduction and abduction of the wrist. The thumb is controlled by the pollicis muscles (Fig. 5.1, FPL).

Hand movements change the musculature of the forearm due to contractions of the muscles. These contractions expand the size of the muscles and pull the tendons, and these changes are reflected in the US image. Furthermore, a particular hand movement changes the image at the specific areas where the involved tendons and muscles are located. Depending on the position of the probe (e.g. wrist or forearm) the observed muscles and tendons will be quite different.

5.5 Study Design and Pilot

The objective of our study is to analyse the best mounting locations of a wearable device for discrete gesture recognition, cross-session accuracy and detection of continuous flexion angle. The wearable uses ultrasound imaging to track the muscles and tendons inside the forearm and with that infer hand pose. There were several variables to consider in the study; the main ones are locations and gesture sets. In order to make the study feasible we ran a pilot study to discard the options that were clearly inferior or not viable, which is presented after this section.

5.5.1 Locations

There are three parameters to consider when deciding the placement of the probe: orientations, proximity to the elbow, and anterior/posterior placement. The following explains each of these variables in more detail:

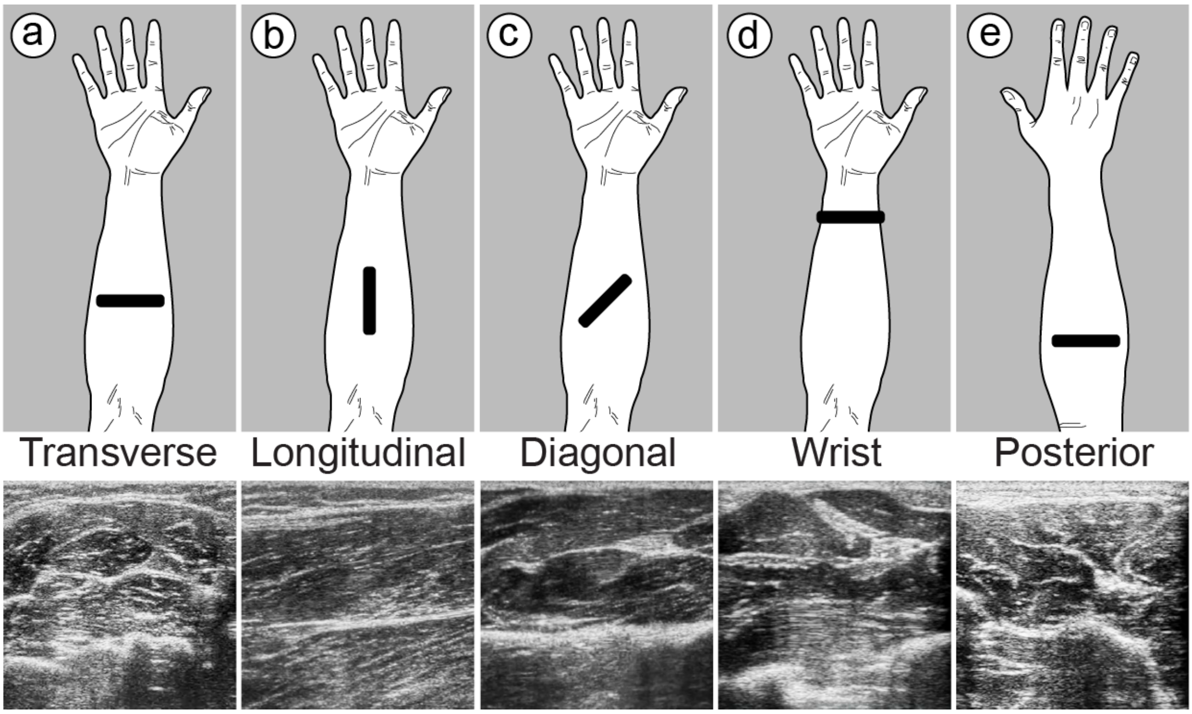

Figure 5.2: The mounting positions of the probe that were compared with their corresponding ultrasound image. a. Transverse, b. Longitudinal, c. Diagonal, d. Wrist, e. Posterior.

Orientation Clinical US scans are usually performed either **Transversally** (Fig. 5.2a) or **Longitudinally** (Fig. 5.2b) along the length of the forearm. In transverse mode, a larger range of muscles can be seen, however the muscle fibre displacement cannot be observed as well as in the longitudinal scan. It is medically uncommon to use the probe **Diagonally** (Fig. 5.2c) but we wanted to test if this position could image a large range of muscles and still observe the displacement of muscle fibres.

Proximal/Distal The muscles are more predominant at the **Proximal** mid-forearm (Fig. 5.2a), whereas the tendons are more visible at the **Distal** wrist location (Fig. 5.2d).

Anterior/Posterior The flexor muscles/tendons are located on the **Anterior** (inside) of the forearm and extensors on the **Posterior** (outside) of the forearm (Fig. 5.2e).

We chose 5 locations to initially test: Transverse, Longitudinal, Diagonal, Wrist and Posterior. These locations are illustrated in Fig. 5.2, along with their corresponding ultrasound images. The figure shows how different the US images look between positions. Although it cannot be seen from static images, the muscle fibre movement differs greatly with the orientation of the probe. Fig. 5.5.3.2a shows the fibre movement when the probe is oriented perpendicular to the flow of the muscle movement (Transverse), whereas Fig. 5.5.3.2b shows that the general movement is along the axis of the forearm in a Longitudinal scan.

Transverse mounting was used by Sikdar et al. [3, 116], Wrist was the location employed by Castellini et al. [18, 19] and a preliminary study tested Posterior [113]. Here, we compare those locations that were used in previous research and we introduce Longitudinal and Diagonal. Longitudinal is the only image that is completely parallel to the muscle and tendon movement. In a preliminary study for continuously detecting ankle angles it showed good accuracy [77] but we expect it to show poor performance for classifying a variety of gestures since there is less coverage of the muscles. We expected Diagonal to be a good compromise between continuous and discrete tracking.

For the purposes of dynamic assessment, clinical radiologists rarely look at images in the axial plane of the wrist, as it is difficult to identify individual tendons and the movement is not as easily appreciated as in the muscle. This was especially true of the the extensor tendons in the posterior compartment of the wrist, as the tendons here are too compact, and very little motion can be observed when compared to the anterior side. Since the wrist is normally considered to be the most ergonomic location to place a wearable device, we included the transverse distal anterior location to have at least one wrist location. If the tendons provide sufficient information for effective gesture recognition, this technique can be integrated with wrist-worn wearables.

In contrast to EMG, in US imaging we are also able to observe muscle relaxation. This means that it is not crucial to image both the anterior and posterior parts as it is possible to infer motion in both directions by inspecting either the flexor or extensor muscles. Consequently, we did not repeat posterior positions with anterior positions.

5.5.2 Gestures

In ultrasound images it is possible to observe the muscles and tendons which control the fingers, thumb and wrist at all locations, albeit more clearly in some locations than others. Consequently, for our gesture set we mixed a representative collection of single digit flexions, multi-finger flexion, wrist flexion and adduction.

We discarded movements that primarily involve intrinsic hand muscles such as pinch gestures and adduction of the fingers. Perhaps these gestures can be detected in the forearm because they are usually accompanied by other characteristic involuntary movements. However, we decided to remove those gestures from this study and focus on the gestures that involve a unique set of muscle movements that originate in the forearm. Fig. 5.3 shows the selected gesture set. We split each of these gestures into flexion and extension phases producing a total of 20 discrete movements.

There are numerous ways to flex a digit since they have at least two joints. We instructed the users to perform a natural flexion which usually involves all the joints but with more emphasis on the proximal interphalangeal joint. We thought that this would hinder the angle estimation since the finger is bent at several positions, different to previous studies in which the fingers were held with splints to force flexion at only one joint [19]. We wanted to avoid unnatural or

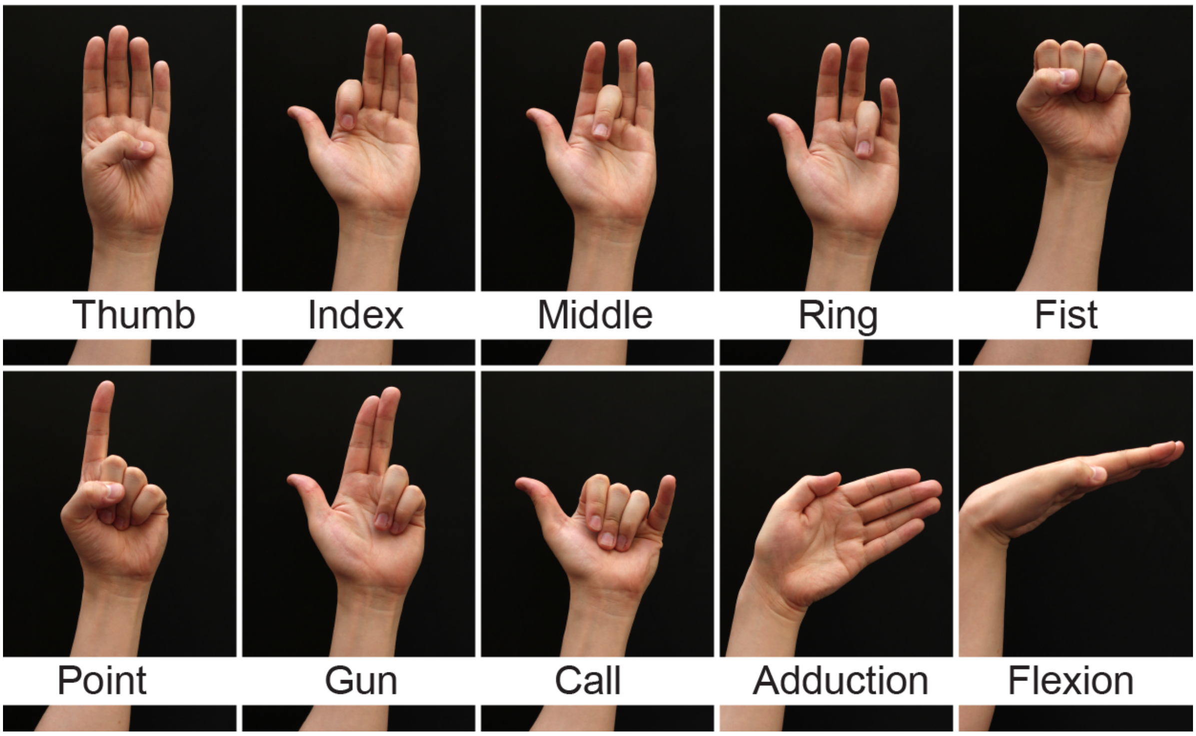

Figure 5.3: The set of gestures from top to bottom and left to right: thumb, index, middle, ring, fist, point, gun, call, wrist adduction and flexion.

uncomfortable gestures as they would not be used in a real situation. For the same reason, we also discarded individual pinky flexion.

With US imaging it is feasible to see the separation between the deep and the superficial muscles. For instance, the boundary between the flexor digitorum superficialis (FDS) and the flexor digitorum profundus (FDP) can be seen as a bright horizontal line in Fig. 5.2b . Consequently, it should be possible to differentiate between flexing of the digit at the metacarpophalangeal and the interphalangeal joints (ie different knuckles).

5.5.3 System

5.5.3.1 Hardware

We used a Toshiba Aplio 80 US imaging machine with a flat, linear probe with an 8MHz central frequency and 12MHz harmonic imaging mode. We used the default musculoskeletal settings with 1 focal point and 4cm of depth imaging for all the positions, except for the Wrist condition where we used 3cm. The machine was an ex-clinical machine but its deficiencies compared to the state of the art reflect the likely limitations of a mobile ultrasound device in resolution and image quality. In order to mount the probe onto the users forearm, we used a 3D printed structure to hold the transducer in place and to strap it around the user's forearm (Fig. 5.4a). We recorded the images from the US machine via a video capture card (Fig. 5.4d). Anagel US gel was applied for coupling between the probe and the user's skin.

In order to measure the ground truth of the hand pose, we created a sensor glove (Fig. 5.4b). We used 5 Spectra Symbol Flex Sensors [41] which were sewn onto each digit of the glove. We used an Arduino to read the resistivity values of each sensor and to transfer the data to the PC. This data was only used for the continuous angle tracking studies. We found that the sensor readings matched the amount of finger flexion linearly.

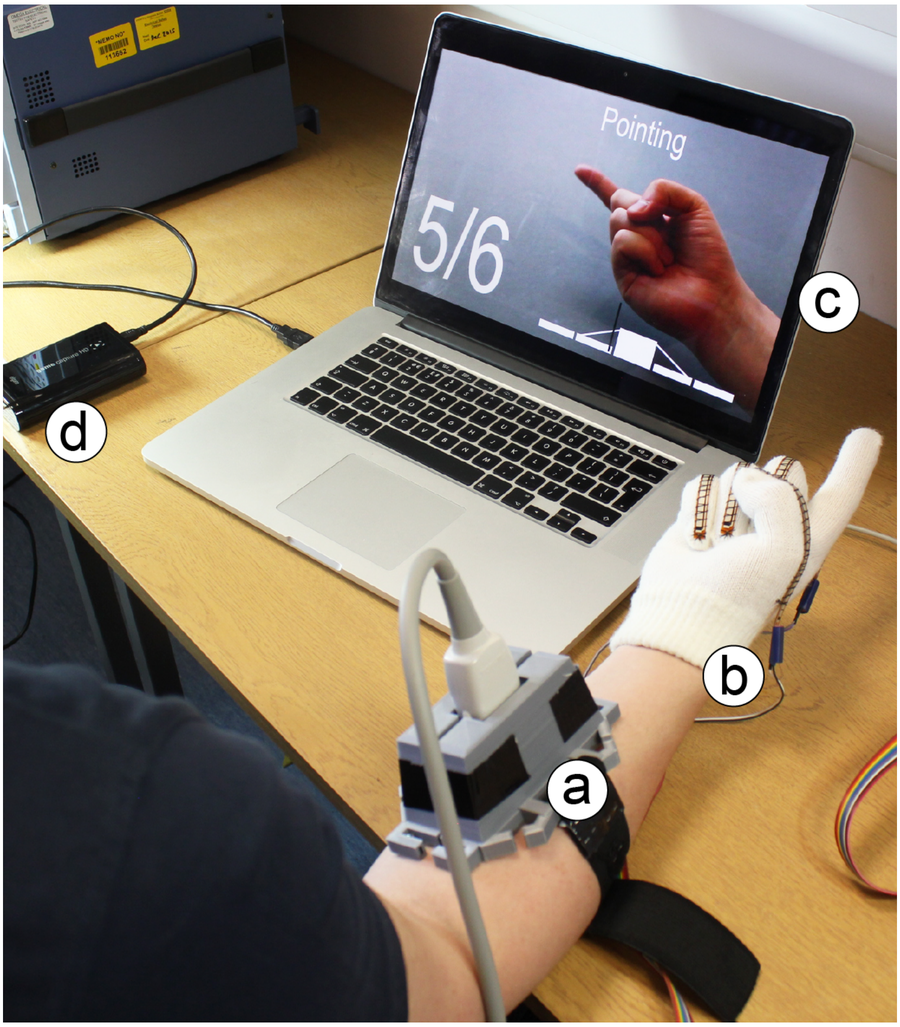

Figure 5.4: The setup of the experiment: a. Transducer, b. Sensor glove, c. Computer d. Video capture device.

5.5.3.2 Software

Since a gesture is actuated by a specific set of muscles, the gestures have unique patterns of activity within the US images. The type of motion depends on the location of the probe, and

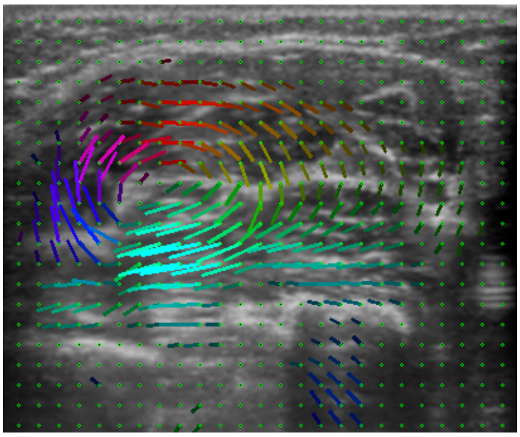

Figure 5.5: Optical flow for the Transverse position during a thumb gesture.

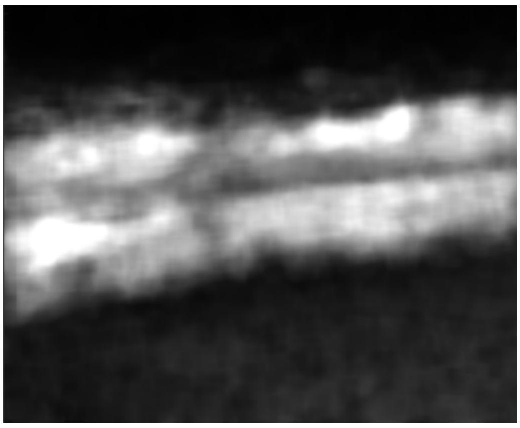

Figure 5.6: Summation of the magnitudes of the flow vectors representing the activity pattern of the index gesture.

the observed motion is caused by reflections/scattering from the muscle fibres. In a longitudinal scan, the motion is primarily along the main axis of the probe, whereas transversal scans have a variety of different motions including rotations or elevations (Fig. 5.5.3.2). Nevertheless, there are patterns in the images wherever the probe is placed. The motion seemed therefore an appropriate feature to classify discrete gestures with.

Several operations must be performed with the ultrasound videos in order to classify discrete gestures or estimate finger angles. A reduced set of features are extracted from a collection of frames and then used to train a machine learning classifier. Later, this trained classifier determines the gesture or angles from a sample set of features derived from unseen images.

For the discrete classification, the first step is to segment the video stream into either flexions or extensions from the neutral position. Then, we extract features for each of the frames and these features are averaged per segment. In the continuous angle inference, the machine is a regressor that is trained with both the features at each frame and the angles from the data glove.

Segmenting the videos for discrete classification was performed by calculating the sum of the magnitudes of the dense optical flow vectors for each pixel of every frame, using Farnebäck's algorithm [40]. A Gaussian blur with a kernel size of 15 pixels was applied beforehand. Fig. 5.5.3.2a shows the sampled points of the optical flow during the index flexing gesture and Fig. 5.7 plots the magnitude of the optical flow against the data glove values. The latter graph shows that the optical flow produces little or no activity during the period of time when the hand is not moving in the flexion hand pose, unlike EMG signals. The videos were split based on a plot such as this, where each gesture can easily be seen with two peaks at the beginning and end. Each gesture in the training set then has a video clip associated to show participants the gesture.

