Exploring the Potential of Wrist-Worn Gesture Sensing

by

Qi Feng Liu

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Computer Science

Waterloo, Ontario, Canada, 2018

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

This thesis includes work and analysis carried out in conjunction with Keiko Katsuragawa, and my supervisor, Edward Lank. The contents of this thesis has been adapted, revised, and extended from previous unaccepted conference submissions.

My contribution includes: implementation of the system, carrying out the complete study, performing literature review, performing quantitative and qualitative analysis. I thank my co-authors for their continuous guidance, their help with writing and revising some of the paper content, and for helping with the creation of some of the experimental figures.

Abstract

This thesis aims to explore the potential of wrist-worn gesture sensing. There has been a large amount of work on gesture recognition in the past utilizing different kinds of sensors. However, gesture sets tested across different work were all different, making it hard to compare them. Also, there has not been enough work on understanding what types of gestures are suitable for wrist-worn devices. Our work addresses these two problems and makes two main contributions compared to previous work: the specification of larger gesture sets, which were verified through an elicitation study generated by combining previous work; and an evaluation of the potential of gesture sensing with wrist-worn sensors.

We developed a gesture recognition system, WristRec, which is a low-cost wrist-worn device utilizing bend sensors for gesture recognition. The design of WristRec aims to measure the tendon movement at the wrist while people perform gestures. We conducted a four-part study to verify the validity of the approach and the extent of gestures which can be detected using a wrist-worn system.

During the initial stage, we verified the feasibility of WristRec using the Dynamic Time Warping (DTW) algorithm to perform gesture classification on a group of 5 gestures, the gesture set of the MYO armband. Next, we conducted an elicitation study to understand the trade-offs between hand, wrist, and arm gestures. The study helped us understand the type of gestures which wrist-worn system should be able to recognize. It also served as the base of our gesture set and our evaluation on the gesture sets used in the previous research. To evaluate the overall potential of wrist-worn recognition, we explored the design of hardware to recognize gestures by contrasting an Inertial measurement unit (IMU) only recognizer (the Serendipity system of Wen *et al.*) with our system. We assessed accuracies on a consensus gesture set and on a 27-gesture referent set, both extracted from the result of our elicitation study.

Finally, we discuss the implications of our work both to the comparative evaluation of systems and to the design of enhanced hardware sensing.

Acknowledgements

I would like to take this opportunity to thank all the people who have helped me along the way.

First, I would like to thank my supervisor Edward Lank, who has provided me with the opportunity to experience HCI research and provided me guidance along the way. I would also like to give special thanks to Keiko Katsuragawa, who was continuously providing insights and helping along the way. I also want to thank Hemant Surale for providing me initial guidance to bend sensors and his help in building a prototype of the device. I would like to thank all of my labmates who have given suggestions to my project.

Next, I would like to thank my readers: Jim Wallace and Daniel Vogel. Thank you for your time in reading my thesis and offering suggestions to make it complete.

Last but not least, I would like to thank every participant who participated in my study. This work cannot be done with their help.

Dedication

This is dedicated to the ones I love.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

In Human-Computer Interaction, a large amount of work exists in gestural interaction, including work in free-space gesture [28, 32, 61, 64], surface gesture [64, 65], motion gesture [52], and hand gesture input [20]. Our interest is specifically in free space hand gestures, which can be characterized as hand movements performed by the wrist and fingers (versus movements of the arm [12], of a device [52], or on a surface [65]). Hand gesture interaction is attractive both because of the dexterity of the hand and the expressivity of pointing and gesturing as a communication modality.

While natural and expressive, there exist two primary challenges with gestural interaction. The first one is about gesture design. Gestures are not self-revealing [65, 30], so gestural interfaces typically stress recall over recognition [64], presenting challenges for gesture design and use. To address these challenges, researchers aim to create gesture sets that "make sense" to end-users. To craft these gesture sets, significant effort has been made into the design [44, 53, 59, 65] and execution [19, 21, 28, 52] of *elicitation studies*. An elicitation study is a type of study that collects from end-users a set of gestures, measures agreement, and then leverages the consensus on sensical gestures as input (alongside technical constraints) to the design of gesture sets.

The second challenge for gestural interaction, and, in particular, free space or 3D gestural interaction, is the capture (sensing) and interpretation (recognition) of gestures. In the commercial realm, camera-based systems like the Leap Motion exist [9, 60], but are poorly suited to mobile contexts because the camera-based nature of their device currently limits placement flexibility [60]. Wearables like the MYO Gesture Control Armband [26, 4] use electromyography

(EMG) and a 9-axis inertial measurement unit (IMU) to sense hand gestures, but the Myo supports only five hand gestures: wave-left, wave-right, fist, spread, and double-tap. Many research systems recognize different small sets of finger and hand gestures [3, 27, 28, 49, 54]: for example, a recent system by Zhang et al. [68] recognizes two gesture sets, one of eight gestures and a second of five gestures; and a second recent system by Wen et al. [63] uses only a smartwatch built-in IMU to recognize a set of five repeated-movement gestures. Many richer gestural sensing technologies that have been explored in the literature are expensive to implement [61] and/or require extensive set-up and have limited sensing range [41, 49, 63].

1.2 Thesis Goals

During the development of the work, we address four research questions:

1.2.1 Can we design a low-cost wristband which can replicate the MYO armband's performance?

Our initial interest in this space was to explore "discount gesture sensing," i.e. we wanted to see how easily we could design a sensing system that would replicate the functionality and accuracy of the MYO gesture sensing architecture. As well, we wanted to see whether low-cost, low-power sensors could be incorporated into, the band of a wearable device such as a smartwatch or fitness tracker and still provide robust recognition. Our initial attempt into this space of discount sensing gave rise to a prototype system, WristRec, which uses bend sensors, sensors whose resistance varies based on a deflection, and a miniaturized Arduino board to design a watch-strap-based hand-gesture-sensing system. The initial WristRec prototype required less than $15 worth of electronics components – four inexpensive bend sensors, a low-power Arduino board, and a Bluetooth module – and was embedded into the strap of a smartwatch. The evolution of this prototype is shown in Figure 1.1. In an initial evaluation, we found that WristRec performed on-par with the commercial MYO system on the five-gesture set of the MYO armband.

1.2.2 What kind of finger and hand gestures are natural to use in daily interaction with electronics?

With the success of our system, our subsequent interest was in more realistic hand gesture input. Unfortunately, in this space, we encountered a challenge: what type of finger- and hand-gestures

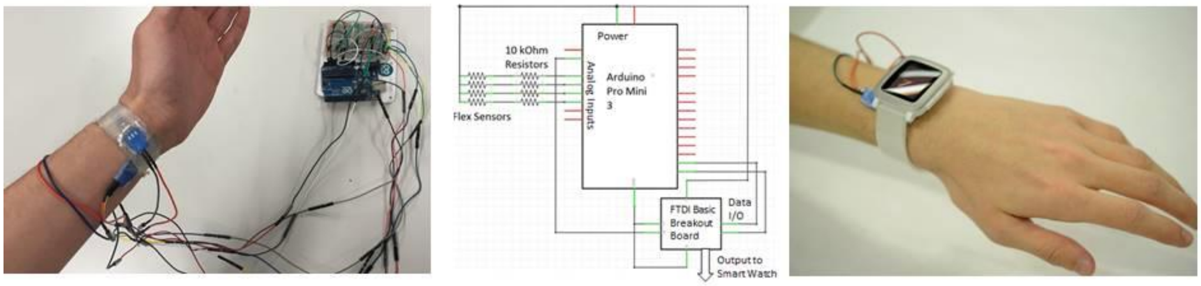

Figure 1.1: Our early design prototype, wiring diagram, and final smart strap prototype.

are natural for integration into a wrist-worn sensor strap? Systems in the literature were typically evaluated on different gesture sets [20, 43, 63, 68], limiting our ability to identify a consensus set of finger- and wrist-gestures appropriate to be considered as a reference set. To address this, our next interest in this work is in furthering our understanding of what wrist- and hand-gestures should be supported. Given this goal, we first present an elicitation study that explores hand-, wrist-, and forearm- manipulations to perform a series of interactions with ubiquitous computing artifacts (with tasks heavily influenced by work by Shimon *et al.*[15] and Ruiz *et al.*[52]). We synthesize the results of our elicitation study, combine the results of elicitation study with past work, and present an aggregated gesture set comprised of 27 gestures which provides full coverage of gestures leveraged in past work [20, 43, 63, 68, 69] and covers over 70% of gestures elicited during our elicitation study, including all gestures that have moderate or high agreement scores.

1.2.3 Is IMU enough to recognize a large set of gestures?

On-going research in the arm, wrist, and hand gesture inputs has expressed a desire for easily wearable recognition systems and, in particular, recognition systems that can be integrated into everyday devices. Many research systems designed around the smartwatch form factor have demonstrated high recognition accuracy on small gesture sets, with set sizes on the order of ten gestures [20, 43, 63, 68, 69]. In particular, the Serendipity system of Wen *et al.* [63] achieves over 80% recognition accuracy using only the Inertial Measurement Unit (IMU) sensors of an off-the-shelf smartwatch. A follow-up question, in light of the success of IMU-only recognition, explores the scalability of the IMU-only recognition and whether or not the IMU-only recognition can be enhanced via the addition of simple wrist-worn sensing technology. We address this question by carefully replicating Wen *et al.*'s Serendipity system, using our wrist movement recognizer, and contrasting IMU-only recognition with augmented recognition.

1.2.4 Is it practical to expect a wrist-worn device to recognize a large gesture set?

Using the 27 gesture set obtained from elicitation study, we also address a follow-on question of practicality. Many of previous systems recognize only a small gesture set, on the order of ten gestures[20, 43, 63, 68, 69]. To determine whether a gesture set comprised of 27 gestures is pragmatic for current wrist-worn configurations, we used the wrist-worn gesture recognizer designed before in combination with a IMU, and demonstrate that our system both performs at near parity with other wrist-worn systems and achieves high recognition rates on our significantly expanded consensus gesture set.

1.3 Contributions

Overall, we identify three contributions in this work.

- Instead of testing with more advanced sensors, we tested whether simple solutions can provide good recognition accuracy. The use of bend sensors showed that non-invasive signal detection integrated into an everyday artifact such as a smartwatch could be combined with an onboard IMU and could be used to sense and recognize more encompassing sets of hand and wrist gestures.
- Our elicitation work both validates the gestures examined in past work and, through elicitation, argues that past gestures from diverse researchers provide a good coverage for gestures gleaned via elicitation.
- We show that using IMU alone is not enough to detect all kinds of gestures. The IMU is good at recognizing gestures which require large movement. To detect gestures like finger-tap or finger-pinch, additional sensors are needed.

1.4 Organization

The thesis is organized as follows:

- Chapter 2 describes previous work on gesture recognition devices. It also gives a description of elicitation studies.
- Chapter 3 describes WristRec, a low-cost device we created for gesture recognition, and a study to verify the device's accuracy.
- Chapter 4 describes an elicitation study we ran to get insights for hand and wrist gestures in different contexts.
- Chapter 5 compares different sensor settings (i.e, IMU only, bend sensor only, and combination of IMU and bend sensors) and gives insights to what types of sensors are needed for a wrist-worn gesture recognition device.
- Chapter 6 discusses the findings from the studies and potential future work.

Chapter 2

Related Work

This chapter discusses previous work on gesture input, focused specifically on work most relevant to the research in this thesis. In the first section, we discuss sensing technologies that have been used in the research community, and which we used or compared our work with. In the second section, we introduce some commercial gesture recognition systems that are currently available. The third section describes research work on gesture recognition devices. The fourth section describes elicitation studies, which are a type of study used to elicit more user-centric ideas about a particular task. Lastly, we talk about the idea of minimalism, which drives much of our design strategy.

2.1 Sensing Technologies Used in Previous Work

2.1.1 Electromyography(EMG)

"Electromyography (EMG) is an electrodiagnostic medicine technique for evaluating and recording the electrical activity produced by skeletal muscles" [6]. In traditional medical usage, EMG is mainly used to analyze the health of muscles by detecting simulated signals passed by nerves [10]. When a person wants to perform an action, a electrical signal is transmitted by the motor neurons which causes the muscle to contract. EMG is a process that can translate these signals into numerical values. Since the signals sent for different gestures may be different, this means that we can potentially recognize gestures by classifying the signals in the muscle.

2.1.2 Inertial Measurement Unit (IMU)

Inertial Measurement Unit (IMU) is a device that can collect information about angular velocity and linear acceleration [2]. It is currently being used in the smartwatch and smartphone to help detect different type of motions.

In a standard Android watch, there are mainly three types of motion data which can be collected: linear acceleration which is measured by an accelerometer sensor; force of gravity which is measured by a gravity sensor; and rate of rotation which is measured by a gyroscope [5]. All these three types of sensors in an IMU would be constantly collecting information at 200HZ in a LG smartwatch which is used in our study. When a user is holding a smartphone or wearing a smartwatch and move his or her hand, those three sensors would sense the corresponding motion and record the corresponding data.

2.2 Commercial Devices

2.2.1 MYO Armband

The MYO armband [26] is a commercial gestural sensing device that uses EMG to reliably sense a small set of wrist- and hand-gestures (see Figure 2.1 for gestures). The armband is worn on the upper arm (typically the dominant arm) and includes a battery and Bluetooth communication to transmit input. MYO armband is one commercial device that can provide gestural control to systems. It can be used to control PowerPoint slides, small flying drones, and even used as VR controller. The MYO armband allows users to perform gestures with their hand and arm. This replaces the traditional button-controller, and give users more options to perform different tasks.

Even though the MYO armband is a great device for interaction, the form factor of the product makes the cost of the device relatively high. To obtain a good accuracy for gesture recognition, the device uses EMG. Since EMG is a more advanced technology which detects electrical signals in muscles, it is driving up the cost of the armband. As of today, MYO armband is currently marketed at $200 USD[4].

Another drawback of MYO armband is that it is a standalone device which people need to wear on their upper arm. Though providing an advantage by keeping away from users' hand, because of the form factor, it is less comfortable if users are wearing it all the time and it is less convenient because users need to carry it around while not using it.

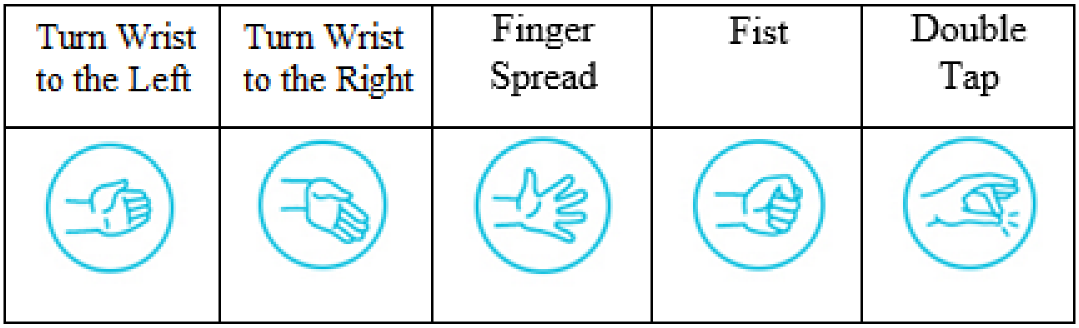

Figure 2.1: Five Myo gestures that can be reliably classified by the MYO armband

2.2.2 Leap Motion

Leap motion [9] is a camera-based solution that can accurately detect hand and finger gestures. It has recently started to provide VR mount so that the device can be combined with VR headsets (HTC Vive and Oculus Rift CV1) to act as an input device for VR.