The machine learning algorithms cannot process all the information contained in an image and thus need a reduced feature set. For the discrete gesture classification, Sikdar et al. used the

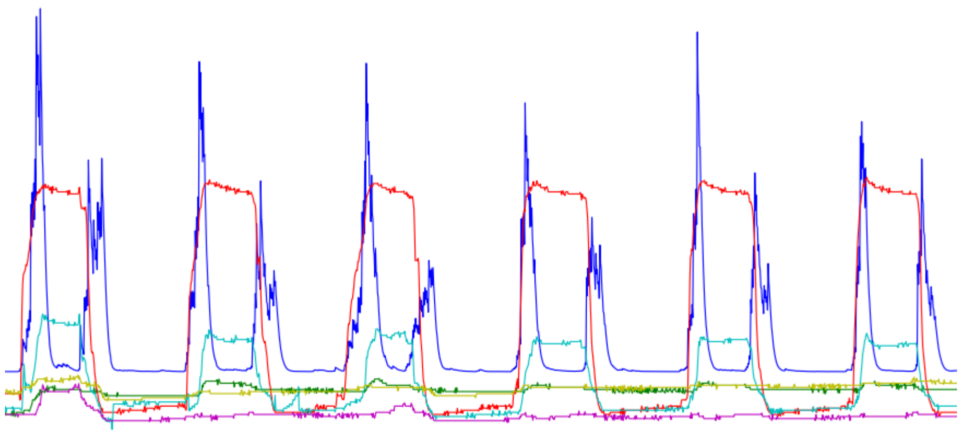

Figure 5.7: The blue line is a plot of the magnitudes of the flow vectors from an ultrasound video of the index finger repeatedly being flexed. The other lines represent the sensor data from the glove. Each channel is a 1D time signal.

accumulated difference in brightness level [3, 116]. In our experiments we used the vertical and horizontal components of the optical flow. For each frame of a video clip, the dense optical flow was calculated, and then the horizontal and vertical optical flow values were averaged within each region based on a grid with 20 pixel spacing. Then the optical flow values were accumulated over every frame in the video clip, giving a final feature vector with a length in the order of hundreds of values. In continuous detection of finger angles, Castellini et al. used the first order surface as features [18], we calculated the moments for each frame (m00, m20 and m21-m12) which are equivalent but more standard in the computer vision community. The moments are computed for each frame, and the regressor trained on moments of each frame with the data glove data.

For machine learning in the discrete gesture recognition task, Sikdar et al. used Nearest Neighbour algorithms [3, 116]. We tested Support Vector Machine's (SVM) since it was shown that they gave better results for discrete gesture discrimination in another study [105]. We also tested a type of Neural Network called Multi-Layer Perceptrons (MLP) since they have often shown very good results and have yet to be applied in this technology. In the continuous angle detection, Castellini et al. used Linear Ridge Regression but suggested to use SVMs as they may yield better results [18]. In addition to testing SVMs for the regression task, we also test a multi-layer perceptron regressor. The accuracies of these algorithms are compared in the pilot study.

A shift correction was used to improve the accuracy for cross-session studies. These errors occur because the wearable can be placed slightly differently to the first time it was worn and thus the anatomical features are not in the same position. To correct this shift, we computed the optical flow between the first frame of a training set and the first frame on the current session.

Classifier	Diag.	Long.	Wrist	Trans.	Post.
MLP	99.875	99.5	99.5	99.25	97.875
SVM	99.625	99.125	99.375	98.75	97.5

Table 5.1: Results for the average discrete gesture classification accuracies for each position, using different classification algorithms.

Classifier	Diag.	Long.	Wrist	Trans.
Participant 1	99.5	97.0	94.0	98.5
Participant 2	98.5	97.0	97.0	98.0
Average	99.0	97.0	95.5	98.25

Table 5.2: 10-fold cross validation classification accuracies for finger flexing at different joints.

The flow was averaged to get a 2D translation and this transformation was applied to the current video to better align the training and sample features.

We used OpenCV [14] for processing the videos and Scikit [97] for the machine learning implementations.

5.5.4 Pilot Study

We eliminated some of the variables from our main experiment through a pilot study, including the least valuable locations, gestures and algorithms. For this pilot, 2 participants performed each of the 10 gestures (with both flexion and extension of each) 20 times at each of the 5 locations. We compared the accuracy of MLP and SVM, with different classifier parameters to find a suitably accurate configuration. We used 10-fold cross validation on the data set in order to have a simple and reliable measure of the accuracy for each case. Table 5.1 is a table of results for the average cross validation accuracy for each location, given a particular classifier, for the discrete gesture recognition case.

Our initial results demonstrated that MLPs had a slight advantage over SVMs in every location, so our main study used MLPs. Our best MLP configuration had 15 neurons in the hidden layer, alpha=0.001 and BFGS for weight optimization. After initial exploration, it was very difficult to find a position for the posterior location that gave good coverage of all the required muscles. This is due to the arrangement of muscles and bones around the posterior side of the arm and the larger surface area, and is likely to be the reason for the Posterior location's weaker performance.

In our regression tests for this same data set, the same outcome was mirrored: Posterior location showed the worst performance, and the MLP neural network regressor surpassed the SVM. For this reason, we eliminated the Posterior position from our main study. The best MLP regressor configuration we used for this had 15 neurons in the hidden layer, alpha=0.0001 and Adam for weight optimization. The Posterior position is also the usual location for screens and interactive elements of wearables (e.g., watch face). Therefore our decision to ignore this location is appropriate for ergonomics as well as efficiency.

We also asked these participants to perform flexions for each of the 5 digits 10 times, flexing at the metacarpo-phalangeal joint and again at the interphalangeal joints. For each of these 10 video clips, they were again split into flexion and extension gestures. Analysis of this data indicated that it was in fact possible to differentiate between flexing at different joints, with an average accuracy of 97.4% across all positions. The different types of flexing exhibit unique areas of muscle activity across the superficial and deep muscles. This shows the potential for US to detect finer differences in hand pose, a feature that is difficult to achieve with EMG since it is difficult to differentiate between signals from the superficial and the deep flexors.

The ability to differentiate between different levels of pressure has been demonstrated in previous studies with ultrasound imaging [47, 94]. We also found discernible differences in the images and could qualitatively infer the amount of pressure with the changes in the ultrasound images.

5.6 User Study

5.6.1 Participants

12 participants aged between 20 and 50 years were recruited to take part in the user study; 8 were male and 4 female.

5.6.2 Task

The participants were asked to sit in front of a computer, wear the sensor glove and the probe mount on the right hand and then to follow the instructions shown on the monitor (Fig. 5.4c). The arm position was as depicted in (Fig. 5.4a). The video showed the user which gesture to perform and an indicator at the bottom gave visual cues for the timing of the gestures.

At each location the participant had to perform a total of 10 gestures, 10 times each. Each 5-second gesture clip prompted the flexion and extension phases, which were then split into those two halves for a total of 10 flexion and extension gestures. Therefore the study involved: 12 participants X 3 locations X 10 gestures X 10 repetitions = 3600 gesture performances. The average study lasted 45 minutes including the time taken to explain the procedure to the participants and to equip the user with the sensors. The data collected from the participants were analysed offline. The data from the pilot study was not used in the main study.

5.6.3 Experimental Design

The conditions for the experiment were the locations of the probe: Diagonal, Longitudinal, Wrist and Transverse. Each of the 12 participants were assigned a pair of conditions (there are 12 ways of picking 2 out of 4 with order) and repeated the last condition. Therefore, each participant performed the gestures on three locations, the second and third locations were the same for testing cross-session performance. At each location, the probe was removed and placed again after a short rest, without any attempt to recalibrate using the US images. In this regard, the short time in between may not seem sufficient for a cross-session study. However, we presume the main issue for cross-session performance is sensor misalignment (as is the case with EMG [7]), as the significant changes in the US images are likely to cause classification problems. A short delay between sessions is enough to simulate the misalignment that would occur with extended delays, and this way we focus solely on the calibration mechanism without interferences from other variables.

5.6.4 Results

Different measures of accuracy were calculated for discrete recognition and continuous angle detection. All experiments are within-user.

Cross validation was performed on the data, using the MLP classifier described in the pilot study to produce classifications. We employed a 10-fold leave-one-out cross-validation strategy, with each fold containing one instance of every gesture, of which there are 20 of. Since we split gestures into flexion and extension, gestures instances of the same type were not temporally adjacent. The average classification percentages for each location are shown in Fig. 5.8. The confusion matrix for the Longitudinal location is shown in Fig. 5.10.

A one-way mixed Analysis of Variance (ANOVA) was conducted to compare the main effect of sensor location on 10-fold MLP classification performance in Diagonal, Longitudinal, Wrist and Traversal conditions. There was a significant effect of location, $F(3,6)=14.44$, $p<0.01$. Post hoc pairwise comparisons used t-tests with Bonferroni corrections to account for multiple comparisons. There was a significant difference between Diagonal ($M=99.78$, $SD=0.09$) and Longitudinal ($M=97.94$, $SD=0.38$), $t=5.67$, $p<0.01$; Diagonal and Transverse ($M=98.94$, $SD=0.19$), $t=4.08$, $p<0.05$; Longitudinal and Wrist ($M=99.78$, $SD=0.12$), $t=-4.30$, $p<0.01$; and Wrist and Transverse, $t=4.08$, $p<0.05$. There was no significant difference between Diagonal and Wrist, or Longitudinal and Transverse. From these results we can establish that for classification accuracy the Diagonal and Wrist conditions were best, followed by Transverse, with the Longitudinal condition last.

The cross-session accuracy is shown in Fig. 5.9 divided by location for both the raw videos and the shift-corrected videos. The cross-session accuracy is obtained by training the classifier with the data from the second session and then classifying the data from the third session which belonged to the same location.

For the continuous detection of finger angles, the main metric was the Normalised Root Mean Square Error (NRMSE) of the predicted angle compared to the real angle. The value was averaged over the five fingers since the errors were similar for all the digits. These values are shown for each location in Fig. 5.11.

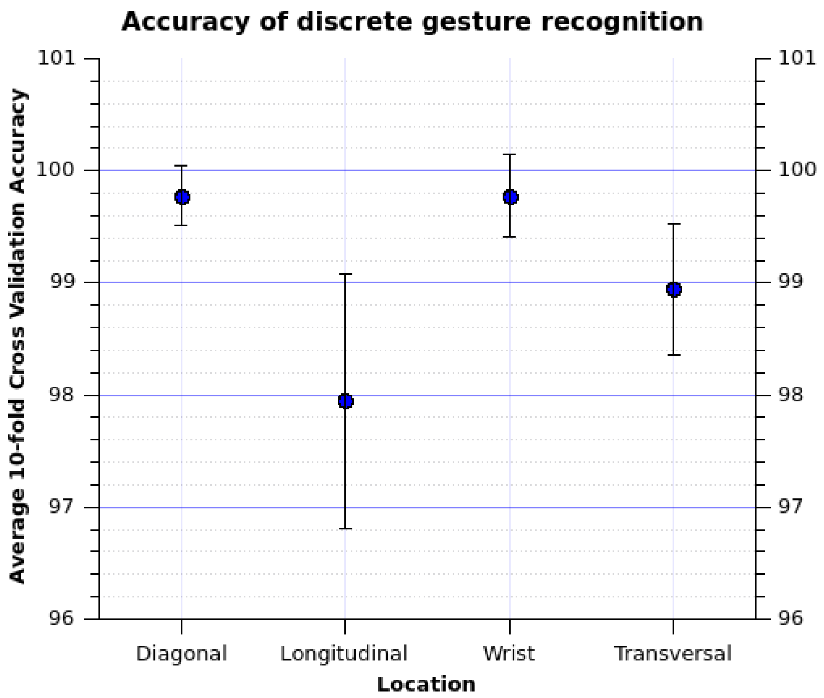

Figure 5.8: Accuracy of the discrete gesture classification. Average 10-fold cross-validation for each location across all participants. Error bars represent standard error.

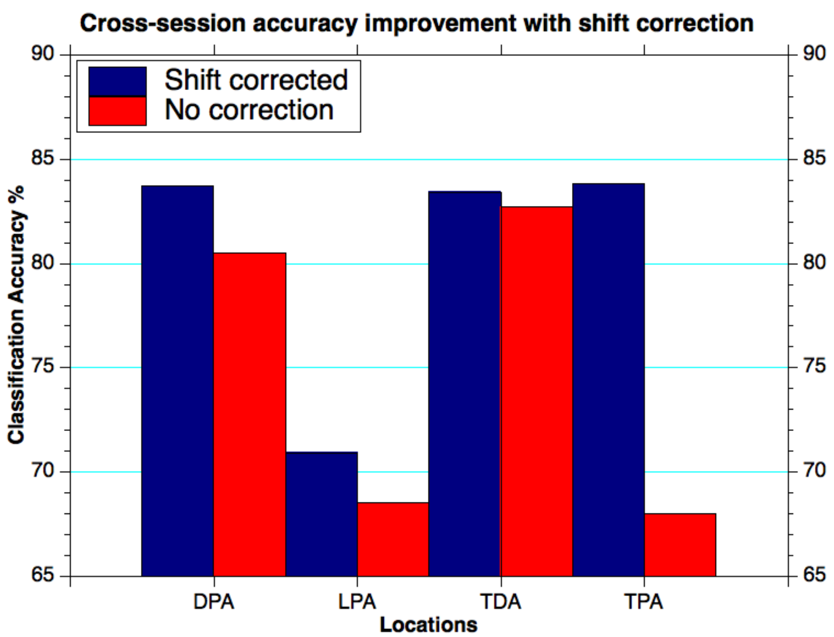

Figure 5.9: A graph plotting the average classification accuracies for the cross-session data at different locations.

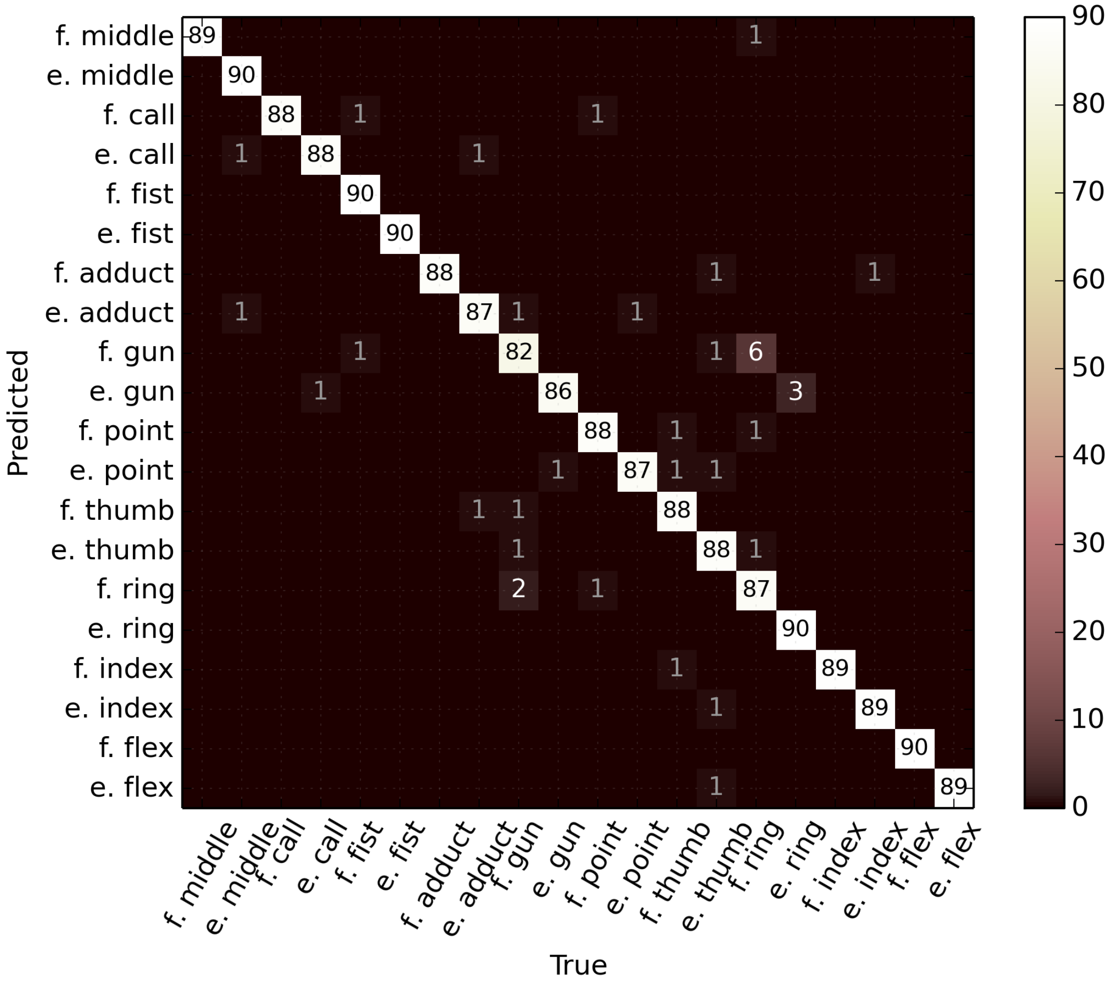

Figure 5.10: A confusion matrix for the Longitudinal position.

5.7 Discussion

The accuracy in discrete gesture recognition is very high, with the average of all locations being above 99%. The Longitudinal position has been statistically shown to perform the worst for this task. A more limited range of muscles are within the view of the probe at this position, supported by the fact that the gun and ring gestures were usually confused (for both flexion and extension - Fig. 5.10) since they involve a common muscle. Other locations showed similar issues, however the issue was more marked in the Longitudinal images. This may also be indicative of inconsistent or insufficient information about the fifth digit muscle across all the locations. In our observations of the images, the thumb and pinky muscles were at the extremities of the images, and it is likely that the pinky could have been out of view. Unexpectedly, the Wrist location offered good results, even when the images seemed to show little movement in comparison to the proximal locations. One possible reason for this is the good coverage of the tendons by the probe,

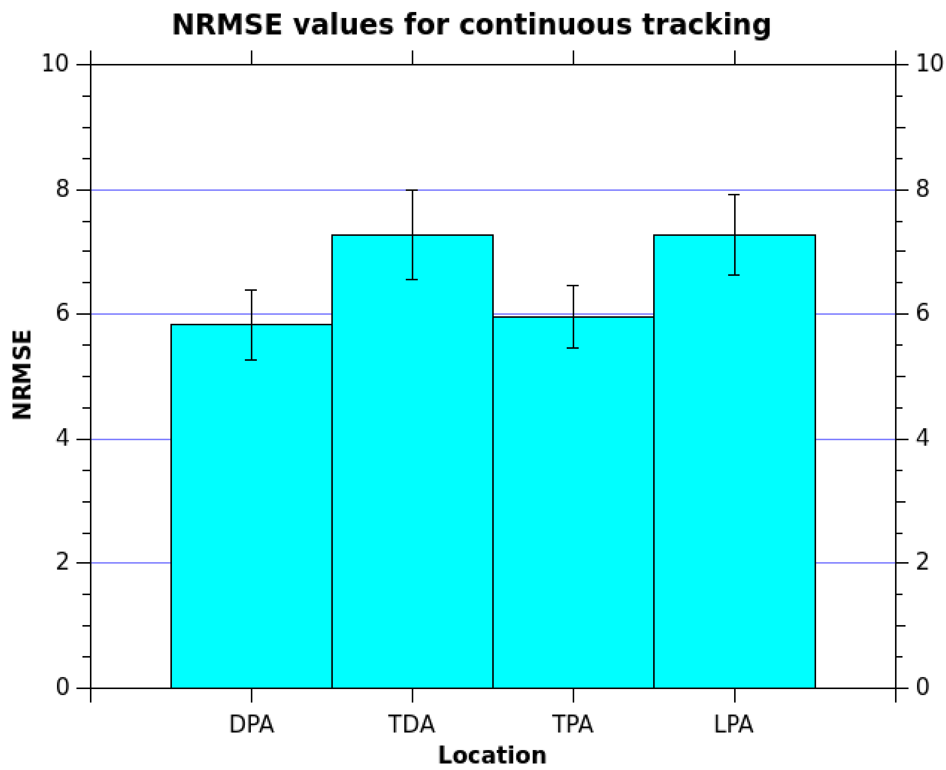

Figure 5.11: NRMSE values of the continuous digit flexion predictions averaged over all digits and for all participants. Error bars represent standard error.

which can easily cover the central part of the wrist due to the smaller surface area. The tendons that move the fingers and wrist have to pass through the carpal tunnel in the wrist making it a concentrated area of information, in contrast to the proximal locations where the muscles and tendons are spread across a larger area.