However, since it is a camera-based solution, the solution suffers from the occlusion problem. Hands must be inside the camera's field of view. Another problem with the camera-based solution is that it is less suitable for mobile contexts because of the limitation of placement flexibility [60].

2.3 Previous Research

2.3.1 Tendon Movement Detection

While performing gestures, different tendons will move and cause visible deformation at the wrist and on the hand. With the current available sensors, is it possible to detect such deformations? If it is, is it possible to classify these different movements? Researchers have tried different ways to detect and classify tendon movements in the past .

BackHand [38] was created by Jhe-Wei *et al.* using 19 strain gauge sensors laid out in two rows at the back of the hand. The system is able to detect 16 gestures chosen from the American Sign Language and popular Asian hand gestures with a 95.8% average accuracy within participants.

Dementyev *et al.* created WristFlex [20], a wrist device consists of an array of force sensitive sensors to measure subtle tendon movements on the wrist. Their system was able to recognize four finger-pinch gestures and spread hand gesture, for a total of 5 gestures, with an accuracy >80%. They also tested recognition accuracy with different hand orientations and was able to achieve a cross-validation accuracy of 94.4%.

2.3.2 Electromyography (EMG)

Alongside commercial systems, researchers have also tried EMG-based sensing solutions [54, 55] for gesture recognition.

In their first work [54], Saponas *et al.* used *BioSemi Active Two* system as their sensing device (www.biosemi.com). They tested pinching gestures (thumb pinches the other 4 fingers, for a total of 4 gestures) under hand-free and hand-full conditions. During the hand-free condition, participants are asked to perform gestures with the hand facing different directions. During the hand-full condition, participants are either holding a coffee mug or pulling up on the handle of a bag. They also tested to use the pinching gestures as a controller for music player. Their system achieved an accuracy of 79% for hand-free condition, 85% while holding a mug, and 88% while carrying a bag.

In their second work [55], Saponas *et al.* created their own device using EMG electrodes with a sports sweatband and a Zigbee wireless radio. They tested two gesture sets. One is pinching index, middle, or ring finger together with their thumb. The second one is pressing down on a surface with one of index, middle, or ring fingers. Participants performed gestures using their right hand and used squeezing left hand as the system activation mechanism. From a eight-person experiment, they demonstrated an average accuracy of 86%.

2.3.3 Combination of Tendon Movement Detection and EMG

In McIntosh *et al.*'s work [43], they combined pressure sensors and EMG sensors to create Empress. Prior to this work, the wrist has not been considered as a good place to attach EMG sensors since there are fewer muscles, and muscles are tightly packed. However, McIntosh *et al.* showed that even though there is less ideal muscle structure on the wrist for EMG, it still can detect fine-grained gestures with high accuracy. However, when they expanded their gesture set to include more coarse gestures, recognition accuracy with EMG starts to drop. Since pressure sensors has been proven to identify gestures [20], they tried to test the performance of EMG and pressure sensors. After testing the two type of sensors separately, they found that pressure sensors are better for wrist and arm movements, while EMG sensors are better for finger movements. In their paper, they came up with a total of 15 gestures as their gesture set. By placing the two types of sensors alongside each other, they were able to detect the gestures with a cross-validation accuracy as high as 96%.

2.3.4 Other Sensors

Other than EMG and tendon movement, various forms of arm- and wrist-worn sensors [16, 27, 33, 36, 49, 63, 68, 69] are also being tested.

Researchers have tried to use camera-based solutions [16, 33, 49] by putting them in places that are less susceptible to the occlusion problem. Bailly *et al.* [16] proposed to put depth sensors on the shoes and point upwards. With this setup, they were able to classify five finger gestures, five arm gestures, and hand movement in free-space with really high accuracy. Digits [33] is another camera-based solution that places cameras on the wrist and tries to detect gestures by building a kinematic hand model for the user. Ren *et al.*[49] attempted to use Kinect camera sensors to detect a group of 10 static hand poses and achieved a mean accuracy of 90.6%.

Researchers are also actively testing different sensors that might be suitable for gesture recognition. Harrison *et al.* designed Skinput [27], an acoustic device placed near the elbow which can detect mechanical vibrations passed through the body. For one hand gestures, they mainly tested tapping and flicking gestures as those gestures would cause vibrations on the arm. Zhang *et al.* proposed Tomo [68, 69], a system that utilizes Electrical Impedance Tomography (EIT) to classify gestures by calculating the ratio of voltage and current between electrodes attached to the arm and wrist. They tested their system on two gesture sets: one consists of four pinch gestures, and the other one consists of seven hand gestures. Recent improvements in ultrasound imaging makes it a possible solution to gesture sensing. Ultrasound is less intrusive compared to gloves, and it is not affected by the occlusion problem which is a common issue in camera-based solutions. By placing an ultrasonographic device on the forearm, and combining image processing and neural networks, McIntosh *et al.* [42] were able to classify a group of 10 gestures, consisting of 8 hand gestures and 2 wrist gestures. Instead of trying to focus on both hand and wrist gestures, Gong *et al.* [24] focused on using wrist as a continuous input modality. By adding proximity sensors, which can "detect the presence of nearby objects" [11], to the wristband of a smartwatch, they were able to detect free-form shapes drawn with wrist with high accuracy.

Other than using other additional sensors, researchers have also tried to utilize sensors that come with everyday electronics. Wen *et al.* [63] used a off-the-shelf smartwatch to detect a group of 5 gestures that requires relatively large hand movement. They classified gestures by combining data from different sensors in the IMU of the smartwatch. Laput *et al.* [36] adopted a similar approach and used a off-the-shelf smartwatch. However, instead of combining sensor data in the IMU, they increased the sampling rate of the IMU to 4k Hz(the original sampling rate of a typical LG watch is 200Hz) and used only the accelerometer data. Their device is able to detect 6 two-handed gestures and 6 one-handed gestures with high accuracies.

2.3.5 Limitations from previous work

Each of these research systems exhibits an ability to achieve a relatively high recognition rate on their provided gesture set during evaluation. However, one thing that we have noted in our analysis of past work is that every system has used a gesture set that, while exhibiting broad similarities, differs in significant ways. For example, some examined finger tapping [36], bending [20] or pinching [54, 68]; others examined static poses [49, 69] versus dynamic movements [36, 63]. Gesture sets vary extensively in size as well: the Serendipity system of Wen et al. [63] was evaluated on a set of 5 gestures, McIntosh et al.'s EMPress system [43] was evaluated on 15 gestures, and Zhang and Harrison [68] evaluated the Tomo system on two different gesture sets – a set of 8 wrist gestures and a set of 5 finger gestures – without reporting unified recognition rates on all 11 gestures (an oversight the authors corrected in follow-on work [69] by combining their 8 and 5 gesture sets into an 11 gesture set). Finally, sometimes, the recognition goals alter the classification problem. For example, Xu et al. designed a system that accepts 39 gestures as input, but the classification problem is to label gestures as either originating with the finger, hand, or arm [66], i.e. a 3-class decision problem, not to recognize the gestures themselves.

Despite on-going interest in wrist-worn gesture capture, we posit that the use of unique gesture sets to evaluate competing technologies presents two drawbacks. First, from the perspective of past systems, it is difficult to analyze the benefits and drawbacks of any one recognizer technology that exists in the literature. If everyone generates over 90% recognition, then why consider, as one example, electrical impedance tomography with all of its complexities [69] when researchers have already obtained 90% recognition using only the IMU on a smartwatch [63] on gesture sets of approximately the same size? Second, and related, from the perspective of new system design and evaluation, we are left suspicious of each new system. Is the new system really necessary? What does the new system add in terms of class of gestures?

One of our goals is an attempt to further the understanding around wrist, hand, and finger gesture input. First, we would like to understand what classes of gestures are needed to be sensed. Alongside this, we investigate the research needed to advance recognition in this field.

2.4 Elicitation Study

Beginning with understanding the desired space of finger, hand, and wrist gesture input, one option is simply to pick one gesture set used in past research, but the question becomes which of the myriad sets of gestures should one select? Should the gestures be dynamic, i.e. require large movement [36, 63]? If so, then this biases against sophisticated techniques that have been shown to detect static pose with high accuracy [20, 69]. Should all gestures be little more than point-and-click operations [32, 39], or are more abstract poses necessary [16, 33, 49, 60, 62]? Perhaps the gold standard for gesture recognition should be various forms of signed letter alphabets [46]? With these questions in mind, it seems to us that there has not been enough study in this area. Therefore, we ran an elicitation study to help with answering these questions.

In the past, software systems are employing gestures created by device designers who are familiar with the device and know the limit of the device. This is good for the system because the gestures would work well for them and achieve the desired result. However, these gestures may not seem intuitive to the general public and it is not necessarily how a general user would perform them. This presents a gap between what designers think users should do and how users would actually do it.

To address issues of heterogeneity in gestural input, elicitation studies [59, 65] are a common technique for understanding potential end-users' concepts of gesture parameters and of how gestural manipulations should map onto commands [19, 52, 60, 65]. Wobrrock et al.[65] are the first to present elicitation study in the field of human-computer interaction to solve the aforementioned problem. In their work for surface computing, such as, tablet, tabletop, instead of coming up with gestures themselves and ask users to validate those gestures, they asked users to come up with gestures and try to form a consensus gesture set.

An elicitation study is a type of participatory design exercise where users are asked to design gestures that "make sense" given a desired system action. In the study, researchers would present participants with a list of tasks and some constraints, such as, the part of body that can be used. For each task, the participants would be asked to come up with two or three gestures which they would be comfortable to perform. By aggregating the gestures from all participants, researchers can come up with a consensus gesture set which is a more user-friendly gesture set. By looking at all the gestures, researchers can also derive a taxonomy of the gestures and some design guidelines, which can be used for future gesture design in a similar task.

While reviewing past work in our problem, we found that one issue with elicitation studies is that there have been many such studies published. The initial question we asked, given the myriad set of studies available, is whether past gesture elicitation studies might provide enough information to us on appropriate gestural interaction. While it is true that many elicitation studies for gestures have been done, our analysis found that these studies either allowed large-scale full arm gestures [53], any body part gestures [21], or strictly limited studies to a very small number of joint movement (i.e., microgestures, using only fingers to perform gestures) which limited the participants' freedom to perform a preferred gesture [19]. Each of these elicitation studies addresses a class of gestures, but there is some evidence that these studies do not fully address the space of gestures that should be considered for wrist-worn sensors. Specifically, Chan et al. [19], while examining microgestures, found that participants showed a desire to use larger arm movements (e.g. hand and wrist) and spatial recognition. Overall, we believe that the class of gestures sensed by wrist-worn recognizers [20, 43, 63, 68, 69] has not been elicited in a way that can be contrasted directly with gesture sets used in past wrist-worn gesture systems.

2.5 Minimalism

Our goal of minimalism may, at first, seem counter-intuitive, but human-computer interaction researchers have a long history of lauding minimalism in design and prototyping. As one example of this is Greenberg and Fitchett's *Phidgets* prototyping tools [25], which were designed to simplify hardware hacking. More recently in the recognition space, systems like the $1 recognizer [65] were designed specifically as a simplification of more complex forms of elastic matching, and researchers have built upon these systems to support, for example, arbitrary orientations [37], multi-stroke input [14], and 3D input [35]. Each of these systems was designed to simplify aspects of deployment and lower costs. The $1 recognizer paper notes this explicitly: the goal was not to beat Tappert's work [56] on elastic matching for handwriting recognition or any of the enhancements to the algorithm to recognize more complex and subtle differences between graphical objects (e.g. [13, 17, 45], but was, instead, compared to other simple unistroke recognizers [51] as a useful tool for prototyping because of its low-cost. Finally, within this domain of minimalism, Wen et al. [63] specifically note the disadvantages of external hardware for gesture sensing because the external hardware drives up cost and reduces wearability.

Chapter 3

WristRec

This chapters introduces a wrist-worn gesture recognition device, WristRec, we created with bend sensors and an Arduino UNO board. We ran an evaluation study to determine the optimal placement for bend sensors, and a validation study with 5 gestures to verify the feasibility of the device by comparing the accuracy with the MYO armband.

3.1 Bend Sensors

As our interest is in low-cost and low-power systems, we identified bend sensors as a cost-effective input sensor. Bend sensors are a type of resistor that changes resistance based on how much it is bent (Figure 3.1). We are not the first to study bend sensors as a gesture recognition mechanism. For example, Tidwell *et al.* and Lin *et al.* [38, 58] attached sensors to fingers/hand and used these sensors to recognize hand gestures. However, attaching sensors to hands decreases the usability of their configuration for daily wear. Other researchers have explored more complex configurations [20, 43] of sensors in wrist-worn configurations. Our goal of a minimally intrusive set of sensors and minimalism distinguish us from these previous work. The bend sensors we used are from Flexpoint Sensor System Inc [8].

3.2 A Simple Wrist-Worn Gesture Recognizer

Our initial wrist-worn sensor system was designed as an exercise to discern how easily we could replicate the functionality of a commercial gestural recognizer such as the MYO armband. Our

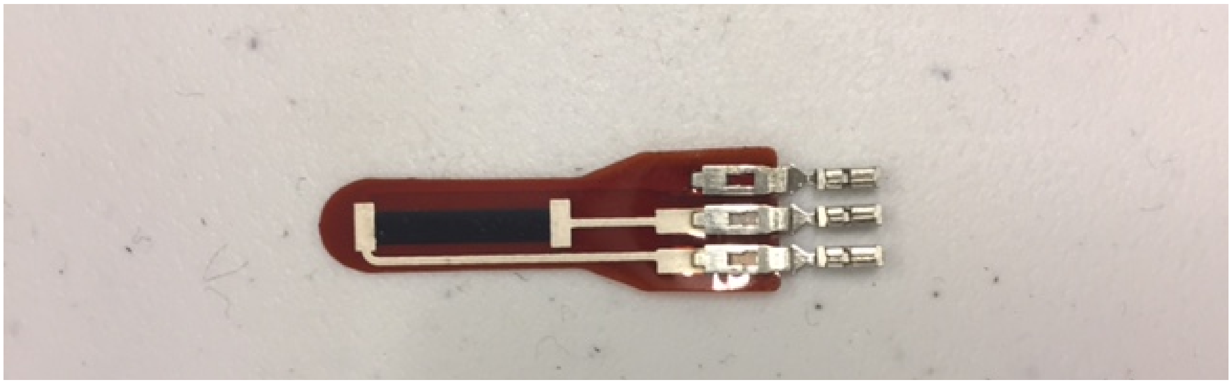

Figure 3.1: A 1-inch Bend Sensor

goal was to build a wrist-worn recognizer or the kind that could be embedded into a smartwatch's strap. To accomplish this, we first conducted a hardware evaluation study to determine whether and how many simple wrist-worn sensors could sense hand and finger movement. We then built and implemented the prototype system shown in Figure 1.1. Finally, we evaluated the system against the MYO on the MYO's gesture set.