The cross-session results show that the loss in accuracy due to probe displacement between sessions affects each location differently. Prior to shift compensation, Transverse and Longitudinal had worse cross-session accuracy when compared to the other locations. This suggests that large shifts occurred between sessions for these positions, and that it may be more difficult to place the probe in the same place between sessions. The shift compensation algorithm improves the cross session accuracy for every location, but this improvement is much larger for Transverse. In contrast, Wrist only improves slightly but that position already offered high cross-session accuracy. This is likely because the smaller size of the wrist allowed less variation in the placement of the probe. Although there were large cross-session errors for Transverse and Longitudinal, the shift can only be compensated well for Transverse. The image shift with Longitudinal cannot be recovered because the image changes differently when the probe shifts laterally, and a simple 2d translation is not enough to correct this change. The other locations are more robust to image changes as displacements perpendicular to the probe do not affect the images significantly.

In summary, the results from the study suggest that there is no significant difference between Diagonal and Wrist in discrete gesture classification, but these are superior to other locations. For ergonomic reasons, Wrist seems like the best location for discrete gestures. For continuous detection of digit angles both Diagonal and Transverse offer the best results. Therefore, the best position for both requirements is Diagonal but if ergonomics is an important factor the Wrist location could be selected for discrete gesture or Transverse for continuous angle detection.

5.8 Limitations and Future Work

There are issues with practicality which are yet to be solved before this method is viable as a wearable gesture recognition device. In this section, we will address these limitations and provide information about the current state of research for each of these problems.

Gel is needed to couple the transducer to the skin and facilitate the transmission of ultrasound into the body. It may not be practical to use gel with an everyday wearable, but there have been recent developments which show that hydrogel pads can be used as an alternative coupling medium [59, 100].

Participants were sedentary during our experiments. Different postures will change the muscular disposition inside the forearm and shift the features. There have been some preliminary studies that investigates classification rates for different arm positions, which show that they do not seem to significantly compromise reliability [3]. Further comprehensive research is required to analyse and mitigate the effects of arm movement during everyday activities. It may be possible to use a similar method to our cross-session calibration for this purpose. Also, with inferences of posture from other sensors (e.g. IMU), specific calibration schemes could be applied.

In our experiments, we used a decade old ex-clinical scanner, we are aware of the elevated price and bulky size of these machines. Recently, emergency point of care ultrasonography has created a need for cheaper and smaller pocket-sized scanners; in fact, several portable devices exist [124]. These handheld devices utilise the power of mobile phones to process the raw data. Although they are quite small, they are still not small enough yet for integration with wearables. The image quality of these portable devices are comparable to the machine used in our studies. These devices are expensive (>$2k) but will decrease in price with acquisition boards integrated into a chip, and as research in transducer technology progresses [12]. As with most electronics, the cost of the raw materials of these devices is low, but the long and difficult fabrication process makes it expensive.

Since the wrist requires a shorter imaging depth, it is possible to use a probe with a higher frequency. This higher frequency may yield better results with the increase in resolution. Higher frequency probes can be smaller in size, making it easier to integrate with wearables.

The probe that was used in this study was large and rigid, hindering its utilisation in wearable scenarios. Transducer fabrication research has developed small and flexible thin-film probes, which should help to reduce the size. Currently, it is possible to build a flexible probe with multiple elements that can be wrapped around the thumb [129]. This technology may also help to alleviate the aforementioned arm movement artefacts. It would seem more suitable if the probe was designed in a way which fits the curvature of the forearm. It has also been shown that piezoelectric elements can be 3D printed, which may allow for a custom-tailored design of a wearable transducer device[25].

The US probe that we used is linear; that is, it has the elements arranged in one dimension. 2D probes that can image 3D volumes (while not moving) is a current interest in ultrasonography research. These probes could improve robustness to shifts in the positioning of the probe, as well as providing more features to classify with.

Always-on gesture detection is desirable for real use. We do not have a dedicated study for proof of gesture spotting, but we found that high levels of activity can be measured by integrating the magnitudes of the optical flow vectors. While we only used this method to split our data, it may also be used as a variable for segmentation. However, this simple approach could be brittle and remains to be tested under realistic conditions, where motion artifacts may cause problems.

Power is a concern for wearable ultrasound imaging devices at the moment. The portable GE Vscan [34] lasts for $\tilde{1}$ hour of continuous use. Battery improvements are expected in the future, but a more interesting approach is to only activate full imaging during gestures, for instance pulsing ultrasound with fewer elements, effectively providing a low-res image to detect gesture onset. A more complete system could have other low-power sensors integrated, such as myoelectric sensors to initially spot the gesture using existing segmentation techniques.

All of the presented experiments are within-user. Much like electromyographic devices, anatomical differences between users creates an interesting challenge for cross-participant use of US wearables. Perhaps a more elaborate classifier and calibration scheme that takes into account the common anatomical features between users, could lead to an effective user-independent system.

Bio identification is could also be another possibility with ultrasound imaging. The veins and other anatomical features are detectable with high frequency probes; previously, it has been shown that the structure of the veins can be used as an unique characteristic for identification [68]. This technique could be used to identify the wearer and load their preferences, without the need for external validation.

5.9 Reflection on Findings and the Key Problems

We have presented a solution for detecting discrete gestures and tracking continuous angles of the fingers using ultrasound images captured from a probe mounted on the forearm. Our novel contributions include findings on the variation in performance between different mounting locations. In contrast to previous studies, we provide results for the cross-session accuracy decrease, and show that a simple calibration algorithm can improve the accuracy, and highlights the usefulness for future work. In conclusion, the performance variation across the tested locations vary somewhat insignificantly, meaning that an ergonomic location such as the wrist may be chosen as the desired location for a wearable US device. However, there are some differences in robustness and continuous angle recognition, and depending on the requirements other locations may be more desirable.

As discussed, the size, complexity and cost of current state-of-the-art devices are still an issue for wearable applications. There is hope that with the current trends of continued decrease in cost and size of these devices, ultrasound imaging for gesture detection will become more feasible in the future. In the next section, we will look into this issue in more depth. We present an alternative probe design that is far smaller and simpler in complexity, which may be an alternative to imaging that still provides enough information for accurate hand pose detection.

The ultrasonic gel that we discussed as a limitation, is actually a key practicality issue of this method. Unfortunately, this is largely ignored by those studying ultrasonography for non-medical purposes. Of course for clinical purposes, the application of gel suffices for the quick scanning time necessary for medical diagnosis by the radiographer. For applications such as wearable device control, this is obviously an issue if the device is to be used for extended periods of time. With EMG, coupling is also an issue, but the difference is there have been dry electrode designs that have been tested successfully. However, there is much less evidence for a dry alternative with US. The closest such work on rigid coupling mediums are polyacrylamide gels [100] for focused ultrasound therapy (not sensing). On the other hand, there are many patents that can be found for partially rigid coupling mediums. Due to the lack of information of rigid gels for ultrasonic coupling (for sensing), we investigate this ourselves in the next part of this chapter.

5.10 Improving the Feasibility of Ultrasonic Hand Tracking Wearables

Ultrasonic imaging suffers from a couple of issues: First and foremost, the propagation of ultrasound into flesh suffers greatly without a suitable coupling medium; Secondly, the complexity of the driving circuitry for medical grade imaging currently renders a wearable version of this infeasible. In part 2 of this chapter, we aim to address these two problems by finding a rigid coupling medium that lasts for significantly longer periods of time; and devising a new sensor configuration to reduce the device complexity, while still retaining the benefits of the technique. Furthermore, a comparison between high and low frequency systems reveal that different devices can be created with this technique for better resolution or convenience respectively.

The main practicality problems of ultrasound are stated below:

1. An ultrasonic coupling medium. Due to the gaps of air and large difference in acoustic impedance between the skin and the imaging probe, most of the signal is lost. This effect

Figure 5.12: The GE Vscan, a modern portable ultrasonic imaging machine.

is greater at higher frequencies, which are used in ultrasonic imaging. To attain a clear image, a liquid gel is typically applied to facilitate the coupling between the skin and the transducer. The gel usually dries up within 20 minutes. Thus any device based on ultrasound imaging will surely be limited by this impracticality, if intended to work for longer periods of time.

2. The sheer complexity of an imaging transducer creates an enormous challenge in reducing the size to a wearable form factor. Commercially available portable scanning machines that can be bought today are still too large for a wearable device [34] (Fig. 5.12). Furthermore, the power consumption and computational complexity of these units are still far too high for today's battery and processing technology.

While one may argue that (2) could eventually be solved with enough time and engineering, (1) presents a fundamental problem of the technique that cannot be solved without further research. Thus, it can be said that (1) is a more critical problem.

The contributions of this work are summarised as follows:

- A rigid hydrogel couplant is found to increase the transmission of ultrasound into flesh. The hydrogel can be used for extended periods of time, for up to 3 hours.

- A new sensor configuration that uses fewer sensors is proposed. This configuration requires far less in terms of computation, efficiency, driving circuitry complexity and cost.

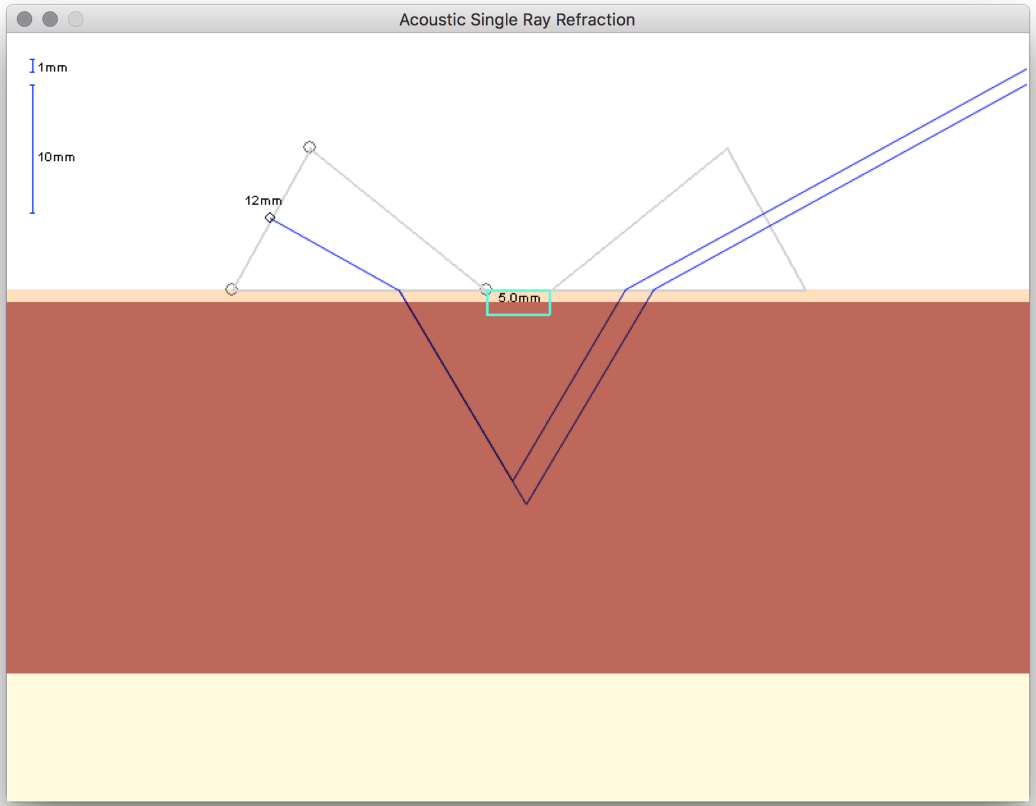

Figure 5.13: A simulation to show the refraction of the main ray from the piezoelectric elements into the flesh.

5.11 Hardware

For the ultrasonic transmitter and receiver, we used 1MHz 10mm diameter piezoceramic discs, with wrap-around electrodes available from Noliac [91] (Also has resonant frequency at 200KHz). The discs needed to be encased in an insulator in order to prevent shortages between the electrodes, for this we used epoxy. For our experiments, we used separate pulse and receive machinery, and thus also separate transmitting and receiving piezo elements.

Consequently, the piezo's needed to propagate the sound waves at an angle into the forearm. After inspecting the anatomy of the forearm through ultrasonic images, we found estimates for the depth where muscle activity can be measured. We then developed software to approximate these angles, and created a silicone mould with the shape and size of wedges determined by the software (Fig. 5.13). The mould was created by first cutting a piece of acrylic with the desired shape, then using silicone (EasyMold) around all but one face. We then cast the epoxy into these

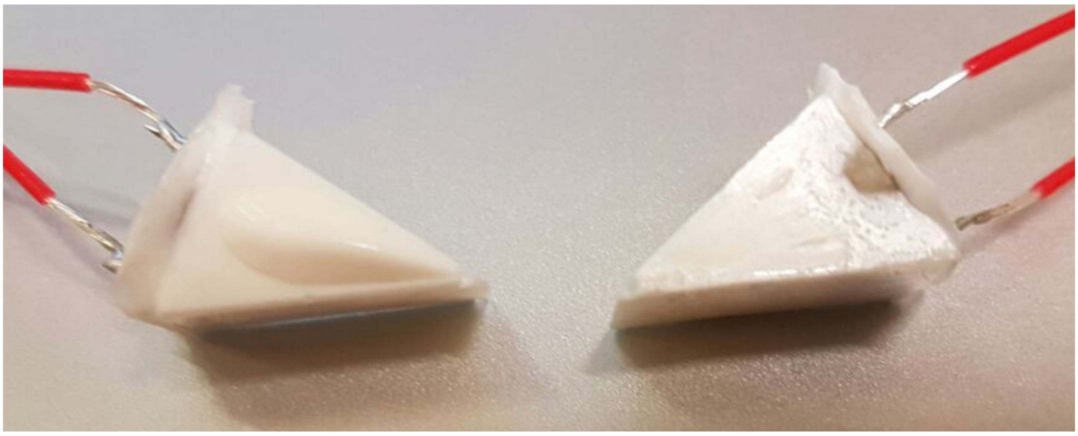

Figure 5.14: Piezoelectric discs, encased in epoxy that are shaped into wedges that are determined by the software.

moulds, with the piezo resting on the open face (Fig. 5.14). We 3D printed a structure to separate elements but hold them at a fixed orientation and distance apart (Fig. 5.16).

For the equipment to generate ultrasonic pulses, we used an Agilent 33220A arbitrary waveform generator (set to pulse 10 periods, sinusoidal, 10V pk-pk). The pulse sync was set to trigger the oscilloscope, Agilent DSO-X 2024A, so that the output of the receiving transducer could be seen directly after the pulse was sent.

5.12 Experiments

5.12.1 Hydrogel

There are many types of hydrogels, but they are all capable of absorbing large quantities of water within their structure (sometimes over 99%) [2]. They retain this water for long periods of time, which is an ideal property for ultrasonic applications that require long scanning times (unlike clinical scans).

Here, we test a calcium alginate hydrogel. This is comprised of alginic acid and calcium chloride. The alginic acid is in fact a biopolymer, which means that it is biodegradable, renewable and food safe. The calcium chloride is the cross-linker, which connects between the negative polymers. We tested numerous quantities of each and settled on using a mixture of 450mg of alginic acid to a 15ml solution of CaCl with water (~0.01 Molar). The mixture was mixed together gradually using a magnetic stirrer. This mixture is then left to hydrate for 24h, after which the mixture is then quite viscous. The mixture is cast into shape using sheets of acrylic, then exposed to a higher concentration of CaCl for a further 24h (approximately 10x higher concentration). The gel is then left in water for 24h to get rid of any unbonded calcium ions. This results in a hydrogel that is mechanically strong enough for our desired application (Fig. 5.15).

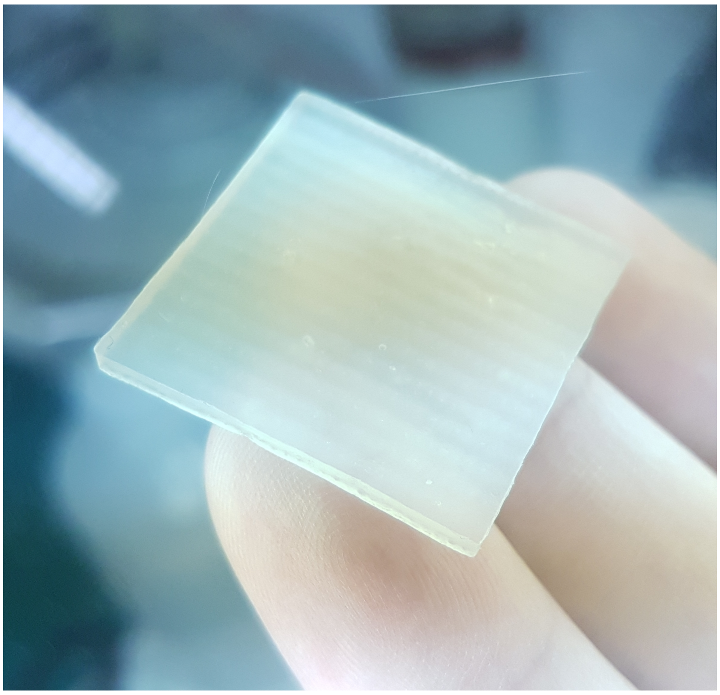

Figure 5.15: Alginate hydrogel.

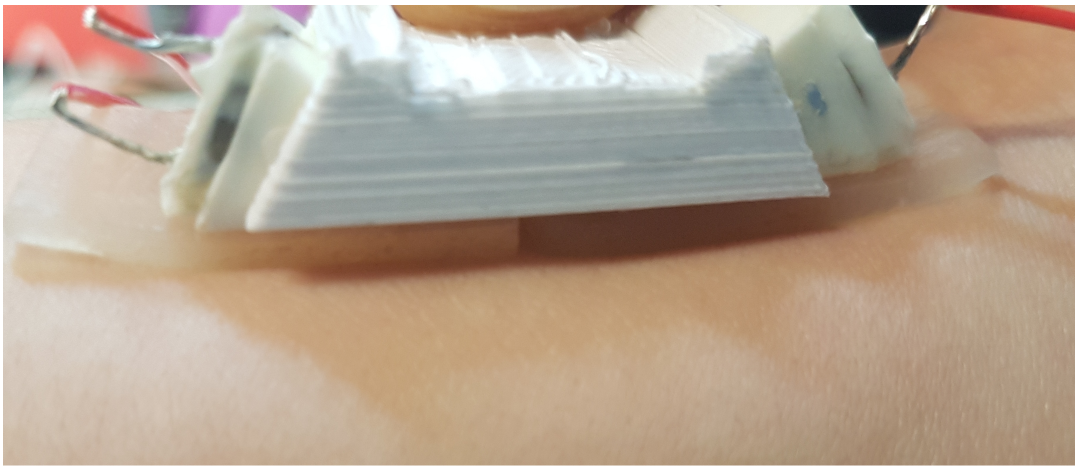

Figure 5.16: The prototype and experimental setup, the piezo discs inside wedges, held together with 3d printed housing, with hydrogels beneath them.

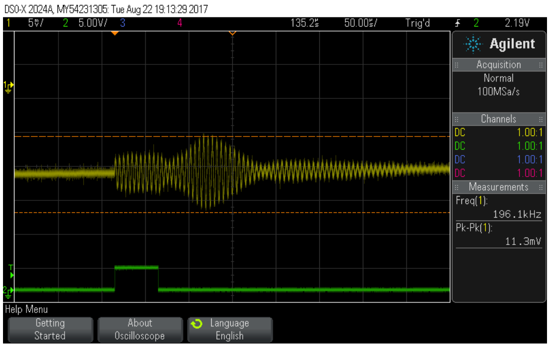

Figure 5.17: 200KHz received signal, without hydrogel.

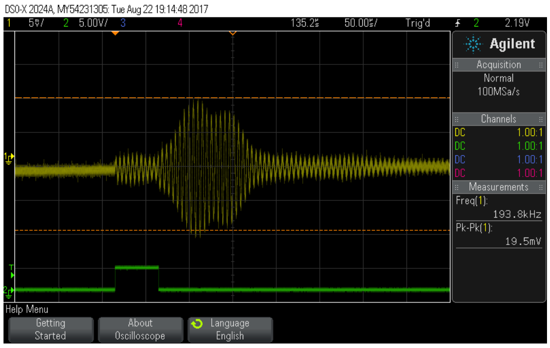

Figure 5.18: 200KHz received signal, with hydrogel.

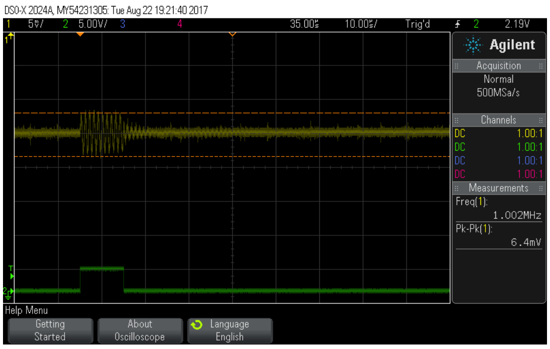

Figure 5.19: 1MHz received signal, without hydrogel.

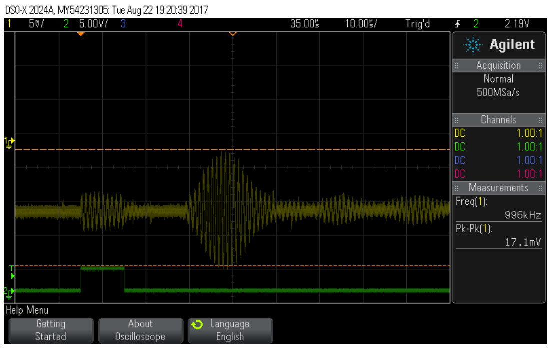

Figure 5.20: 1MHz received signal, with hydrogel.

Figure 5.21: Testing the effectiveness of a hydrogel coupling medium at 200KHz and 1MHz. The line underneath shows the trigger duration from the pulser.

Table 5.3 shows our tests with the hydrogel on the wrist. We compare the peak-peak voltage of the received echo with and without the hydrogel. The signal captured from the oscilloscope can be seen in Fig. 5.21. The hydrogel was measured after every hour, for 3 hours.

5.12.2 Finger flexion

The signal varies quite significantly between the different frequencies. On one hand, the 1MHz signal captures many more features of the muscles due to the higher resolution. However, the signal to noise ratio is better when 200KHz is used. Therefore, the system is likely to be more robust when using lower frequencies. In either case, flexing an individual finger results in a gradual change in the signal whether using 200KHz or 1MHz.

Fig. 5.22 shows the scope readings when exciting the transducers at 1MHz, using the hydrogel couplant. The line at the top shows when the hand is open, and progressive flexing of the index finger are measured until it is fully flexed (bottom line). The purple rectangles highlight the

	Without hydrogel	With hydrogel
200KHz 1MHz	11.3mV < 2mV	19.5mV 17.1mV
1 hours later		
200KHz 1MHz	n/a n/a	18.1mV 12.5mV
2 hours later		
200KHz 1MHz	n/a n/a	15.3mV 7.8mV
3 hours later		
200KHz 1MHz	n/a n/a	11.8mV 4.0mV

Table 5.3: The highest peak-peak voltage of the received signal

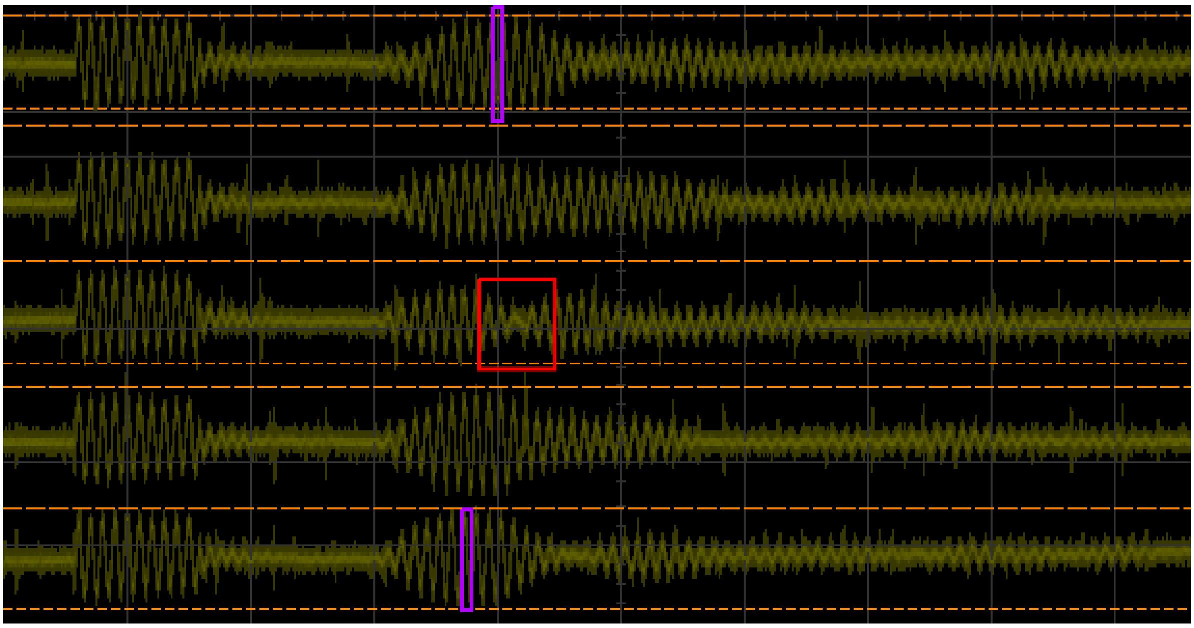

Figure 5.22: Gradual flexing of the index finger, from relaxed (top) to fully flex (bottom).

peaks of the first reflection; a slight shift in phase can be seen throughout flexing. There are also other signs in the signal that infer changes in muscle activity, as highlighted by the red square. This qualitatively shows that it is possible to detect finger flexion with a single pair of emitting and receiving piezo elements, unlike imaging transducers which are comprised of hundreds of elements [81].

5.13 Discussion and Future Work

The hydrogel experiments confirmed what was previously known about higher frequency ultrasound applications, which is that without gel, the signal cannot be observed. The hydrogel improves the signal to noise ratio for both 200KHz and 1MHz, but importantly the signal can be clearly seen at 1MHz. The hydrogel does decrease in performance over long duration's and would require a method of re-hydration for day to day usage, but this is a large improvement on the current standard which dries rapidly. Future work in this area should look into other types of hydrogels, and vary parameters to improve the working duration.

The single emitter and receiver configuration demonstrates the possibility for a system to accurately determine the amount of digit flexion. This is a starting point for what could be a highly accurate hand tracking wearable and would be very useful for high fidelity hand tracking applications like augmented reality, where a wearable system could be a necessity. Such a system would also be far less complex than medical ultrasonic imaging systems, which uses hundreds of elements. However, the question remains as to whether it is possible to use multiple pairs of emitters and receivers in order to estimate the angles for each finger individually. This is the obvious next step for such a system. The next iteration of the system would probably use smaller piezos, so that they can be formed into a line along the wrist, where each piezo pair concentrates on a particular digit.

When the device is used at 200KHz, the received signal might lack detail for individual finger flexion across multiple fingers, but would probably still work for discrete gesture detection. Surprisingly, the signal can still be seen relatively clearly without the gel at this low frequency. This opens up two separate branches of the technology: one for discrete gesture detection without the need for gel; another for higher fidelity hand tracking for AR/VR, continuous variable control or implicit gesture detection (weakly performed) but requires gel.

5.14 Conclusion of Feasibility Experiments

We have shown that there are alternatives to traditional transducer design and liquid gels, with experiments to show their benefits. These alternatives give greater hope regarding the feasibility of high fidelity wearable ultrasonic hand tracking wearables. This work provides other researchers the opportunity to explore multiple avenues of research. Low frequency ultrasonic systems up to 200KHz can be explored for convenient gesture detection without gel. High frequency systems affording higher resolution hand tracking can be pursued, requiring further research into hydrogel and signal processing.

5.15 Reflection and adjustments to the requirements

The main findings of this chapter were that ultrasonographic device placement is not a significant variable and therefore many positions and orientations performs well. Importantly, this includes the wrist location, even performing favourably in some tests. We confirmed it's effectiveness in detecting continuous finger angles, which provides utility in the applications that EMG performs poorly at. As mentioned at the end of the previous chapter, these are applications that involve implicit gesture detection, weakly performed or subtle gestures and control of continuous hand tracking for control of continuous variables or for interaction in AR/VR. Then we went further to show that US systems are fairly robust, with a simple algorithm to mitigate device shifting when used between sessions.

This is what we initially set out to investigate, but perhaps the more unique and interesting findings stem from the analysis of whether current state of the art systems can integrate with wearable form factors yet. The answer to this was no: it is clear that US imaging systems are very complicated and required a lot of electronics to drive the transducers. A more fundamental problem with ultrasound devices was identified, which is the use of gel coupling. Gel coupling is absolutely necessary to get any images from the device when used at high frequencies, which are typically used in musculoskeletal imaging. While gradual improvements to engineering may eventually allow integration of high frequency, high element-count transducers into wearables, the issue of gel is one which does not seem to change in time without active research into material science of rigid gels. With very little knowledge of material science, but with the help of colleagues familiar with hydrogels, I was able to fabricate my own and report findings which indicate the possibilities of using a partially rigid coupling, longer lasting medium for ultrasonic imaging. Coupled with the findings of using lower-frequency piezoelectric transducers in different configurations, the feasibility of integrating such a device with wearables is now not such a far-fetched idea. But a fully working prototype that uses rigid and non-wet gels is still far away from complete.

The issue of device coupling to the skin is not an issue that is limited to EMG and ultrasound. Looking more deeply into the literature, there are also coupling issues with pressure sensing and also infrared devices. But in particular, dry electrodes that are used in EMG and EIT, must be very tight against the skin in order to maintain good electrical conductivity. Furthermore, any loosening in tightness will affect the measurements, likely inducing a worse signal to noise ratio as we noticed ourselves during the pilot study in the EMG chapter. Unfortunately, this phenomenon has not been reported in much of the research in EMG or EIT. Although, there are a few examples of work that try to increase the comfort of dry electrodes over extended periods of time, since the pressure of the electrodes onto the skin are so tight (but necessarily so) that they leave a mark [24]. Thus far, we have not considered the comfort of the wearable to be of much variation between the different methods. However, given: how different the coupling requirements are between methods; the different options (dry vs wet) within methods, the trade-off between

Technique	Mobility	Placement & Size	Fidelity	Comfort	Complexity
EMG	Average	Wrist	Good	Poor	Simple
Ultrasound	Poor	**Wrist**	Very good	Poor	Very complex
Infrared	Highly	Thin & wrist	Poor	**???**	Very simple
Pressure	Highly	Thin & wrist	Poor	Poor	Very simple
EIT	Highly	Thin & wrist	Average	Poor	Fairly complex
Optical	Average	Wrist (but obtrusive)	Limited	Good	Fairly complex
Glove	Highly	Obtrusive	Excellent	Good	Simple

Table 5.4: An updated table of analysis for gesture recognition techniques, with respect to the requirements.

device tightness and signal quality, we must now consider this to be one of the requirements. Therefore, we add a new requirement to our previously amended list of requirements, changes marked in bold:

- Mobility → Wearable.
- Placement & Size → Ergonomics, practicality and social acceptability.
- Accuracy & Range of Gestures & Continuous and Weak → Usefulness.
- Continuous control or subtly performed gestures.
- **Device tightness → Comfort.**
- Complexity → Immediate feasibility.

We will now update our table of analysis with this new requirement, but also with our findings from the earlier parts of this chapter.

Admittedly, the device tightness may not be the only factor that affects comfort, but it is the main factor that we have observed during our hands-on experiences with the technologies. We will assume that the dry electrode approach is the only reasonable option for wearable devices. Therefore, in the table, EMG and ultrasound methods have been categorised as having poor comfort due to the electrodes requiring tight skin contact. EIT use dry electrodes similar to EMG and have also been marked as requiring tight skin contact. During our testing with pressure sensors, we also noticed that the sensors needed to be pressed tightly against the skin or else the measurements would be less reliable. Of course with the camera/optical approach, there is no need for coupling of the sensor to the forearm, therefore it is good in this regard. Similarly with the glove. However, we did not have any hands on experience with infrared, so whether this technology works well without requiring a tight coupling of sensors to the skin is question to be answered. Therefore, our next logical step looks into filling in this blank spot in the table, i.e., investigating whether infrared can be used without being worn so tightly. In addition, we also look to see if we can improve the accuracy and range of gestures using infrared, as the fidelity is poorer when compared to EMG, US or EIT.

Chapter 6

Infrared

6.1 Introduction

Having discovered that the coupling of electrodes to the skin is often too tight, this challenges EMG, ultrasound and EIT as technologies that are comfortable for long term use. This is evidenced in Chen et al. work as they design for softer, more comfortable electrodes [24]. Although this problem is often ignored by researchers in this field, it is sometimes picked up by reviewers of off-the-shelf devices. For instance, here is an excerpt from a review of the "Myo" armband, by Thalmic Labs[1]:

> "The sensors need to be tight against the skin to provide accurate feedback ... And when I say the fit need to be tight I mean it. It's comfortable enough to wear initially, but I do find myself itching to get it off after half an hour or so. When you do you're left with a skin imprint resembling a tribal tattoo from the '90s."
>
> "Verdict: It's not very comfortable and sometimes the controls aren't particularly refined"

The Myo (at the least in the version that is reviewed here, reviewed in 11/2015), uses dry electrodes in order to circumvent the practicality and longevity issues of gel electrodes. As the reviewer points out, based on his experience, the electrodes need to be tightly pressed against the skin in order to work sufficiently well. After using the device somewhat uncomfortably, an imprint is left on the skin, as seen in figure 6.1.

[1] An excerpt from a user review of the Myo armband, accessed 05/04/2018: `http://www.trustedreviews.com/reviews/myo`

Figure 6.1: An imprint of the dry electrodes on the skin of the user, after having it worn it for a short session (less than 1 hour).

Infrared sensing technology for gesture detection is different because it does not require good contact for electrical conductivity or acoustic impedance matching. However, whether infrared works with skin-tight contact remains a question that is to be answered within this chapter.

In the following sections, we present SensIR, a bracelet that uses near-infrared sensing to infer hand gestures. The bracelet is composed of pairs of infrared emitters and receivers that are used to measure both the transmission and reflection of light through/off the wrist.

Previous work have used arrays of infrared (IR) emitters and receivers around the wrist to measure the amount of reflected infrared light to infer wrist deformations. However, these systems only take a single receiver measurement per emitter. We propose to take measurements between all the possible combinations of emitters and receivers, capturing not only the reflected light but also the amount of light transmitted through the wrist. Transmission of infrared light through human tissue is relatively high and does not pose any danger at the levels that we use.

Unlike systems such as EMG or EIT, infrared does not rely on high electrical conductivity between the sensor and the skin. This allows infrared systems to be far more comfortable, as we show in this paper, the sensors do not need to be skin-tight, nor do they need to occupy much space.

In the following sections, we present the hardware and software behind SensIR. We conducted a user study in which the system provided an accuracy of 93% for 12 gestures. We demonstrate the importance of including the transmitted light features and analyse the performance of different bracelet arrangements to guide the design of wearable gesture recognisers that use SensIR's approach. Finally, we explore robustness issues that should be considered for using the device in real scenarios.

The content of this chapter is a revised version of work I have published in the Proceedings of the 2017 UIST Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17): "SensIR: Detecting Hand Gestures with a Wearable Bracelet using Infrared Transmission and Reflection" [80].

6.2 Related Work

In this section we focus on describing previous work on infrared gesture detection devices. Detailed descriptions of the related work can be found in chapter 3.

Although we name the technique infrared, this is different to what we call "optical" methods (as we have defined in this thesis at least), which use cameras and can be used to determine the position and flexion of fingers [101]. For instance, Digits [65] uses an IR camera attached to the wrist. A problem of these approaches is that the camera requires a direct line of sight to the fingers and therefore camera approaches suffer from occlusions by the hand.

While performing gestures, the shape of our wrist and forearm changes. This can be measured to infer hand gestures. A more detailed explanation of this can be found in the literature review in section 3.1.1, as well as an overview of methods reliant on wrist deformations.

Infrared distance sensors have been utilised in past work in order to detect such wrist deformations. Fukui et al. created a prototype with 150 pairs of emitter-receivers were placed around the wrist to detect 5 gestures with 70% accuracy [42]. Hamid et al. [53] showed qualitatively that it would be possible to differentiate between 10 gestures. Ogata et al. [92] augmented a smartwatch with 12 sensors to detect 9 different skin deformations around the watch. Gong et al. [46] placed 12 sensors around the wrist for detecting 8 gestures with 89% accuracy. IR sensors are inexpensive and easy to integrate into small wearable systems and have better robustness against coupling conditions. However, the accuracy and amount of gestures seem limited compared to other methods.

Previous IR approaches only emit and receive with the same sensor. In this work, we demonstrate that using infrared light enables a novel possibility that can be used to improve accuracy. Infrared is a non-ionizing radiation and penetrates through flesh with relatively high transmission (Fig. 6.2). We propose to emit and receive with all possible combinations of transmitters and receivers as light passes diffusely through human tissue, obtaining exponentially more features to analyse.

Near-infrared diffuse tomography is a medical imaging technique [132] in which infrared lasers are pulsed at picoseconds periods through optical fibers to image human tissue with low-resolution. We think that a similar principle can be applied to obtain a wearable gesture detector that is accurate, inexpensive, ergonomic and resistant to bad coupling.

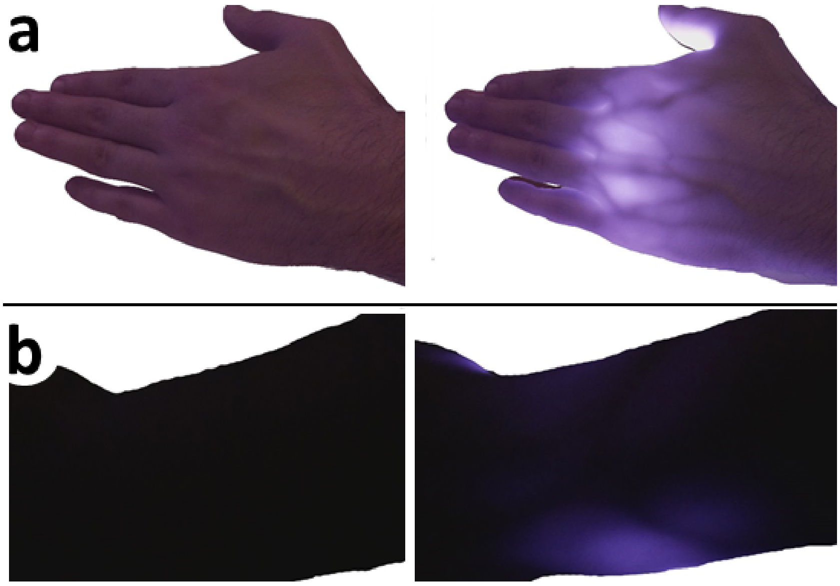

Figure 6.2: Images captured with a camera sensitive to infrared. On the left there is no infrared illumination, on the right an infrared source is placed behind the hand, showing partial transmission through the hand (a) and wrist (b).