3.2.1 Hardware Evaluation

The design of WristRec was driven by a hardware evaluation study that identified a minimal set of sensors needed to support gesture recognition. We used the MYO gesture set, depicted in Figure 2.1, for the initial hardware evaluation. To find out the minimum number of sensors and their optimal placement, we conducted a pilot study. We recruited 5 participants (4 male and 1 female; all right-handed). For each participant, we used a Fitbit sleep band (as shown in Figure 3.2) to measure the length of each participant's left (non-dominant) wrist, which is the wrist most people wear their watch on. We then marked the band with 24-equal-distance marks as test locations for bend sensors.

For testing, we glued one axial and one tangential bend sensor to the sleep band (Figure 3.3). We then rotated the sleep band around each participant's wrist so that all 48 positions (24 locations for two orientation) were tested. Participants performed the 5 gestures of the MYO gesture set 5 times for a total of 5 x 5 = 25 gestures. Each gesture was performed in a 1.5 seconds interval. The participants were given two sound cues: one to indicate the start time to perform the gesture and one to indicate the end time to stop the gesture.

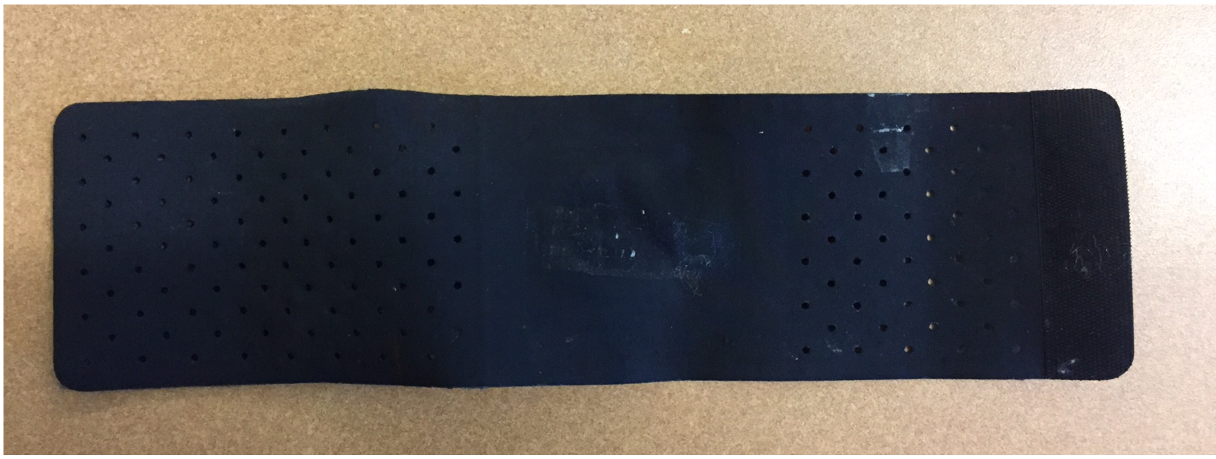

Figure 3.2: Fitbit Sleep band

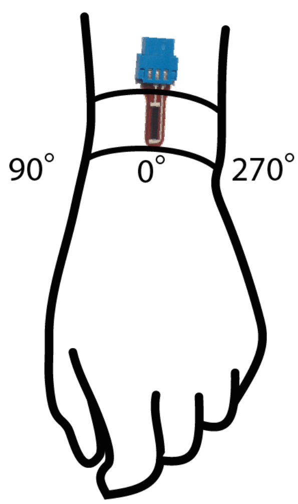

Figure 3.3: Sensor orientations at 0 offset. On the left, axial orientation; in the middle, tangential orientation; and on the right, tilt orientation.

At the end of the pilot study, we analyzed the bend sensor data from all 48 positions for all participants. By looking at which positions would result in the largest sensor reading difference compared to the resting position, we identified three possible combinations, each with 4 sensors, which seemed the most promising locations to support recognition. Considering the top of the wrist (center of the back of the hand) as 0°, these combinations were:

1. 4 axial sensors at 0°, 90°, 180°, and 270°(4 Axial);
2. 4 tangential sensors at 0°, 90°, 180°, and 270°(4 Tangent);
3. 2 axial at 0°and 180°and 2 tangential at 90°and 270°(2 Axial and 2 Tangent).

Given that there seemed to be a benefit to both axial and tangential orientations, we hypothesized that a 45°offset between axial and tangential might also be effective for sensing, so we added a fourth option which is positioning four sensors at a 45°offset, a tilted orientation (4-Tilt). Figure 3.3 depicts a bend sensor positioned at 0°with, from left to right, axial, tangential, and titled alignment.

Consider the third item in the list above, two axial and two tangential sensors located as indicated (0°/180°for axial). Figure 3.4 depicts example sensor data for this arrangement, and it is an example of the data used to drive our sensor placement. Tangential sensor readings are depicted at the top and axial sensor readings are located at the bottom. Note that axial sensor readings are particularly significant for left and right gestures, and that tangential sensors seem to work well for spread, fist, and double tap in this configuration.

3.2.2 Hardware Design

We envision that the most feasible mechanism for ubiquitous hand gesture sensing is the smart strap of a smartwatch or fitness tracker. Figure 1.1 on page 2 shows our prototype WristRec. In Figure 1.1 left, our initial prototype WristRec includes four bend sensors positioned around the wrist. The bend sensors are held in four 3D-printed sensor holders and a jute string fed through small holes allows resizing to accommodate wrists of different sizes. The bend sensors are sending data to a Arduino UNO board which in turn sends the data to the computer through a USB connection. The Arduino board is collected data at 100Hz. We used this prototype in our evaluation of recognition accuracy. Figure 1.1 center depicts the hardware configuration of an untethered and more elegant prototype. Finally, Figure 1.1 right depicts the current WristRec prototype. We envision that each smartwatch strap will be sized such that sensors are positioned at the correct locations around the wrist of a user, i.e. each WristRec is custom-fit.

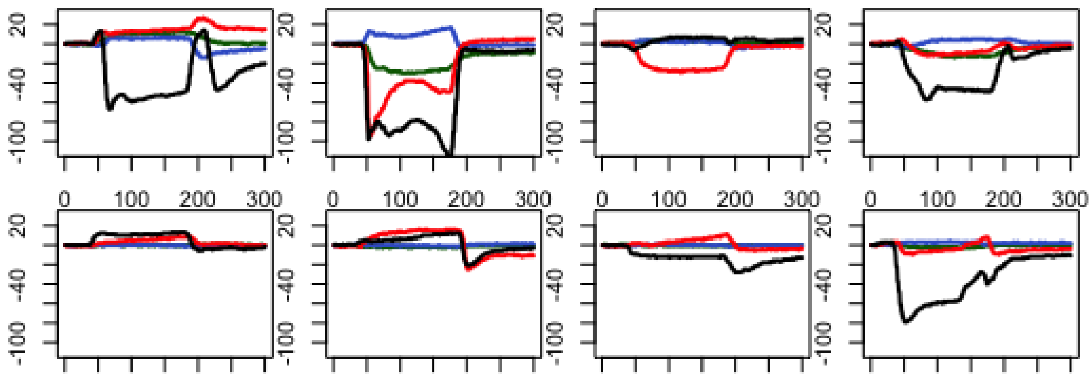

Figure 3.4: Bend sensor data for one instance of one participant's logged performance of each of the gestures. Each color represents data from one of the four sensors. From top left, Flex, Extend, Spread and Fist. From bottom left, Index, Middle, Ring and Pinky finger bend gestures.