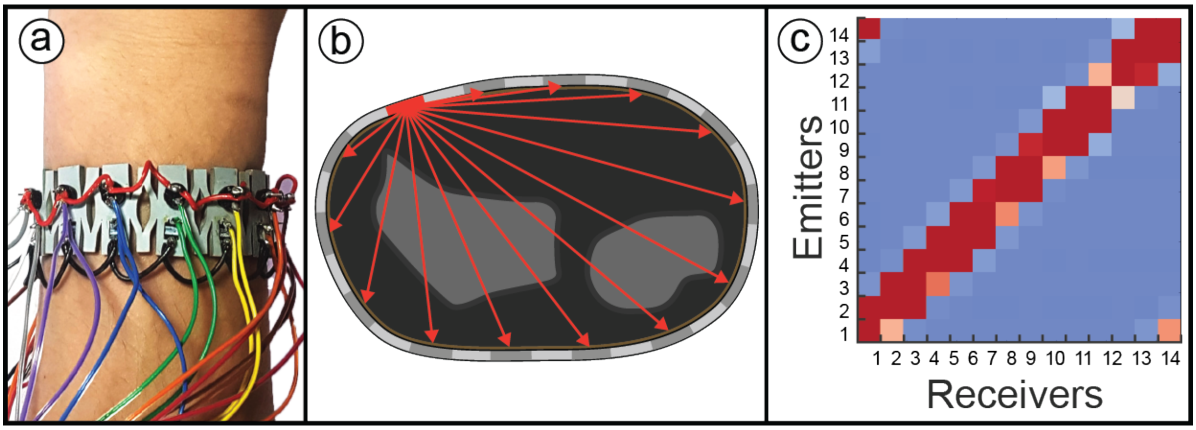

Figure 6.3: a) SensIR worn by the user. b) An emitter is on and all the receivers capture the light level. c) This is repeated for all the emitters to produce a full matrix of data.

6.3 SensIR

SensIR is a bracelet made of 14 segments that is placed around the users wrist (Fig. 6.3.a) Each segment has an infrared emitter and receiver, measurements between all possible pairs are captured (196 measurements) (Fig. 6.3.b). When the emitter/receiver pair are close, most of the light is reflected from the skin. When they are opposite to each other, the light transmits diffusely through the body. The light levels (i.e. features) are fed into a neural network that infers the current gesture.

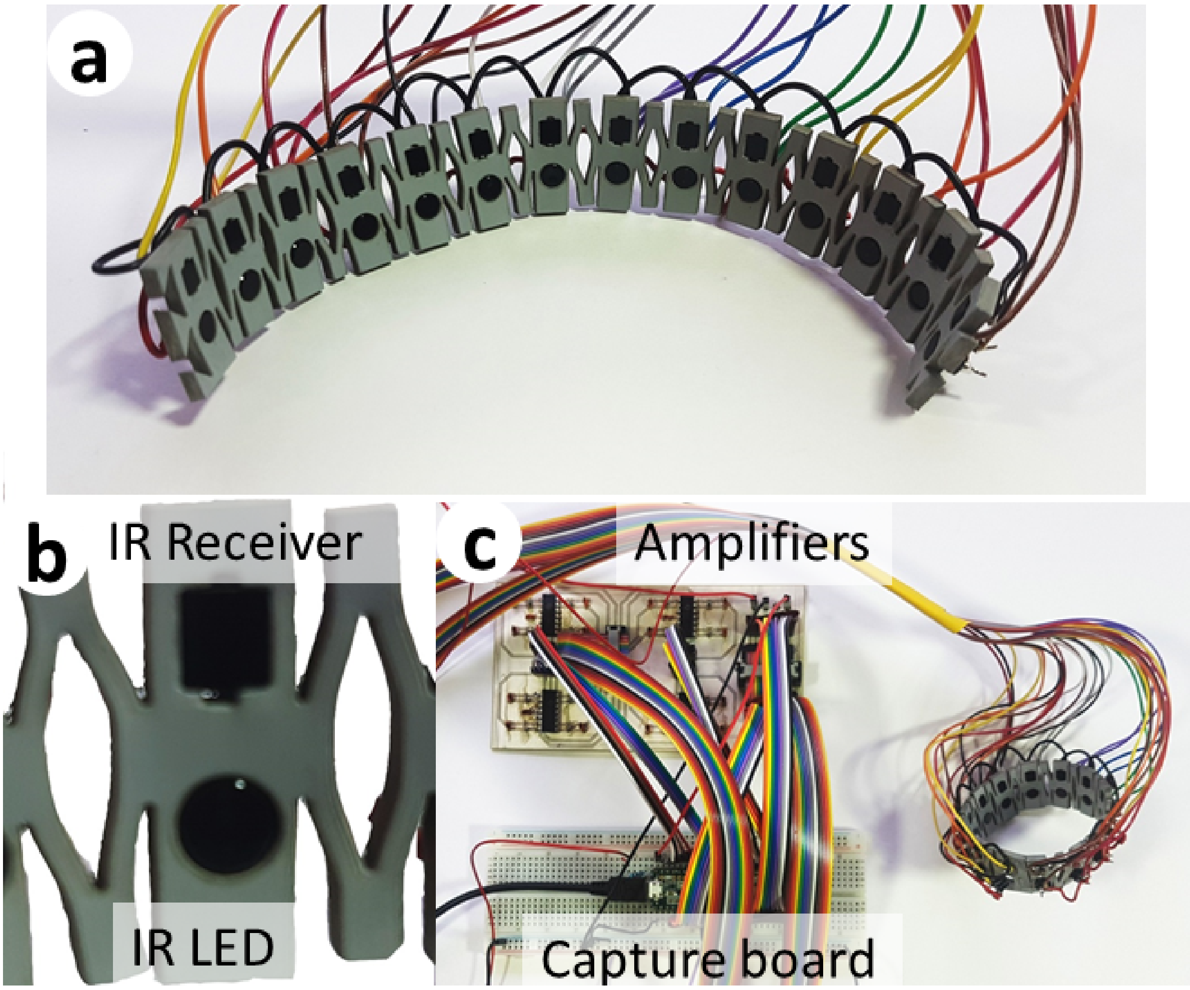

Figure 6.4: a) bracelet made of 14 segments. b) individual segments with an emitter and receiver. c) circuit board used to amplify, sample and send the received signals to the computer.

6.3.1 Hardware

In this section, we will describe the hardware of the system. This will hopefully serve as a guide for people to follow should they wish to replicate our prototype. To this end, we recommend they use the supplementary material found in the ACM digital library[2]. This material contains design files and software which will be referenced in the following paragraphs.

The bracelet was lasercut in one piece of laser rubber (Hobarts) with holes for the emitters and receivers. Each segment had a pattern of curls to permit stretching and bending of the bracelet (Fig. 6.4.a). The design file for this can be found in the source materials as `"bracelet_cut_rubber.ai"`. The emitting elements were LEDs (Osram Opto SFH 4556P, 860nm) and the receiving elements photodiodes (Osram Opto BPW34 FA IR) (Fig. 6.4.b) Opamps (LM324N) with a gain of 1M were used as trans-impedance amplifiers to measure the current

[2]Found under "Source Materials": `https://doi.org/10.1145/3126594.3126604`

from the diodes, a transistor array was used to amplify the control signal for the LEDs. A Teensy 3.6 was used to control the LEDs and read the values from the photodiodes (Fig. 6.4.c) using its internal ADC of 16 bits and a reference voltage of 3.3V. We waited 10ms to enable the LEDs to reach full power and then read with the photodiodes. All 196 measurements were taken 20 times per second. The system consumes 110mA operating at 4.5V (450mW) and 63% of the consumption comes from the microcontroller.

6.3.2 Software

The Teensy microcontroller controlled the switching of the LEDs and the sampling from the photodiodes. The measurements are sent to a PC running a neural network algorithm to classify the features into gestures. We used a multilayer-perceptron (MLP) classifier in one-versus-rest mode (Scikit-learn) with 1 hidden layer of 24 neurons, L-BFGS training algorithm and an alpha parameter of 0.05.

6.4 User Studies

6.4.1 Main Study

6.4.1.1 Procedure

Our main study explored the accuracy of the system in different configurations.

6.4.1.2 Participants

10 participants took part in the study, aged between 24 and 32 (6 male, 4 female).

6.4.1.3 Task

They were seated in a chair with the bracelet worn on the wrist of their dominant hand. They were asked to perform the gestures, in the sequence shown in Fig. 6.5, 10 times. That is, 10 participants X 12 gestures X 10 repetitions = 1200 gestures.

6.4.1.4 Experimental Design

The study and analysis of the data was performed within-user, and training of the classifier is user dependent due to anatomical differences between users. We chose to use a variety of finger and wrist gestures that are commonly found in related work.

6.4.1.5 Data Processing

Cross validation was performed using the MLP classifier described earlier employing a 10-fold leave-one-out cross-validation scheme, with each fold containing one instance of every gesture.

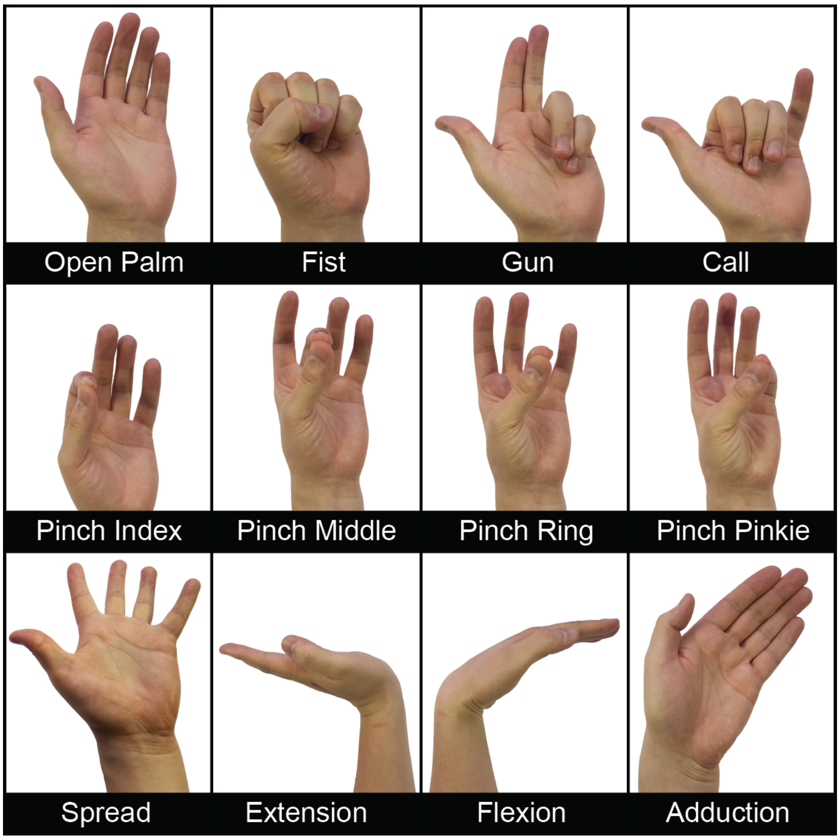

Figure 6.5: Gesture set used for the study.

Since the gestures were performed in sequence, gestures were not temporally adjacent across the test and training sets.

6.4.1.6 Results

The obtained accuracy was 93.3% (SD=3.49), the confusion matrix is shown in Fig. 6.6. Most confusions were between the pinch gestures, presumably due to their common muscle groups used to perform them. Surprisingly, misclassifications occurred between the open palm and pinch gestures. This suggests that the features for the pinch gestures are also less pronounced, we noticed that during the study some participants performed the pinch gestures with less emphasis than others.

The accuracy for different arrangements and number of sensors is presented in Fig. 6.7. For each configuration there are two results, in the first one all the features were used (i.e. both reflective and transmissive); in the other, only the reflective features were used (as used in previous work). The accuracy was significantly greater using all the features 93.3%(SD=3.4) than with only reflective 68.3%(SD=27.0), $t(9)=2.964$, $p=0.016$ for 14-segments, 89.0%(SD=5.9) > 63%(SD=26.9), $t(9)=2.921$, $p=0.017$ for the Smartwatch, 84.0%(SD=9.6) > 50.8%(SD=29.9),

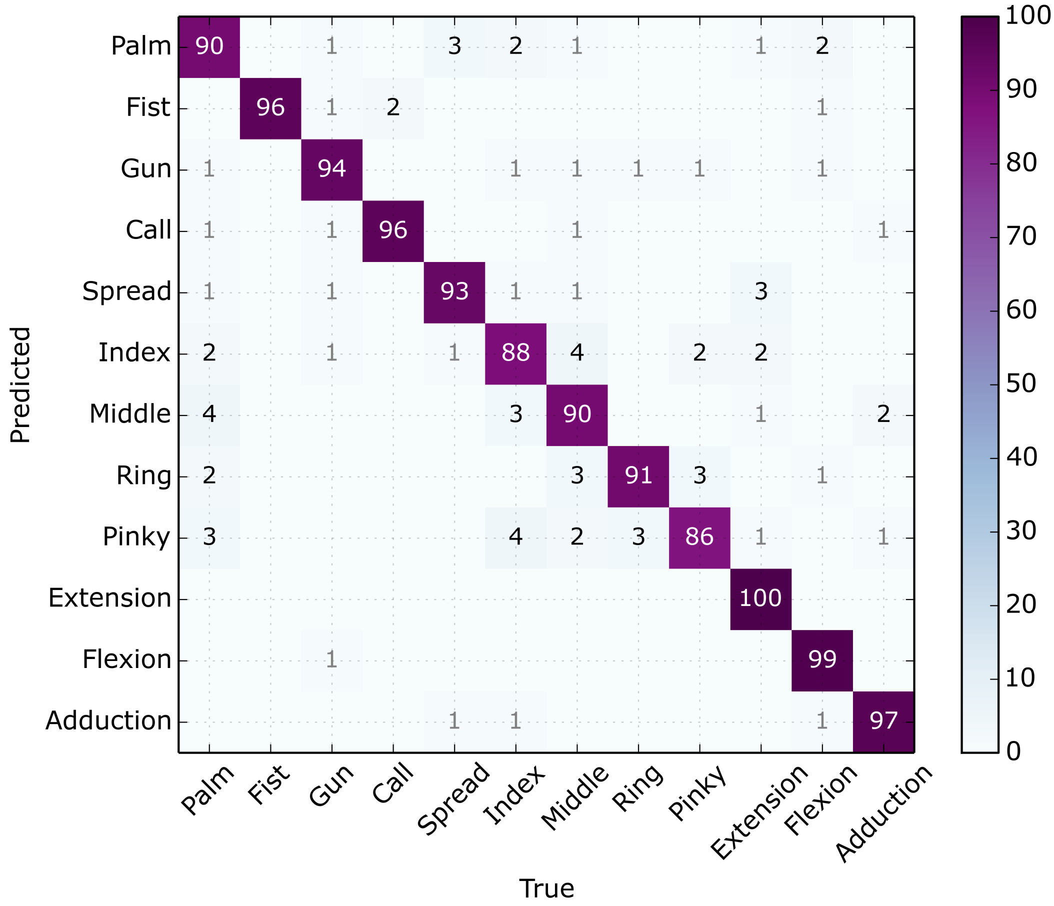

Figure 6.6: Confusion matrix for all gesture classifications accumulated in the cross-validation.

t(9)=3.002, p=0.015 for 7-segments, and not significant 51.6%(SD=15.6) <> 40.6%(SD=24.5), t(9)=0.948, p=0.368 for 4-segments. Although the 14-segments configuration obtained the best results, the Smartwatch and 7-segments arrangements still provide good accuracy. This could enable integrating SensIR into the strap of existing wearables or to reduce the cost and power usage with 7-segments. For all the arrangements, using all the features provided significantly better results than using only reflection, except for the 4-segments configuration in which both accuracies were not adequate for usage in a real system.

6.4.2 Secondary Studies

These additional preliminary investigations into practicality issues that would be encountered during real scenarios. For these studies, we used data from a single participant only (Male, aged 25).

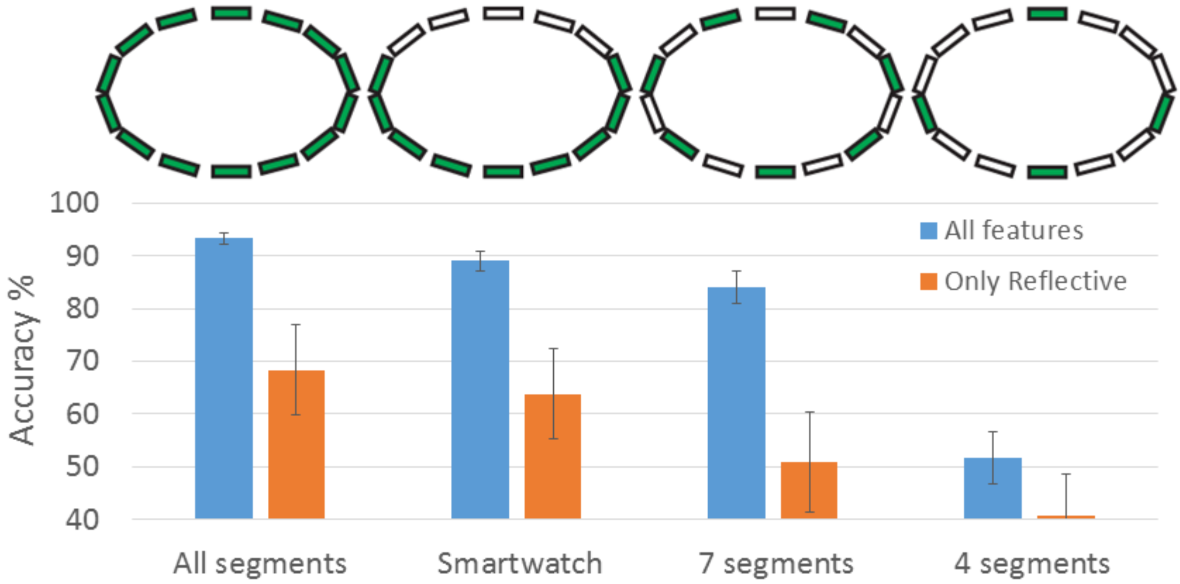

Figure 6.7: Accuracy in gesture recognition for different arrangements. For each, the accuracy is split into a system that uses all the features (like SensIR) and a system that only uses reflective measures (like previous systems). Error bars represent standard deviation.

6.4.2.1 Non-Sedentary Study

Procedure

In a real scenario, the users are likely to move their arm around. As an initial investigation into the performance reduction due to arm movements, we conducted a study which purposefully introduced arm elevations and rotations as part of the classifier training procedure. We tested for 3 different arm elevations: arm pointed towards the floor, at 45 degrees, and perpendicular to the floor.

Results

The classifier was still able to detect the same 12 gestures with 86.1% accuracy. We used 4 training rounds at each of the 3 arm positions, using a 12-fold cross validation method. For 3 different forearm rotations (palms facing upwards, facing inwards and facing downwards), the accuracy drops to 79.2%. The forearm rotation causes greater classification error likely due to stronger morphological changes inside the wrist.

6.4.2.2 Calibration

Procedure

Small sensors misalignment happen while the bracelet is worn or when the user takes off the wristband and puts it back again, this shifts the features causing errors in the classifier. We included a study to determine if the open palm position can be used to calibrate the orientation of

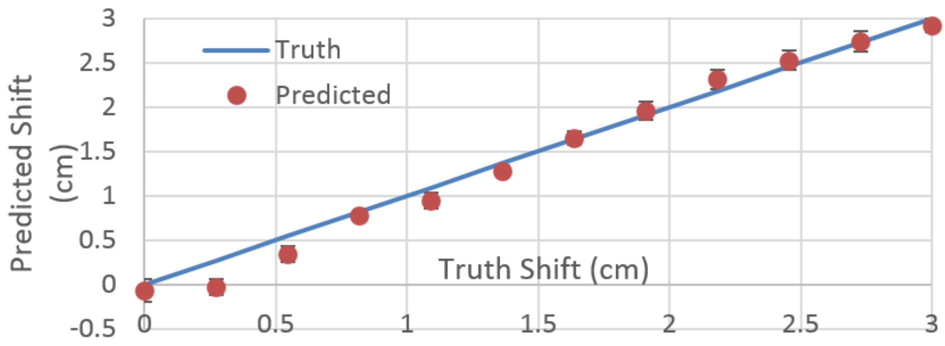

Figure 6.8: A plot of the averages of the orientation shift estimation.

the device. We gathered data from 12 different placements rotating the bracelet on the wrist with displacements of 2.5mm over a range of 3cm.

Results

We used a neural network regressor to estimate the orientation, giving an NRMSD of 0.186 (Fig. 6.8). This suggests that it is possible to determine the orientation of the bracelet and thus correct for small shifts in the sensors alignment.

6.4.2.3 Coupling

Procedure

Infrared systems are inherently more resistant to bad coupling than EMG, pressure, impedance or ultrasonic methods, since neither emitting or receiving components require direct contact with the skin. To test this, we lifted the entire array off the skin in numerous ways: by using rubber bands between the skin and the array and wearing the bracelet on top of thick, clear latex.

Results

Neither of these additions to the experimental procedure seemed to affect the accuracy, both still achieving an accuracy of above 90%.

6.5 Discussion

Our results indicate that using all the combinations of emitters and receivers outperforms previous configurations where only one measurement is taken per emitter/receiver pair. The additional data provided by this new configuration provides extra reflective measurements when the emitter/receiver pair are close, and transmission measurements which indicate orientation and distances between distant emitters and receivers.

SensIR seems comparable in accuracy and gesture sets to EMG or electrical impedance tomography with the added advantage of robustness against bad coupling. However, there are many factors such as the machine learning algorithm, number of sensors and study differences that affect the classification outcome. We only claim that SensIR is an improvement over previous

IR methods. However, more research should still be conducted with this approach such as cross-session and cross-participant studies.

Good cross-session performance is an important requirement often ignored. As found in similar techniques, the main difficulty is sensor misalignment between different sessions. An algorithm to detect and correct placement shift would improve this significantly. Given the accuracy of the calibration shown in the last study, it would be feasible to estimate the orientation of the device and rotate the measurements accordingly.

Commonly in gesture recognition studies, the effects of arm movements that occur in non-sedentary scenarios are ignored. We have shown that the system is still capable of recognising a high number of gestures under conditions that might occur in real situations, by including such data during training. Further work is required here to make the system more robust to these motion artefacts, especially regarding arm rotations.

Even a tightly fitted band could be insufficient to prevent interference's from powerful sources of infrared like the sun. A solution could be to take a measurement of the receivers without emitting, and then use the differential to cancel out background IR.

6.6 Conclusion

We have presented a novel technique for detecting hand gestures using a wearable bracelet, designed to be integrated with wrist form-factor devices. The bracelet is composed of 14 segments, each of them able to emit and measure infrared light. Our user study has shown that measuring with all the possible combinations of emitter/receiver is superior to previous IR systems that capture only reflections. Additional preliminary studies highlight areas for future improvement in robustness and coupling, with suggested solutions for calibration. We anticipate that this work will stimulate more research using SensIR since it leads to a significant increase in accuracy without additional hardware.

6.7 Reflection and final adjustments to the requirements

In this chapter, we've demonstrated that the accuracy and range of infrared gesture recognition devices are close to EMG. In this work we achieved 93.3% accuracy for 12 gestures against the 86.6% for 15 gestures of EMG (note: only EMG, not combined EMG and FSR, from chapter 4). Although the narrative of the paper "SensIR" had more emphasis on the accuracy improvement due to the novelty of transmission and reflection, the greater implications of the work - at least in the context of this thesis - are in the small findings towards the latter parts of the paper. As we originally set out to investigate, we did find that the infrared sensing mechanism was largely undisturbed by a loosening of the bracelet. This was tested by introducing a 1mm gap between the skin and the bracelet, and also tested by putting a variety of clear materials underneath the bracelet. Although we cannot prove with certainty that this is more comfortable without a user

Technique	Mobility	Placement & Size	Fidelity	Comfort	Complexity
EMG	Average	Wrist	Good	Poor	Simple
Ultrasound	Poor	Wrist	Very good	Poor	Very complex
Infrared	Highly	Thin & wrist	**Good**	**Good**	Very simple
Pressure	Highly	Thin & wrist	Poor	Poor	Very simple
EIT	Highly	Thin & wrist	Average	Poor	Fairly complex
Optical	Average	Wrist (but obtrusive)	Limited	Good	Fairly complex
Glove	Highly	Obtrusive	Excellent	Good	Simple

Table 6.1: An updated table of analysis for gesture recognition techniques, with respect to the requirements.

study to acquire feedback on the comfort of the device, it is reasonable to assume that a device that is less tight than electrode-type devices (which leave imprints on the skin) is preferable. This has great implications for where the technology sits from a practicality point of view. We will explain these implications in more detail in the final discussion chapter. For now, we will update the table of analysis to include our findings about infra-red regarding it's increased accuracy (from "poor" to "good") and better comfort due to our findings. As per usual, the changes are marked in bold:

We also investigated device robustness in these last two chapters. Robustness was tested here in two ways. The first was to test sensor shifting, which happens primarily between sessions (as the device is placed slightly differently each time it is worn), but shifting of the sensor position can also happen during a session. The second way is testing against natural movement, which puts the body into different postures that may affect the readings. As tested with the infrared prototype, different arm positions indeed lowers the overall accuracy of system. The robustness of the device is also a factor to consider in the requirements. However, the required robustness is likely to change depending on the application. For example, if the device is to be used during sports where the body moves a lot, of course then robustness is of importance here. But for all applications, a certain level of robustness is required, as none of the envisioned applications assume that the body is sedentary. Also, for any application, cross-session shifts will be a problem. Therefore, although we haven't accounted for robustness into our analysis yet, this should be a requirement to be considered in future work. Our final list of requirements are updated here accordingly:

- Mobility → Wearable.
- Placement & Size → Ergonomics, practicality and social acceptability.
- Accuracy & Range of Gestures & Continuous and Weak → Usefulness.
- Continuous control or subtly performed gestures.
- Device tightness → Comfort.

- **Robustness to movement and cross-session.**
- Complexity → Immediate feasibility.

In the next chapter, we will discuss more about the implications of our final table of analysis and requirements. Following this, we will give important future work for each of the methods described within this thesis, with a particular interest to their practical weaknesses that are highlighted in the table of analysis.

Chapter 7

Discussion

7.1 Final Analysis of Techniques

We now have the final revision of the requirements and table of analysis. This is by no means exhaustive, but takes into account the most prevalent methods and important aspects of the requirements that have been found through experimentation. The table has been re-stated below for ease of viewing. The optical and glove methods are faded because the placement of these devices cannot conform to socially acceptable form factors, as desired by people [96]. If this is not an issue however, then the glove is still by far the best method.

7.1.1 Overall Best Method

By rather simply inspecting the comfort column, there is a clear winner in this category. None of the alternatives have yet overcome the problem of device tightness. In fact, the infrared method performs at least as well as any other method in mobility, complexity, placement and size. The fidelity of this method doesn't quite reach the levels that ultrasound imaging can achieve, but in this regard it is still fairly good for discrete gesture recognition. At present, the most practical method with the highest accuracy is therefore infrared sensing. If comfort is ignored as a requirement here, infrared is still a very good choice, above EMG and EIT all-around. It should still be noted that EMG can support better fidelity, at the expense of a large number of electrodes covering a large surface area. However, such a spread of electrodes goes against the placement and size constraints that make it a practical and unobtrusive device. EIT has also only very recently been applied to gesture recognition. By measuring the internal structures of the forearm, this should be able to attain high levels of accuracy, although this has not yet been demonstrated to the levels of fidelity that infrared or EMG have shown.

Technique	Mobility	Placement & Size	Fidelity	**Comfort**	Complexity
EMG	Average	Wrist	Good	Poor	Simple
Ultrasound	Poor	Wrist	Very good	Poor	Very complex
Infrared	Highly	Thin & wrist	Good	Good	Very simple
Pressure	Highly	Thin & wrist	Poor	Poor	Very simple
EIT	Highly	Thin & wrist	Average	Poor	Fairly complex
Optical	Average	Wrist (but obtrusive)	Limited	Good	Fairly complex
Glove	Highly	Obtrusive	Excellent	Good	Simple

Table 7.1: The final table of analysis for wearable gesture recognition techniques, with respect to the requirements.

One criticism of the infrared method is that we have not tested nor seen literature that shows this method is able to track continuous hand movements or weakly performed gestures. The only technique that is able to support this well enough at the minute, is ultrasound. Ultrasound imaging has the most promising technology for the future, but there are still serious practical issues that need to be fixed as discussed in chapter 5. We will discuss future directions for ultrasound towards the end of this chapter.

7.2 Areas for future improvement

Using this table, we can see the weaknesses of each method with respect to the requirements. For each method, we will briefly talk about the issues that need improvement, in order of importance:

7.2.1 EMG

Comfort As pointed out in the previous chapter, the dry electrodes need to be very tight against the skin, making the device uncomfortable to use for extended periods of time. This fundamental issue is shared across numerous techniques, yet the problem is often disregarded. Material science seems to be the only approach so far that is tackling this problem [24].

Placement & Size The next important aspect is the size of an EMG device. Even though we tested it at the wrist, the prototype was not thin enough to be integrated with the typical widths of watch straps or bracelets. Further research into the minimum inter-electrode distances will be necessary to tell the absolute minimum width of a strap or bracelet.

Mobility EMG still uses quite a bit of power, and the size of the electrodes and circuitry make current prototypes or even off-the-shelf systems fairly large. Advancements in power and circuit design should alleviate some of these problems.

Fidelity Denser electrode designs (which also depends on minimum inter-electrode distances at the wrist) should enable higher numbers of electrodes. This creates more features with which to classify with, which improves the overall classification rate [7].

7.2.2 Ultrasound

Comfort Although not strictly a comfort issue, it is still an issue that is related to coupling between the skin and the sensor. The fundamental issue with ultrasound is with the necessity of the coupling gel. As we found out, partially rigid hydrogels could be a solution to this. However, further research in this area is definitely a priority for enabling wearable ultrasonic imaging devices. Such research is not only beneficial to HCI, but there are also clinical benefits for wearable ultrasound devices. Commercially available is the "ProbeFix" create by Usono, bringing "lengthy and stable fixation of an ultrasound probe to the body. The device facilitates hands-off workflow, continuous measurement and reproducibility of the ultrasound image." [121].

Complexity The other main issue with ultrasound systems is the complexity of the device. The circuits are difficult to design and fabricate, as are the probes themselves. This process makes the device costly, far more than would be reasonable for a technology that is only part of a wearable device. Further engineering to reduce the size and design of better integrated circuitry are expected to continue. These engineering efforts are mostly driven by the recent demands for point of care ultrasound for clinical use [34].

Mobility The devices are large, due to the same reasons as described above for complexity.

7.2.3 Infrared

Fidelity The only part which should be improved on is the accuracy and range of gestures. Currently only supports discrete gestures, whether this can track continuous movements is a big question. A tomographic approach that creates images of the forearm, similar to EIT, should be possible, especially because diffuse optical tomorgraphy is already a well known technique. Perhaps also to be investigated further is the robustness of the device with respect to: natural body movements which change the arm pose; shifting of the device due to it being loose or cross-session shifting; external noise. Otherwise, a solid technology that is good in every other aspect.

7.2.4 Pressure

Fidelity Pressure sensing methods seem poor in fidelity when used by themselves, at least when compared to the state of the art in EMG [7]. Perhaps a higher resolution array of sensors are able to push this boundary higher. As it stands, this technology seems to be a good complementary technique. As we discovered in chapter 4, EMG and pressure are complementary sensing techniques. Adding pressure sensing to electrodes could be a simple way to add an extra dimension of features for other methods such as EMG, EIT, or ultrasound.

Comfort On the other hand, in order for pressure sensing to work well, there has to be some tightness between the sensors and the skin. This is necessary because without any baseline pressure, the sensors have nothing to measure the differences of. This inherent tightness of the pressure sensing technology makes this somewhat impractical for long term usage.

7.2.5 EIT

Comfort Similarly with EMG, the electrode tightness against the skin is a clear issue for comfort. The same improvements in material science for dry EMG electrodes could pave way for a solution here as well.

Fidelity The accuracy and range of gestures with this technique have not yet reached the levels attained by EMG or infrared. However, the data acquired from the deformations within the forearm seem to be detailed. With further improvements to the hardware, imaging algorithm and gesture classification algorithm, we should expect improvements in fidelity.

7.2.6 Optical

Placement & Size As explained in section 3.2, the protrusion of the camera makes the device conspicuous. This is a physical limit with camera based devices. Additionally there are a limit range of gestures that can be captures with this method.

7.2.7 Glove

Placement & Size The glove has obvious issues with conspicuity.

Accuracy & Range of Gestures The glove can replicate many degrees of freedom of the hand. However, there is still not a single implementation that is able to track every degree of freedom of the hand, which remains the ultimate challenge.

7.3 Limitations

We have previously discussed the limitations of individual techniques. Here, we explain the issues that were common to most, if not all of them. This includes some problems that we investigated, such as sensor misalignment between sessions, also those that we did not acknowledge such as user-independent usage.

7.3.1 Robustness

In our previous two chapters, we investigated robustness issues. The robustness of the device in this context implies that there are common factors that affect the measurements of the sensors, which ultimately affects the recognition of gestures. There are robustness issues that are common to almost all technologies, such as calibration of the device placement. Then there are issues that only affect some techniques.

7.3.1.1 Sensor Misalignment Between Sessions

Robustness concerns that are common among all techniques is the placement of the device, which is likely to differ each time the device is worn. This is also known as "cross-session", "between-session" and also "inter-session". For optical and glove type devices, this is not much of a concern since the hand is measured directly, visually or through contact with the appendages respectively. However, for many of the techniques we investigated, a slight displacement of the device can produce a large change in the measured area of the forearm. Using EMG as an example of this, the sensors are placed in areas close to muscles of interest. In current work in this area, classifiers are trained on prior data collected from a specific position. Thus, an offset in position has an affect on the classification of hand poses, due to the sensors now collecting data from different sets of muscles.

This issue can be formulated as a calibration problem. Indeed, in chapter 5, we showed that a calibration to detect the amount of shifting is trivial in the case of ultrasonography; the calibration mechanism being a simple image analysis for transformations. We also demonstrated that with the infra-red system, a change in orientation of the device between sessions can be detected to a certain degree of accuracy, depending on the density of the sensors. However, for EMG this is a more difficult problem, requiring far more complex calibration mechanisms. The latest research in EMG using deep neural networks has tried to improve cross-session gesture recognition accuracy through the use of deep domain adaptation [32].

Calibration due to shifting of the sensors is still to be explored in electrical impedance tomography and pressure sensing techniques.

7.3.1.2 Sensor or Musculature Shifting During a Session

Movement of the device during a session can also occur if the device is worn loosely, or if the user is doing a highly physical activity. This phenomenon is more difficult to test, with so many varieties of human movement that can affect how the device moves. In addition to this, the calibration needs to happen quickly and on-the-fly, if there is a requirement to detect gestures during the activity.

Not only can the sensors shift during a session, but simply by moving the arms position - such as it's elevation - can affect the musculature inside the forearm [61]. This in turn can have an effect on the classification of gestures, especially those that rely on images as is the case with ultrasonography or electrical impedance tomography. We confirmed in our previous chapter, that forearm elevation and rotation have a negative impact on gesture recognition performance. Rotations seemed to have a stronger negative influence, which makes sense intuitively due to the much bigger morphological changes than happen within the forearm.

This is no doubt the more difficult robustness problem of the two. There needs to be research in this area in the future if there is to be progress made towards making these technologies practical in the real world. Inertial Measurement Unit's will be of great utility when developing systems to compensate for the natural movement of the arm, i.e. knowing the arms elevation and rotation.

7.3.2 Anatomical Uniqueness and It's Effect on Cross-Participant Performance

There is no doubt a significant difference in the anatomy across individuals. There are obvious differences such as the size of the muscles and bones, thickness of fat, but there are also missing muscles or tendons [110] amongst certain individuals, which is just an example of just how unique the anatomy can be per individual. This creates challenges in creating a gesture recognition device that works across all kinds of people. Not only is this a challenge for recognition software, but also in designing and fabricating a device that is able to fit anyone. Even in form factors such as gloves, it is not trivial to design a device that can fit different sized hands and maintain similar levels of tracking precision across them. Most gloves with high fidelity tracking often place sensors on or around the joints in the fingers, which of course are shifted in people with differently sized hands.

One possible solution to this problem would be to scan the individuals anatomy as part of an initial setup procedure, and find the closest such anatomy in a database of pre-existing training data. This might broadly categorise traits such as gender, age and body fat percentage. However, this is a challenge in research due to the necessity of needing a large enough data set of participants to test this principle.

Such scanning would be possible using ultrasonography that gives detailed images of the underlying muscular structure. Data taken from sensors can give inferences for similarities between anatomies, but cannot be better than a direct imaging technique such as US.

This is a major challenge that is yet to be tackled in any of the methods that depend on information from forearm musculature.

7.4 Trade-off: Sensor Surface Area vs Accuracy

There is a clear theme that occurs in many of the techniques, where more sensors lead to an increase in performance. This is an expected relationship, nonetheless there is evidence of this in

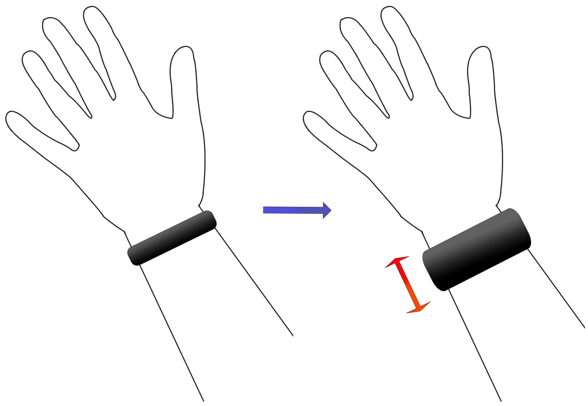

Figure 7.1: A diagram to illustrate the axis in which the device could expand in to improve performance, at the cost of ergonomics.

state-of-the-art high density EMG [7], and also own studies in this thesis support this idea (in chapter 6, testing different arrangements of infrared sensors).

It is far easier to add more sensors in the infrared setup than it is for EMG. This is in part due to the fact that EMG requires a minimum distance between electrodes for effective use [28]. In addition to this, EMG would benefit more from being placed on a different muscle group. This means that more EMG sensors placed within the same surface area benefits less than if it were spread out. This creates challenges in effective EMG design for increasing performance in this way.