3.2.3 Software

All programs are written in Processing 2.2.1 with the built-in library "Arduino (Firmata)". For classifying gestures, we used a simple Dynamic Time Warping (DTW) algorithm [57]. DTW is an algorithm used to calculate the difference between two temporal sequence data which may vary in speed. This fits our data very well because each gesture is performed in a fixed time period. However, the speed to perform the same gesture may be different from time to time. Using DTW, we can determine whether two gestures are the same gesture by finding whether they have the same pattern in their time series data.

The system consists of a calibration stage and a classification stage for each user.

During calibration, each gesture was performed 5 times. The system calculates the warp distance between each candidate gesture in the 5-gesture set and uses the most representative sample as the template for the gesture, i.e. the template that is at minimum average distance from the other four candidate templates.

During classification, the system first standardizes the values for all sensors and also passes the values through a low pass filter. Then, it runs the DTW algorithm to determine the model gesture best matching the current time series data.

As shown in Figure 3.4, when performing gestures, there is a significant spike in sensor data. These spikes are typically absent from sensor input when a user's hand is at rest, i.e. they are highly indicative that the user is performing a gesture.

3.2.4 Evaluation

We performed a two-session experiment over two days to evaluate our WristRec sensor against the MYO gesture sensing arm band. On the first day, participants performed gestures using WristRec; on the second day, they performed the same set of gestures using the MYO arm band. We did this because, with the MYO device, a training process is required that would have resulted in an experiment which is too long if done in a single session.

Participants

We recruited 12 participants (9 male, 3 female; 1 left-handed) from the general student body of our institution. All participants ran the study using their non-dominant hand. All participants were remunerated with $10 after the completion of the experiment.

Method

The first block of the study, which took place on day 1, is for evaluating gestural input using the WristRec prototype. As noted earlier, in our pilot testing, we identified four candidate sensor location combination for bend sensors in WristRec (4-Axial, 4-Tangent, 2-Axial/2-Tangent and 4-Tilt), described in Section 3.2.1. We used a within-subject experiment to verify the performance of the four different sensor location combination.

Participants performed gestures based on sound cues. A one minute training session was run to familiarize the participants with the sound cues. Next, for each sensor location combination, the procedure was as follows:

1. The participant was cued to perform gestures shown on a computer display screen and performed gestures to generate a model. We collected a total of 25 training data sets (5 for each gesture) and the order of the gestures was randomized.

2. In the second step, as the participants performed gestures, the program classified the gesture based on the models generated in step 1. Then the recognition result was shown on the computer display. 25 gestures (5 of each gesture type) were performed in total to obtain test data.

3. Repeat 1 and 2 for all 4 sensor settings. The order of the sensor settings was randomized, and all 12 participants were tested with a different ordering.

The second block of the study took place on day 2. We assessed the accuracy of the MYO device on its gesture set. 10 out of the 12 participants participated in this second block of the study (two participants did not return). Participants first calibrated the MYO arm band using MYO's calibration procedure, and performed 25 gestures (5 each of each kind), again in random order.

Results

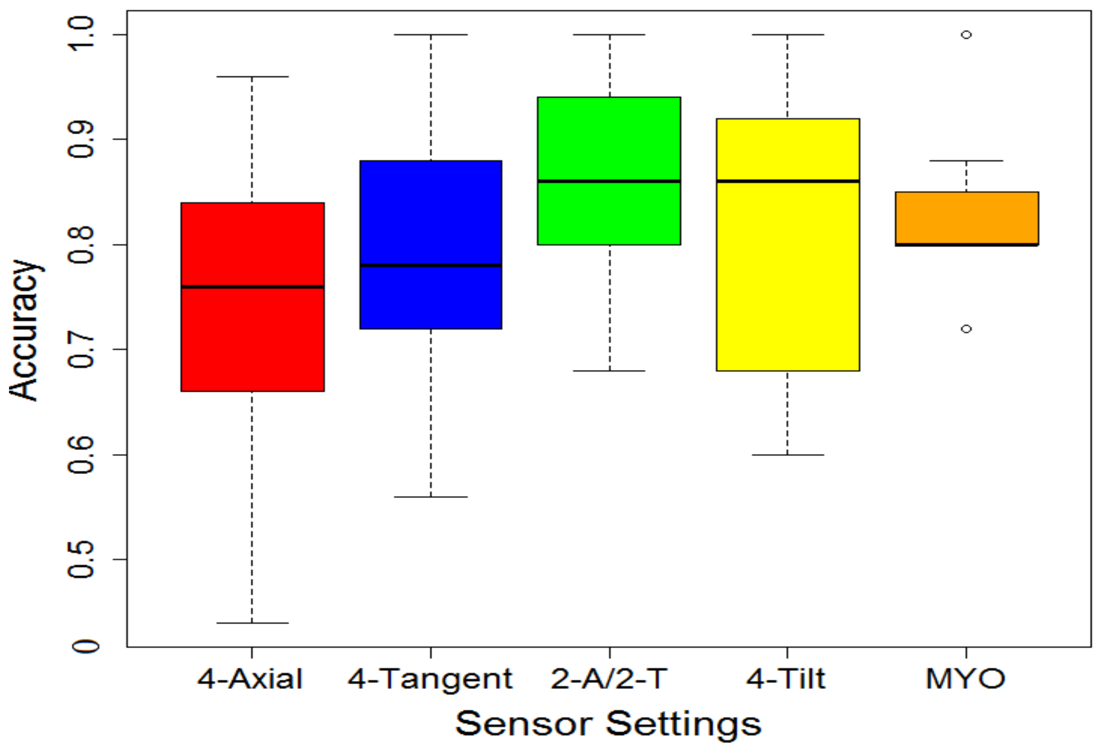

Figure 3.5: Overall recognition rates for each of the four configurations of the bend sensors in WristRec.

Recognition accuracy for our WristRec prototype and for the MYO is calculated by dividing the number of correctly identified gestures by the total number of gestures performed. Figure 3.5 depicts the recognition accuracy for each of the sensor locations identified in our pilot study and for the MYO. As shown, the two-axial-two-tangential and the four-tilt combination give highest

	Left	Fingers Spread	Right	Fist	Double Tab
Left	58	1	0	0	1
Fingers Spread	0	51	3	5	1
Right	1	1	52	1	5
Fist	0	1	1	50	8
Double Tap	0	2	0	11	47

Table 3.1: Confusion matrix for the MYO Gesture Set using our WristRec.

median accuracy; however, the two-axial-two-tangential mean accuracy is highest (86% versus 82%), a result of higher variance in the four-tilt sensor positioning. Given these results, our final WristRec prototype, Figure 1.1 on the right, uses the two-axial-two-tangential bend sensor positioning combination, 2-A/2-T in Figure 3.3. Table 3.1 shows the confusion matrix for our final sensor positions. The most difficult gesture to distinguish is the double tap gesture, and this gesture is frequently confused with the fist gesture.

Comparing this accuracy with the accuracy for MYO, we see mean MYO accuracy of 82% (SD: ±8.27%) for the 10 out of 12 participants who participated in block 2 of our study. A paired t-test (n = 10) indicates that these results are not statistically significant.

3.2.5 Discussion

Our early results with our simple gesture recognizer prototype, WristRec, are promising. By carefully tailoring sensing locations from user input, we find that WristRec performs on par with the Myo arm band (86% for our configuration versus 82% for the Myo, not significant). While the lack of counter-balancing in our initial study may have negatively impacted the performance of our system or of the Myo, from the perspective of feasibility, our data provides evidence that our system has the potential to recognize the Myo gesture set with good reliability and with hardware that can be embedded into an everyday device such as a watch or fitness tracker (in contrast to the Myo which is a special purpose device worn on the upper arm).

3.3 Summary

In this chapter, we showed that WristRec is capable of recognizing a small gesture set currently being used in one commercial product. We proved this by first determining the optimal placement for sensors, and then running a two-phase study to compare the accuracy of our wrist band and MYO armband.

Given our early success, a related question is how well our prototype system can work on other possible gesture sets.

There are many different hand and finger recognizers in past research literature. One challenge in evaluating any gesture recognition device is that, when comparing research contributions, it seems difficult to do head-to-head comparisons. This is because many systems use unique gesture sets (perhaps tuned to the configuration of sensors used as input). For example, in seven different wrist-worn gesture sensing devices described in references [3, 20, 27, 28, 43, 58, 68], we found only limited overlap in gestures. At the same time, some systems have also reported high recognition accuracy using only the sensors contained in a modern smartwatch [63]. Our earlier result focused on five gestures, a common number for these gesture input systems, but any real-world system undoubtedly should recognize a larger set.

In the next chapter, we will try to tackle the problem of inconsistency in gestures tested in previous research by running an elicitation study. This will give us insights into what kind of gestures are acceptable to general users. With the elicitation study result, we will also have a consensus gesture set on which we can draw insights to wrist-worn gesture sensing.

Chapter 4

Elicitation Study

Given the lack of consensus on a reference gesture set for finger- and wrist-gesture input, in this chapter, we describe the results of an elicitation study to understand gestural input. In early pilot studies eliciting gestural input, we noted that, to perform gestures with hands, there are three joints on a human arm that define these gestural manipulations: fingers, wrist, and elbow. Free-space gestures also comprise arm gestures which also involve shoulder movement, but we explicitly focus our elicitation study away from arm gestures and toward hand and finger gestures both because of the gorilla arm effect [12] and potential for public embarrassment [50].

One open question is whether past gesture elicitation studies might provide information to us, thus obviating the need to perform yet another elicitation study. While it is true that many elicitation study for gestures have been done, our analysis found that these studies either allowed large-scale full arm gestures [53], any body part gestures [21], or strictly limited studies to a very small number of joint movement (i.e., microgestures) which limited the participants' freedom to perform a preferred gesture [19].

Our interest is specifically in hand- and wrist-gestures. We feel that there is a role for these gestures in input. For example, in studying microgestures, Chan et al. [19] found that participants showed a desire to use larger arm movements and spatial recognition. One challenge with allowing free-form large arm movements is that users often do not consider the long-term impacts of fatigue for these gestures [53]. Our approach to elicitation is most similar to Vatavu *et al.* [60] who drove their elicitation study by restricting users to gestures that could be recognized with the Leap Motion [9], which limits the gestures to only hand gestures within a spatially constrained region. Given our wrist-worn recognizer and a smartwatch with on-board IMU, we restrict gestural input to movements of the fingers, wrist, and elbow. This allows any hand, wrist, or forearm movement to be considered as a gesture.

We synthesize the results of our elicitation study, analyze the gestures created, and discuss the similarities and differences between the elicited gestures and the gesture sets used in past work on wrist-worn gesture sensing [20, 43, 63, 68, 69].

4.1 Participants

We recruited 16 participants (10 male, 6 female) to participate in our elicitation study. All participants used their non-preferred hand (left hand) to wear a LG G Watch R smartwatch. We informed them that they could perform gestures by moving their fingers, wrist or elbow to issue commands.

4.2 Elicited Tasks

While it is important to consider both task and context in elicitation studies, it is also the case that participants frequently overload gestures by considering context-sensitive interpretations (e.g. the phone-to-mouth gesture in smartphone motion gestures varies depending on whether the phone is ringing [52], drag with one finger in surface gestures varies depending on whether the gesture starts on the object or not [65]). We chose three contexts for our elicitation study: interacting with an external display, interacting with a wearable device such as a smartwatch, and controlling a smartphone while the phone is not in the user's hand. For contexts, we considered on-watch interaction versus watch-as-input interaction: in real-world usage, sometimes watch input is directed to a watch application, sometimes watch input is directed to another device (e.g. phone). For eyes-based input versus eyes-free input, it refers to whether the user needs to look at the display in order to interact with the application. To limit the length of the study, we restricted our application domain to map navigation for the first two contexts; for the third context, a smartphone while the phone is not in the user's hand, we considered audio-only interactions and selected two applications to control: the phone application and a music application. Our tasks are shown in Table 4.1 with the corresponding texts shown in brackets. Our task selection was heavily influenced by common tasks in Shimon's and Ruiz's papers [15, 52]. There were a total of 23 task commands elicited from the study.

Application	Task
1. Map Application on Smartwatch Display (on-watch & eyes-based)	Pan Left Pan Right Zoom In Zoom Out Activate Help Go to Home Scroll Up Scroll Down
2. Map Application on External Display (watch-as-input & eyes-based)	Pan Left Pan Right Zoom In Zoom Out Activate Help Go to Home Scroll Up Scroll Down
3. Phone Application (audio only) (on-watch & eyes-free)	Answer Call Hangup Call Ignore Call
4. Music Application (audio only) (on-watch & eyes-free)	Play Stop Play Next Play Previous

Table 4.1: Tasks for Elicitation Study.

4.3 Procedure

At the beginning of the experiment, the researcher explained the study and asked the participants to wear an LG G smartwatch on the hand on which he/she usually wears a watch. We asked participants to wear the watch as a placebo, thus encouraging them to bias toward gestures that could, conceivably, be sensed by treating the watch as a quasi "magic brick" [52].

All participants were shown a video showing three possible joints which they could use to generate gestures (finger joints, wrist joint, and elbow joint) and two example gestures for each joint. For finger movement, they were shown "make fist" and "telephone hand sign"; for wrist movement, they were shown "wrist sway left" and "wrist sway right"; for elbow movement, they were shown "elbow sway left" and "elbow sway right". We also showed two combination movements, "make fist then elbow sway to left" and "make fist then wrist sway right". The participants were informed they would be asked to come up with a preferred gesture for each task using any of three joints. They were also told that the video was merely an illustration of the possible joints to be used, and they could create any gesture they wish using finger, hand, or forearm movements. We also asked that within an application domain (e.g. a map app on smartwatch), they could not reuse gestures. However, we did inform them that they could re-use gestures after switching application domains.

To simulate the map application, a Google map was shown both on the smartwatch and on a nearby projection screen. All application domains were counter-balanced using a Latin square and the tasks within an application domain were also counter-balanced.

For each task, we asked participants to come up with two gestures, though if they could not come up with a second gesture, they were allowed to skip the second gesture. We did this to ensure that we elicited a large set of possible gestures within the time limit; at the same time, we wanted to avoid frustrating participants if they could not come up with a second reasonable gesture. We also encouraged participants to vary the joint used in each gesture elicited to encourage participants to fully explore the richness of finger, hand, and forearm movement.