Infrared, ultrasound, pressure and EIT are all better in this regard due to the minimum spacing being less of an issue. Especially for ultrasound, the piezoelectric elements are very small, allowing many of them to be packed within a tighter space. It is exciting to think about the opportunities that 3D imaging may bring to hand pose detectiong, as we briefly mentioned in the literature review (chapter 3,section 3.5). This kind of imaging can be done in a stationary manner using densely packed probes. Of course, with this increase in complexity, the feasibility of a mobile and wearable device using this technology decreases even further.

Another example of highly dense ultrasound scanners, are those found in mobile phones for fingerprint detection. These highly dense ultrasound scanners, used in the most recently manufactured phones, scan the contours of the skin of the user. These sensors work at extremely high frequencies in order to image features as small as fingerprints. The high frequency also means that the ultrasound does not penetrate beyond the skin. The sensors have actually been developed as a chip, which is what allows them to be integrated into phones [119]. This is a result of low power usage and low fidelity of the images produced. Perhaps in some future work, there could be a similar chip, but a bit larger in size and works at a slightly lower frequency. Such a chip may enable imaging of superficial tendons or muscles in a small and efficient package.

Fukui et al. already proved that an extremely dense band of infrared sensors can be fabricated with current technology, creating a band of 150 infrared sensors [42]. At the time of writing, this work was published 7 years ago. The physical limits of this are still to be tested with today's technology.

However, if the thinness of the device is not of utmost importance, then performance increase may be created by expanding along the axis of the forearm, as depicted in figure 7.1. Different form factors, such as integration with sleeves of clothing, may benefit from experimentation in this way, since the ergonomy of the device is not compromised.

7.5 Benchmarks for Gesture Recognition and Threshold for Accuracy

Benchmarks have been created for certain techniques, such as CapgMyo, a database of EMG data using a high density array of electrodes [33]. This allows people to use different algorithms with the data to compare to other work. Throughout this thesis, we have similarly compared within each technique using the same gesture set and in some cases, the same data.

But we also had different sets of gestures across techniques. In hindsight, this makes it difficult to compare between different techniques. The reason for the changes of gesture sets is due to experiential findings of participants having difficulty in performing particular gestures during studies. Part of the difficulty in choosing a benchmark for gesture recognition is because of the varying degrees of control that people have with their hands. For some, the individual movement of all digits is easy, but for some, due to a lack of training or anatomy, is very difficult or impossible. It is for this reason that we removed the pinky flexing only gesture in the latter chapters due to the extreme variation in performance. The differences in anatomy is also a factor that must be taken into account when comparing techniques. Even if the same benchmark is used between studies, if one is performed with participants that have good hand dexterity and the other is not, it becomes impossible to draw any meaningful conclusions.

Therefore, as it stands, to have a fair benchmark, it must be done with the same participants. To avoid this, there may be a scheme that allows one to assess the dexterity of the participant. For example, one may imagine a simple test to see whether they are able to make all the gestures precisely enough. The study description of participant information may also include details of the participants body fat percentage and age.

So then what gestures should be chosen? This depends entirely on the envisioned application. For explicit control, some minor set of discrete gestures might suffice. On the other hand, implicit interaction might require recognition of more subtle and difficult to detect gestures. In the end, we should be claiming that one technique is better than another but only in the desired application. Even in this thesis, although we do not make direct comparisons using data, we have made observations that lead us to believe that ultrasound is better for applications that require continuous control.

Technique	Mobility	Placement & Size	Fidelity	Comfort	Complexity
EMG	Average	**Wrist**	**Good**	Poor	Simple
Ultrasound	**Poor**	**Wrist**	**Very good**	**Poor**	**Very complex**
Infrared	Highly	Thin & wrist	**Good**	**Good**	Very simple
Pressure	Highly	Thin & wrist	Poor	Poor	Very simple
EIT	Highly	Thin & wrist	Average	Poor	Fairly complex
Optical	Average	Wrist (but obtrusive)	Limited	Good	Fairly complex
Glove	Highly	Obtrusive	Excellent	Good	Simple

Table 7.2: The final table of analysis for wearable gesture recognition techniques, with respect to the requirements.

Then we must consider what an acceptable accuracy threshold should be. Certainly for a low number of discrete gestures, we would expect recognition rates of 100% to be acceptable. It is not clear whether users would be happy if recognition rates were only 99%. Would the 1 in 100 times when it goes wrong, be so much to make the user avoid using the technology? Certainly in the technology that we use today, we are perhaps not 100% accurate in the way that we use it. This is an open question for gesture research.

As more discrete gestures get introduced, there is also the possibility of the user having difficulty in performing the gestures in the same manner, potentially making the boundary between gestures difficult to carve out. In this case, a 100% recognition rate is probably beyond what is to be expected. It may also be difficult for the user to recognise whether it is a fault of the system, or the fault of the user not performing the gesture in the correct way.

7.6 Contributions

To give a quick overview of the work that has been carried out in this thesis, we will highlight in the table of analysis (Table 7.2) which aspects we have improved for each of the three techniques we worked on.

For EMG, we demonstrated that it works at the wrist position. We also improved the accuracy by adding pressure sensors. We explored ultrasound imaging the most out of the three technologies. We examined the differences in position and orientation, showed excellent performance at discrete and continuous gesture classification at every location including the wrist. We found critical practicality issues that prevents such a device from being realised today. We went further to investigate these issues of sensor coupling and complexity, showing possible approaches to mitigate these issues. For infrared, we vastly improved the accuracy using a novel technological innovation. We then showed that it is possible to use this device without it being air-tight against the skin. In addition to these, we improved cross-session shifts for ultrasound and infrared, and also investigated non-sedentary positions for infrared.

Besides these individual contributions within each technique, we created a holistic view of all current technologies in the literature review. This view particularly criticises the practicality concerns of each technology. We used this to guide us to either fill in the unknown gaps in the literature or to improve aspects such as the accuracy. We hope that this table of analysis can be used by others. Primarily for the field of research, our intentions are to let researchers in this field at least see the practicality issues and at least discuss them in any future work. Even more hopeful, is the idea that researchers are able to use this table to produce important future work that tackles the weaknesses of these techniques, as described above. People who wish to use gesture recognition wearables can also use this table as a guideline for which technology to use, given their requirements. Sadly, we have come to the conclusion that there is not yet a viable method that enables applications in implicit gesture interaction, subtle interaction or continuous control. We also go into detail at the end of the three chapters for future work in each of the methods. This includes sensor fusion, different arrangements of sensors, further measures against issues of robustness. We highlight the difficulties in approaching the robustness concerns that are also critical to the use of this technology in the real world.

Chapter 8

Conclusion

At the beginning of the thesis, we formulated some requirements for satisfying a wearable gesture recogniser. We then analysed existing techniques in the literature review and created a table, which in comparison to our final table, looked very different and uncertain on some aspects. The work conducted in this thesis has now made clear these open questions regarding performance changes due to positional variation. By the end of the thesis, the landscape of gesture recognition techniques has changed tremendously through our findings.

Our contributions to infrared sensing and identification of key practicality issues that have not been mentioned elsewhere in literature, led us to conclude that infrared is currently the best method (and by best, we mean the one that satisfies every criteria in the table to a good level). It also happens to be the only technique that does not suffer from skin coupling tightness. This technique shows great commercial viability, as the complexity of the system is low and easy to fabricate.

We discussed robustness issues that we observed during some of our experiments. These occur either because of movement of the sensors or morphological shifting of the musculature due to arm elevation or rotation. In ultrasonography and infrared, we find methods to mitigate these effects on gesture classification accuracy. These robustness problems, that we clearly identify in the discussion, are an important area to investigate in the future.

Towards the end of the thesis, we also discuss the problem of anatomical uniqueness. Not only does this uniqueness create obvious challenges in fabricating devices per user, but also in classification of gestures across users. A solution to this problem would make user-specific training minimal.

Some methods are only just emerging, e.g. EIT, mechanomyography. In time, it is likely that new and novel methods are to be discovered. It is important to assess these new emerging technologies not just on their accuracy, but also in these other aspects of practicality. We hope that this thesis provides a new perspective for others researching wearable hand gesture detection, and gives guidance and inspiration to those seeking improvements to these methods

BIBLIOGRAPHY

[1] S. AHMAD AND P. CHAPPELL, *Surface emg pattern analysis of the wrist muscles at different speeds of contraction*, Journal of medical engineering & technology, 33 (2009), pp. 376–385.

[2] E. M. AHMED, *Hydrogel: Preparation, characterization, and applications: A review*, Journal of Advanced Research, 6 (2015), pp. 105 – 121.

[3] N. AKHLAGHI, C. A. BAKER, M. LAHLOU, H. ZAFAR, K. G. MURTHY, H. S. RANGWALA, J. KOSECKA, W. M. JOINER, J. J. PANCRAZIO, AND S. SIKDAR, *Real-time classification of hand motions using ultrasound imaging of forearm muscles*, IEEE Transactions on Biomedical Engineering, 63 (2016), pp. 1687–1698.

[4] D. ALIAKSEYEU AND J. MASON, *Tap sensor: Evaluating a new physical user interface for connected lighting*, in Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, UbiComp '16, New York, NY, USA, 2016, ACM, pp. 1614–1619.

[5] D. ALIAKSEYEU, B. MEERBEEK, J. MASON, A. LUCERO, T. OZCELEBI, AND H. PIHLAJANIEMI, *Beyond the switch: Explicit and implicit interaction with light*, in Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational, NordiCHI '14, New York, NY, USA, 2014, ACM, pp. 785–788.

[6] A. ALKAN AND M. GÜNAY, *Identification of emg signals using discriminant analysis and svm classifier*, Expert Syst. Appl., 39 (2012), pp. 44–47.

[7] C. AMMA, T. KRINGS, J. BÖER, AND T. SCHULTZ, *Advancing muscle-computer interfaces with high-density electromyography*, in Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, New York, NY, USA, 2015, ACM, pp. 929–938.

[8] D. ARCHER, *Unspoken diversity: Cultural differences in gestures*, Qualitative Sociology, 20 (1997), pp. 79–105.

[9] H. J. Baek, H. J. Lee, Y. G. Lim, and K. S. Park, *Conductive polymer foam surface improves the performance of a capacitive eeg electrode*, Biomedical Engineering, IEEE Transactions on, 59 (2012), pp. 3422–3431.

[10] G. Bailly, J. Müller, M. Rohs, D. Wigdor, and S. Kratz, *Shoesense: a new perspective on gestural interaction and wearable applications*, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2012, pp. 1239–1248.

[11] A. B. Barreto, S. D. Scargle, and M. Adjouadi, *A practical emg-based human-computer interface for users with motor disabilities*, Journal of rehabilitation research and development, 37 (2000), p. 53.

[12] A. Bhuyan, J. W. Choe, B. C. Lee, I. O. Wygant, A. Nikoozadeh, Ö. Oralkan, and B. T. Khuri-Yakub, *Integrated circuits for volumetric ultrasound imaging with 2-d cmut arrays*, IEEE transactions on biomedical circuits and systems, 7 (2013), pp. 796–804.

[13] R. A. Bolt, *“put-that-there”: Voice and gesture at the graphics interface*, SIGGRAPH Comput. Graph., 14 (1980), pp. 262–270.

[14] G. Bradski et al., *The opencv library*, Doctor Dobbs Journal, 25 (2000), pp. 120–126.

[15] J. Byrnes, *Unexploded ordnance detection and mitigation*, Springer Science & Business Media, 2008.

[16] M. Caon, L. Angelini, O. A. Khaled, D. Aliakseyeu, J. Mason, and E. Mugellini, *Tangible interaction with light in the iot*, in Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, UbiComp '16, New York, NY, USA, 2016, ACM, pp. 1646–1651.

[17] E. A. Cartmill, S. Beilock, and S. Goldin-Meadow, *A word in the hand: action, gesture and mental representation in humans and non-human primates*, Phil. Trans. R. Soc. B, 367 (2012), pp. 129–143.

[18] C. Castellini and G. Passig, *Ultrasound image features of the wrist are linearly related to finger positions*, in Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on, IEEE, 2011, pp. 2108–2114.

[19] C. Castellini, G. Passig, and E. Zarka, *Using ultrasound images of the forearm to predict finger positions*, Neural Systems and Rehabilitation Engineering, IEEE Transactions on, 20 (2012), pp. 788–797.

[20] C.-C. Chang and C.-J. Lin, *LIBSVM: A library for support vector machines*, ACM Transactions on Intelligent Systems and Technology, 2 (2011), pp. 27:1–27:27.

Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[21] X. Chen, X. Zhang, Z.-Y. Zhao, J.-H. Yang, V. Lantz, and K.-Q. Wang, *Hand gesture recognition research based on surface emg sensors and 2d-accelerometers*, in Wearable Computers, 2007 11th IEEE International Symposium on, IEEE, 2007, pp. 11–14.

[22] X. Chen, Y.-P. Zheng, J.-Y. Guo, and J. Shi, *Sonomyography (smg) control for powered prosthetic hand: a study with normal subjects*, Ultrasound in medicine & biology, 36 (2010), pp. 1076–1088.

[23] Y. Chen, X. Su, F. Tian, J. Huang, X. L. Zhang, G. Dai, and H. Wang, *Pactolus: A method for mid-air gesture segmentation within emg*, in Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '16, New York, NY, USA, 2016, ACM, pp. 1760–1765.

[24] Y.-H. Chen, M. O. de Beeck, L. Vanderheyden, E. Carrette, V. Mihajlović, K. Vanstreels, B. Grundlehner, S. Gadeyne, P. Boon, and C. Van Hoof, *Soft, comfortable polymer dry electrodes for high quality ecg and eeg recording*, Sensors, 14 (2014), pp. 23758–23780.

[25] Z. Chen, X. Song, L. Lei, X. Chen, C. Fei, C. T. Chiu, X. Qian, T. Ma, Y. Yang, K. Shung, et al., *3d printing of piezoelectric element for energy focusing and ultrasonic sensing*, Nano Energy, 27 (2016), pp. 78–86.

[26] J. Choi, B.-K. Seo, and J.-I. Park, *Robust hand detection for augmented reality interface*, in Proceedings of the 8th International Conference on Virtual Reality Continuum and Its Applications in Industry, VRCAI '09, New York, NY, USA, 2009, ACM, pp. 319–321.

[27] K. Coleman, *Electromyography based human-computer-interface to induce movement in elderly persons with movement impairments*, in Proceedings of the 2001 EC/NSF Workshop on Universal Accessibility of Ubiquitous Computing: Providing for the Elderly, WUAUC'01, New York, NY, USA, 2001, ACM, pp. 75–79.

[28] C. J. De Luca, M. Kuznetsov, L. D. Gilmore, and S. H. Roy, *Inter-electrode spacing of surface emg sensors: reduction of crosstalk contamination during voluntary contractions*, Journal of biomechanics, 45 (2012), pp. 555–561.

[29] A. Dementyev and J. A. Paradiso, *Wristflex: Low-power gesture input with wrist-worn pressure sensors*, in Proceedings of the 27th annual ACM symposium on User interface software and technology, ACM, 2014, pp. 161–166.

[30] P. DeSaix, *Anatomy and Physiology, Volume 1*, OpenStax College (Rice University), 2013.

[31] R. DRAKE, A. W. VOGL, AND A. W. MITCHELL, *Gray's Anatomy for Students E-Book*, Elsevier Health Sciences, 2009.

[32] Y. DU, W. JIN, W. WEI, Y. HU, AND W. GENG, *Surface emg-based inter-session gesture recognition enhanced by deep domain adaptation*, Sensors, 17 (2017), p. 458.

[33] ——, *Surface EMG-based inter-session gesture recognition enhanced by deep domain adaptation*, Sensors, 17 (2017), p. 458.

[34] G. ELECTRIC, *GE Vscan*, 2017.
`http://www3.gehealthcare.com/en/products/categories/ultrasound/vscan_family`.

[35] I. ELECTRONICS, *Fsr400 short*, 2015.
`https://www.interlinkelectronics.com/fsr-400-short`.

[36] G. ELKOURA AND K. SINGH, *Handrix: Animating the human hand*, in Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, SCA '03, Aire-la-Ville, Switzerland, Switzerland, 2003, Eurographics Association, pp. 110–119.

[37] A. ERANKI, L. BELLINI, L. PROSSER, C. STANLEY, D. BLAND, K. ALTER, D. DAMIANO, AND S. SIKDAR, *Measurement of tendon velocities using vector tissue doppler imaging: A feasibility study*, in 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, Aug 2010, pp. 5310–5313.

[38] C.-S. FAHN AND H. SUN, *Development of a data glove with reducing sensors based on magnetic induction*, IEEE Transactions on Industrial Electronics, 52 (2005), pp. 585–594.

[39] D. FARINA, T. LORRAIN, F. NEGRO, AND N. JIANG, *High-density emg e-textile systems for the control of active prostheses*, in Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE, IEEE, 2010, pp. 3591–3593.

[40] G. FARNEBÄCK, *Two-frame motion estimation based on polynomial expansion*, in Image analysis, Springer, 2003, pp. 363–370.

[41] *Spectra Symbol Flex Sensor 4.5*.
`http://www.spectrasymbol.com/wp-content/themes/spectra/images/datasheets/FlexSensor.pdf`.

[42] R. FUKUI, M. WATANABE, T. GYOTA, M. SHIMOSAKA, AND T. SATO, *Hand shape classification with a wrist contour sensor: development of a prototype device*, in Proceedings of the 13th international conference on Ubiquitous computing, ACM, 2011, pp. 311–314.

[43] F. GARZOTTO AND M. VALORIANI, *"don't touch the oven": Motion-based touchless interaction with household appliances*, in Proceedings of the International Working Conference on Advanced Visual Interfaces, AVI '12, New York, NY, USA, 2012, ACM, pp. 721–724.

[44] M. GAZZONI, N. CELADON, D. MASTRAPASQUA, M. PALEARI, V. MARGARIA, AND P. ARIANO, *Quantifying forearm muscle activity during wrist and finger movements by means of multi-channel electromyography*, PloS one, 9 (2014), p. e109943.

[45] W. GENG, Y. DU, W. JIN, W. WEI, Y. HU, AND J. LI, *Gesture recognition by instantaneous surface emg images*, Scientific reports, 6 (2016), p. 36571.

[46] J. GONG, X.-D. YANG, AND P. IRANI, *Wristwhirl: One-handed continuous smartwatch input using wrist gestures*, in Proceedings of the 29th Annual Symposium on User Interface Software and Technology, ACM, 2016, pp. 861–872.

[47] D. S. GONZÁLEZ AND C. CASTELLINI, *A realistic implementation of ultrasound imaging as a human-machine interface for upper-limb amputees*, Frontiers in neurorobotics, 7 (2013).

[48] J. A. GRAHAM AND S. HEYWOOD, *The effects of elimination of hand gestures and of verbal codability on speech performance*, European Journal of Social Psychology, 5 (1975), pp. 189–195.

[49] C. GRIERSON AND J. SCHIEFELBEIN, *The Arabidopsis Book*, American Society of Plant Biologist, 2002.

[50] J. GUMMESON, B. PRIYANTH, AND J. LIU, *An energy harvesting wearable ring platform for gesture input on surfaces*, in Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys '14, New York, NY, USA, 2014, ACM, pp. 162–175.

[51] J.-Y. GUO, Y.-P. ZHENG, Q.-H. HUANG, X. CHEN, AND J.-F. HE, *Comparison of sonomyography and electromyography of forearm muscles in the guided wrist extension*, in Medical Devices and Biosensors, 2008. ISSS-MDBS 2008. 5th International Summer School and Symposium on, IEEE, 2008, pp. 235–238.

[52] J.-Y. GUO, Y.-P. ZHENG, L. P. KENNEY, A. BOWEN, D. HOWARD, AND J. J. CANDERLE, *A comparative evaluation of sonomyography, electromyography, force, and wrist angle in a discrete tracking task*, Ultrasound in medicine & biology, 37 (2011), pp. 884–891.

[53] H. HAMID MUHAMMED AND J. RAGHAVENDRA, *Optomyography (omg): A novel technique for the detection of muscle surface displacement using photoelectric sensors*, in 10th International Conference on Bioelectromagnetism, ISBEM 2015, vol. 10, 2015.

[54] T. HERMANN, T. HENNING, AND H. RITTER, *Gesture desk-an integrated multi-modal gestural workplace for sonification*, in Gesture Workshop, Springer, 2003, pp. 369–379.

[55] N. HETTIARACHCHI, Z. JU, AND H. LIU, *A new wearable ultrasound muscle activity sensing system for dexterous prosthetic control*, in Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on, IEEE, 2015, pp. 1415–1420.

[56] P. HODGES, L. PENGEL, R. HERBERT, AND S. GANDEVIA, *Measurement of muscle contraction with ultrasound imaging*, Muscle & nerve, 27 (2003), pp. 682–692.

[57] M. HOPMANN, P. SALAMIN, N. CHAUVIN, F. VEXO, AND D. THALMANN, *Natural activation for gesture recognition systems*, in CHI '11 Extended Abstracts on Human Factors in Computing Systems, CHI EA '11, New York, NY, USA, 2011, ACM, pp. 173–183.

[58] A. B. HOSTETTER AND M. W. ALIBALI, *Visible embodiment: Gestures as simulated action*, Psychonomic Bulletin & Review, 15 (2008), pp. 495–514.

[59] *HydroAid Ultrasound Hydrogel Pad*.
`http://www.civco.com/mmi/ultrasound/accessories/acoustic-standoffs/Disposable-AquaFlex-Acoustic-Standoff-Pads-610-1322.htm`.

[60] M. A. ISLAM, K. SUNDARAJ, R. B. AHMAD, AND N. U. AHAMED, *Mechanomyogram for muscle function assessment: a review*, PloS one, 8 (2013), p. e58902.

[61] M. JOCHUMSEN, A. WARIS, AND E. N. KAMAVUAKO, *The effect of arm position on classification of hand gestures with intramuscular emg*, Biomedical Signal Processing and Control, (2018).

[62] M. KARAM ET AL., *A taxonomy of gestures in human computer interactions*, (2005).

[63] S. S. KEERTHI AND C.-J. LIN, *Asymptotic behaviors of support vector machines with gaussian kernel*, Neural Comput., 15 (2003), pp. 1667–1689.

[64] F. KERBER, M. LÖCHTEFELD, A. KRÜGER, J. MCINTOSH, C. MCNEILL, AND M. FRASER, *Understanding same-side interactions with wrist-worn devices*, in Proceedings of the 9th Nordic Conference on Human-Computer Interaction, NordiCHI '16, New York, NY, USA, 2016, ACM, pp. 28:1–28:10.

[65] D. KIM, O. HILLIGES, S. IZADI, A. D. BUTLER, J. CHEN, I. OIKONOMIDIS, AND P. OLIVIER, *Digits: freehand 3d interactions anywhere using a wrist-worn gloveless sensor*, in Proceedings of the 25th annual ACM symposium on User interface software and technology, ACM, 2012, pp. 167–176.

[66] J.-H. KIM, N. D. THANG, AND T.-S. KIM, *3-d hand motion tracking and gesture recognition using a data glove*, in Industrial Electronics, 2009. ISIE 2009. IEEE International Symposium on, IEEE, 2009, pp. 1013–1018.

[67] D. K. H. KOHLSDORF AND T. E. STARNER, *Magic summoning: Towards automatic suggesting and testing of gestures with low probability of false positives during use*, J. Mach. Learn. Res., 14 (2013), pp. 209–242.

[68] M. KONO, H. UEKI, AND S.-I. UMEMURA, *Near-infrared finger vein patterns for personal identification*, Applied Optics, 41 (2002), pp. 7429–7436.

[69] P. KONRAD, *The abc of emg*, A practical introduction to kinesiological electromyography, 1 (2005), pp. 30–35.

[70] J.-W. H. KORSTANJE, R. W. SELLES, H. J. STAM, S. E. HOVIUS, AND J. G. BOSCH, *Development and validation of ultrasound speckle tracking to quantify tendon displacement*, Journal of biomechanics, 43 (2010), pp. 1373–1379.

[71] P. KUMAR, J. VERMA, AND S. PRASAD, *Hand data glove: a wearable real-time device for human-computer interaction*, International Journal of Advanced Science and Technology, 43 (2012).

[72] T. LABS, *How do you wear the Myo armband?*
`https://support.getmyo.com/hc/en-us/articles/201169525-How-do-you-wear-the-Myo-armband-.`
Accessed: 2018-06-26.

[73] ——, *Tech specs for the Myo armband.*
`https://www.myo.com/techspecs.`
Accessed: 2018-06-26.

[74] G. LAPUT, R. XIAO, AND C. HARRISON, *Viband: High-fidelity bio-acoustic sensing using commodity smartwatch accelerometers*, in Proceedings of the 29th Annual Symposium on User Interface Software and Technology, ACM, 2016, pp. 321–333.

[75] B. LI, P. HATHAIPONTALUK, AND S. LUO, *Intelligent oven in smart home environment*, in Research Challenges in Computer Science, 2009. ICRCCS'09. International Conference on, IEEE, 2009, pp. 247–250.

[76] J. W. LOCKHART, T. PULICKAL, AND G. M. WEISS, *Applications of mobile activity recognition*, in Proceedings of the 2012 ACM Conference on Ubiquitous Computing, ACM, 2012, pp. 1054–1058.

[77] I. D. LORAM, C. N. MAGANARIS, AND M. LAKIE, *Use of ultrasound to make noninvasive in vivo measurement of continuous changes in human muscle contractile length*, Journal of applied physiology, 100 (2006), pp. 1311–1323.

[78] S. MANN, *Wearable computing: A first step toward personal imaging*, Computer, 30 (1997), pp. 25–32.

[79] J. MCINTOSH AND M. FRASER, *Improving the feasibility of ultrasonic hand tracking wearables*, in Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces, ISS '17, New York, NY, USA, 2017, ACM, pp. 342–347.

[80] J. MCINTOSH, A. MARZO, AND M. FRASER, *Sensir: Detecting hand gestures with a wearable bracelet using infrared transmission and reflection*, in Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, ACM, 2017, pp. 593–597.

[81] J. MCINTOSH, A. MARZO, M. FRASER, AND C. PHILLIPS, *Echoflex: Hand gesture recognition using ultrasound imaging*, in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17, New York, NY, USA, 2017, ACM, pp. 1923–1934.

[82] J. MCINTOSH, C. MCNEILL, M. FRASER, F. KERBER, M. LÖCHTEFELD, AND A. KRÜGER, *Empress: Practical hand gesture classification with wrist-mounted emg and pressure sensing*, in Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, ACM, 2016, pp. 2332–2342.

[83] S. MICHAU, P. MAUCHAMP, AND R. DUFAIT, *Piezocomposite 30mhz linear array for medical imaging: design challenges and performances evaluation of a 128 elements array*, in Ultrasonics Symposium, 2004 IEEE, vol. 2, IEEE, 2004, pp. 898–901.

[84] Z. MO AND U. NEUMANN, *Real-time hand pose recognition using low-resolution depth images*, in Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, vol. 2, IEEE, 2006, pp. 1499–1505.

[85] I. MOON, M. LEE, J. CHU, AND M. MUN, *Wearable emg-based hci for electric-powered wheelchair users with motor disabilities*, in Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on, IEEE, 2005, pp. 2649–2654.

[86] C. L. MOORE AND J. A. COPEL, *Point-of-care ultrasonography*, New England Journal of Medicine, 364 (2011), pp. 749–757.
PMID: 21345104.

[87] A. MUJIBIYA, X. CAO, D. S. TAN, D. MORRIS, S. N. PATEL, AND J. REKIMOTO, *The sound of touch: On-body touch and gesture sensing based on transdermal ultrasound propagation*, in Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces, ITS '13, New York, NY, USA, 2013, ACM, pp. 189–198.

[88] A. C. MYERS, H. HUANG, AND Y. ZHU, *Wearable silver nanowire dry electrodes for electrophysiological sensing*, RSC Advances, 5 (2015), pp. 11627–11632.

[89] A. NG, J. WILLIAMSON, AND S. BREWSTER, *The effects of encumbrance and mobility on touch-based gesture interactions for mobile phones*, in Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '15, New York, NY, USA, 2015, ACM, pp. 536–546.

[90] J. G. NGEO, T. TAMEI, AND T. SHIBATA, *Continuous and simultaneous estimation of finger kinematics using inputs from an emg-to-muscle activation model*, Journal of neuroengineering and rehabilitation, 11 (2014), p. 122.

[91] NOLIAC, *Noliac Piezoelectric discs*, 2017.
`http://www.noliac.com/products/components/disc/`.

[92] M. OGATA AND M. IMAI, *Skinwatch: skin gesture interaction for smart watch*, in Proceedings of the 6th Augmented Human International Conference, ACM, 2015, pp. 21–24.

[93] C. ORIZIO AND M. GOBBO, *Mechanomyography*, Wiley Encyclopedia of Biomedical Engineering, (2006).

[94] V. ORTENZI, S. TARANTINO, C. CASTELLINI, AND C. CIPRIANI, *Ultrasound imaging for hand prosthesis control: a comparative study of features and classification methods*, in Rehabilitation Robotics (ICORR), 2015 IEEE International Conference on, IEEE, 2015, pp. 1–6.

[95] D. PARK, Y. S. LEE, S. SONG, I. RHIU, S. KWON, Y. AN, AND M. H. YUN, *User centered gesture development for smart lighting*, in Proceedings of HCI Korea, HCIK '16, South Korea, 2016, Hanbit Media, Inc., pp. 146–150.

[96] M. PATEMAN, D. HARRISON, P. MARSHALL, AND M. E. CECCHINATO, *The role of aesthetics and design: Wearables in situ*, in Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, ACM, 2018, p. LBW518.

[97] F. PEDREGOSA, G. VAROQUAUX, A. GRAMFORT, V. MICHEL, B. THIRION, O. GRISEL, M. BLONDEL, P. PRETTENHOFER, R. WEISS, V. DUBOURG, ET AL., *Scikit-learn: Machine learning in python*, The Journal of Machine Learning Research, 12 (2011), pp. 2825–2830.

[98] A. PHINYOMARK, P. PHUKPATTARANONT, AND C. LIMSAKUL, *Feature reduction and selection for {EMG} signal classification*, Expert Systems with Applications, 39 (2012), pp. 7420 – 7431.

[99] H. POHL AND R. MURRAY-SMITH, *Focused and casual interactions: Allowing users to vary their level of engagement*, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13, New York, NY, USA, 2013, ACM, pp. 2223–2232.

[100] A. F. PROKOP, S. VAEZY, M. L. NOBLE, P. J. KACZKOWSKI, R. W. MARTIN, AND L. A. CRUM, *Polyacrylamide gel as an acoustic coupling medium for focused ultrasound therapy*, Ultrasound in medicine and biology, 29 (2003), pp. 1351–1358.

[101] S. S. RAUTARAY AND A. AGRAWAL, *Vision based hand gesture recognition for human computer interaction: a survey*, Artificial Intelligence Review, 43 (2015), pp. 1–54.

[102] A. RIENER, A. FERSCHA, F. BACHMAIR, P. HAGMÜLLER, A. LEMME, D. MUTTENTHALER, D. PÜHRINGER, H. ROGNER, A. TAPPE, AND F. WEGER, *Standardization of the in-car gesture interaction space*, in Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '13, New York, NY, USA, 2013, ACM, pp. 14–21.

[103] A. A. ROSBOROUGH, *Gesture, meaning-making, and embodiment: Second language learning in an elementary classroom*, Journal of Pedagogy, 5 (2014), pp. 227–250.

[104] S. ROSEN AND P. HOWELL, *Signals and systems for speech and hearing*, vol. 29, Brill, 2011.

[105] M. ROSSI, S. BENATTI, E. FARELLA, AND L. BENINI, *Hybrid emg classifier based on hmm and svm for hand gesture recognition in prosthetics*, in Industrial Technology (ICIT), 2015 IEEE International Conference on, IEEE, 2015, pp. 1700–1705.

[106] T. S. SAPONAS, D. S. TAN, D. MORRIS, R. BALAKRISHNAN, J. TURNER, AND J. A. LANDAY, *Enabling always-available input with muscle-computer interfaces*, in Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology, UIST '09, New York, NY, USA, 2009, ACM, pp. 167–176.

[107] J. SCHEIBLE, T. OJALA, AND P. COULTON, *Mobitoss: A novel gesture based interface for creating and sharing mobile multimedia art on large public displays*, in Proceedings of the 16th ACM International Conference on Multimedia, MM '08, New York, NY, USA, 2008, ACM, pp. 957–960.

[108] A. SCHMIDT, *Implicit human computer interaction through context*, Personal technologies, 4 (2000), pp. 191–199.

[109] S. A. SEAH, M. OBRIST, A. ROUDAUT, AND S. SUBRAMANIAN, *Need for touch in human space exploration: towards the design of a morphing haptic glove–exoskin*, in Human-Computer Interaction, Springer, 2015, pp. 18–36.

[110] S. J. SEBASTIN, M. E. PUHAINDRAN, A. Y. T. LIM, I. J. LIM, AND W. H. BEE, *The prevalence of absence of the palmaris longus - a study in a chinese population and a review of the literature*, Journal of Hand Surgery, 30 (2005), pp. 525–527.

[111] T. SEYED, X.-D. YANG, AND D. VOGEL, *Doppio: A reconfigurable dual-face smartwatch for tangible interaction*, in Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, New York, NY, USA, 2016, ACM, pp. 4675–4686.

[112] T. SHARP, C. KESKIN, D. ROBERTSON, J. TAYLOR, J. SHOTTON, D. KIM, C. RHEMANN, I. LEICHTER, A. VINNIKOV, Y. WEI, D. FREEDMAN, P. KOHLI, E. KRUPKA, A. FITZGIBBON, AND S. IZADI, *Accurate, robust, and flexible real-time hand tracking*, in Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15, New York, NY, USA, 2015, ACM, pp. 3633–3642.

[113] J. SHI, S.-X. HU, Z. LIU, J.-Y. GUO, Y.-J. ZHOU, AND Y.-P. ZHENG, *Recognition of finger flexion from ultrasound image with optical flow: A preliminary study*, in Biomedical Engineering and Computer Science (ICBECS), 2010 International Conference on, IEEE, 2010, pp. 1–4.

[114] T. SHIBATA, D. AFERGAN, D. KONG, B. F. YUKSEL, I. S. MACKENZIE, AND R. J. JACOB, *Driftboard: A panning-based text entry technique for ultra-small touchscreens*, in Proceedings of the 29th Annual Symposium on User Interface Software and Technology, UIST '16, New York, NY, USA, 2016, ACM, pp. 575–582.

[115] T. SHIBATA, D. AFERGAN, D. KONG, B. F. YUKSEL, S. MACKENZIE, AND R. J. JACOB, *Text entry for ultra-small touchscreens using a fixed cursor and movable keyboard*, in Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '16, New York, NY, USA, 2016, ACM, pp. 3770–3773.

[116] S. SIKDAR, H. RANGWALA, E. B. EASTLAKE, I. A. HUNT, A. J. NELSON, J. DEVANATHAN, A. SHIN, AND J. J. PANCRAZIO, *Novel method for predicting dexterous individual finger movements by imaging muscle activity using a wearable ultrasonic system*, Neural Systems and Rehabilitation Engineering, IEEE Transactions on, 22 (2014), pp. 69–76.

[117] R. J. SMITH, F. TENORE, D. HUBERDEAU, R. E. CUMMINGS, AND N. V. THAKOR, *Continuous decoding of finger position from surface emg signals for the control of powered prostheses*, in Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE, IEEE, 2008, pp. 197–200.

[118] E. A. SORENSON, R. M. SANNER, R. D. HOWARD, D. L. AKIN, B. LINGO, AND D. CADOGAN, *Development of a power-assisted space suit glove joint*, tech. rep., SAE Technical Paper, 1997.

[119] H. TANG, Y. LU, S. FUNG, J. TSAI, M. DANEMAN, D. HORSLEY, AND B. BOSER, *Pulse-echo ultrasonic fingerprint sensor on a chip*, in Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS), 2015 Transducers-2015 18th International Conference on, IEEE, 2015, pp. 674–677.

[120] A. TECHNOLOGIES, *Muscle sensor v3*, 2015.
`http://www.advancertechnologies.com/p/muscle-sensor-v3.html`.

[121] USONO, *ProbeFix*.
`http://usono.com/probefix/`.
Accessed: 2018-06-28.

[122] R. Y. WANG AND J. POPOVIĆ, *Real-time hand-tracking with a color glove*, in ACM transactions on graphics (TOG), vol. 28, ACM, 2009, p. 63.

[123] H. WEN, J. RAMOS ROJAS, AND A. K. DEY, *Serendipity: Finger gesture recognition using an off-the-shelf smartwatch*, in Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, ACM, 2016, pp. 3847–3851.

[124] J. WOJTCZAK AND P. BONADONNA, *Pocket mobile smartphone system for the point-of-care submandibular ultrasonography*, The American journal of emergency medicine, 31 (2013), pp. 573–577.

[125] K. WOLF, Y. ABDELRAHMAN, D. SCHMID, T. DINGLER, AND A. SCHMIDT, *Effects of camera position and media type on lifelogging images*, in Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia, MUM '15, New York, NY, USA, 2015, ACM, pp. 234–244.

[126] K. WOLF, A. SCHMIDT, A. BEXHETI, AND M. LANGHEINRICH, *Lifelogging: You're wearing a camera?*, IEEE Pervasive Computing, 13 (2014), pp. 8–12.

[127] M. T. WOLF, C. ASSAD, A. STOICA, K. YOU, H. JETHANI, M. T. VERNACCHIA, J. FROMM, AND Y. IWASHITA, *Decoding static and dynamic arm and hand gestures from the jpl biosleeve*, in Aerospace Conference, 2013 IEEE, IEEE, 2013, pp. 1–9.

[128] S. YAMAKAWA AND T. NOJIMA, *A proposal for a mmg-based hand gesture recognition method*, in Adjunct proceedings of the 25th annual ACM symposium on User interface software and technology, ACM, 2012, pp. 89–90.

[129] Y. YANG, H. TIAN, B. YAN, H. SUN, C. WU, Y. SHU, L.-G. WANG, AND T.-L. REN, *A flexible piezoelectric micromachined ultrasound transducer*, RSC Advances, 3 (2013), pp. 24900–24905.

[130] Y. ZHANG AND C. HARRISON, *Tomo: Wearable, low-cost electrical impedance tomography for hand gesture recognition*, in Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology, ACM, 2015, pp. 167–173.

[131] Y. ZHANG, R. XIAO, AND C. HARRISON, *Advancing hand gesture recognition with high resolution electrical impedance tomography*, in Proceedings of the 29th Annual Symposium on User Interface Software and Technology, UIST '16, New York, NY, USA, 2016, ACM, pp. 843–850.

[132] H. ZHAO, F. GAO, Y. TANIKAWA, K. HOMMA, AND Y. YAMADA, *Time-resolved diffuse optical tomographic imaging for the provision of both anatomical and functional information about biological tissue*, Applied optics, 44 (2005), pp. 1905–1916.

[133] Y.-P. ZHENG, M. CHAN, J. SHI, X. CHEN, AND Q.-H. HUANG, *Sonomyography: Monitoring morphological changes of forearm muscles in actions with the feasibility for the control of powered prosthesis*, Medical engineering & physics, 28 (2006), pp. 405–415.