After identifying gestures for each task, participants were asked to perform the gestures and to describe the gestures and the joints involved. Participants were also asked to rate each given gesture on a 7-point scale (1 being least- in the category, and 7 being most- in the category) on three categories: Fatigue, Naturalness of the Gesture, and potential Embarrassment in public settings. They were told that the rating would be used to compare the different gestures. At the end of the study, in a semi-structured discussion, participants were asked their preference for joints from most preferred to least preferred for gestural interaction.

The study was video and audio recorded for the post-study analysis.

4.4 Elicitation Study Results

Often the goal of an elicitation study is to develop a consensus set of gestures, but this can be challenging because of typically relatively low consensus scores [59]. As well, while the presentation of technology as a "magic brick" [52] can be useful for eliciting mental models, we believe that insights are particularly useful when interpreted flexibly, as one input to gesture set design alongside a host of other constraints [64].

In our case, the primary goal was to leverage our elicitation study to drive insight into the validation of gesture sets used by others in evaluation. We first explored agreement scores. However, to further address the question of hand and finger gesture input, rather than develop a taxonomy that includes aspects of context and mapping [52, 65], our focus on sensing and recognition technology guides our analysis. Alongside agreement scores, we explore the joints manipulated by participants (including preference for joints) and the parameters used to discriminate between gestures. We also examine participants subjective feedback on the suitability of the gestures selected broken out by specific joints used.

4.4.1 Agreement

Because participants could choose from different joints and a total of 7 different possible combinations of joints to perform gestures, gestures exhibited very low consensus [60]. As a result, we grouped similar gestures together in a task, then calculated the agreement score following guidelines from Chan et al., Morris et al. and Piumsomboon et al. [19, 44, 48]. Specifically, we used the following guidelines from related work to group the gestures:

- If gestures differ only by direction, i.e. "turn wrist to left" and "turn wrist to right", we grouped these gestures together [44].
- As per Chan *et al.* [19], gestures that use two or fewer fingers (i.e. hand faces down, index finger moves up and index plus middle finger moves up) were grouped together and gestures that use three or more fingers (i.e. all five fingers touching together and first three fingers touching together) were grouped together.
- We leveraged Piumsomboon's definition of gesture similarity, i.e. path gestures that have consistent directionality although the gesture is performed with different static hand poses are grouped together [48], for example fist-moves-right and open-hand-moves-right.

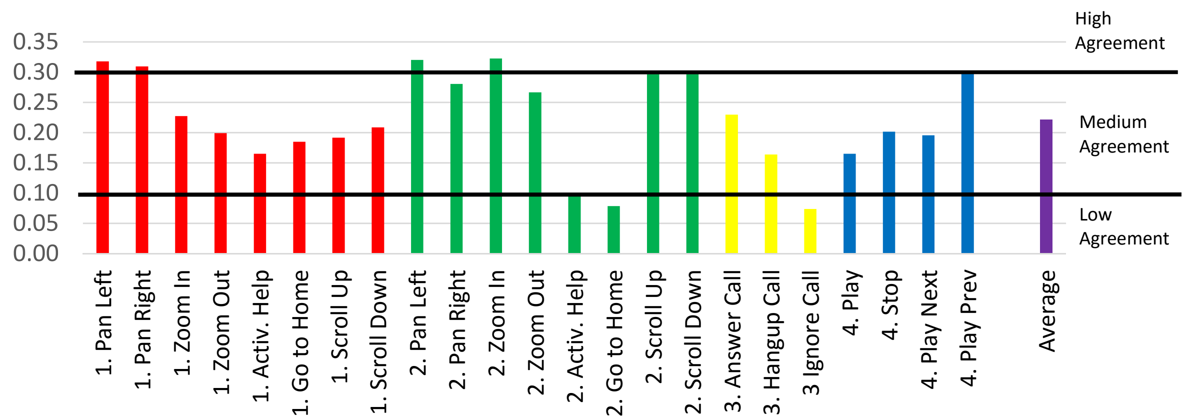

Figure 4.1: Agreement score for all 24 tasks. 1-4 represent the applications as shown in Table 4.1: Map Application on Smartwatch Display (1), Map Application on External Display (2), Phone Application (3), and Music Application (4).

Grouping the above three gesture types together helps us to understand whether moderate or high agreement exists on, respectively, wrist movements, finger movements or poses, and hand poses. This, then, can guide the design of gesture sensing by targeting types of movement.

To measure agreement on our grouped gesture sets, we used the formula proposed by Vatavu *et al.* [59] to calculate the agreement score of our grouped gestures:

$$AR(r) = \frac{|P|}{|P|-1} \sum_{P_i \subseteq P} \left(\frac{|P_i|}{|P|}\right)^2 - \frac{1}{|P|-1}$$

where P is the set of all proposals for a task, $|P|$ is the size of the set, and P_i are subsets of identical gestures from P.

Using Vatavu's interpretation of agreement values, our results ranged from 0.071 (low agreement, $AR \leq 0.100$) to 0.323 (high agreement, $0.300 < AR \leq 0.500$) as shown in Figure 4.1. Most gestures exhibited moderate agreement, $0.10 \leq AR \leq 0.30$, an expected result [59]. Gestures with moderate or high agreement included all of wrist movements, finger poses and movements, and hand poses.

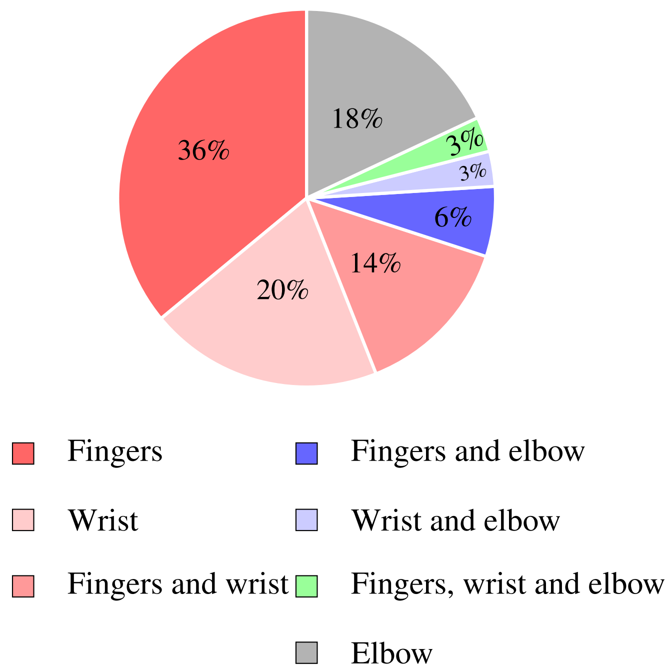

Figure 4.2: Distribution of joint usage for all gestures.

4.4.2 Joint Analysis

The next question we asked was whether there was a bias toward one joint or another during analysis. The distribution of joints used for each gesture are shown in Figure 4.2. We can see from this figure that 74% of the gestures involve only one joint, 23% of the gestures involve two joints, and 3% of the gestures involve three joints. Of these single joint gestures, fingers are the most common, then wrist, and elbow is least common. Fingers and Wrist is the most common multi-joint combination. Elicited joint preferences during interviews triangulate well with this data: the most preferred joint is fingers (12/16 participants, followed by wrist 5/16, and none chose elbow as their first choice). Note that one participant chose both wrist and finger as his first choice. Finger and wrist movements accounted for over 70% of all elicited gestures.

Difference	No. of participants
Opposite Movement	16
Number of Fingers	15
Finger ⇔ Wrist	14
Wrist ⇔ Elbow	14
Post-motion Hand Posture	13
Wrist ⇔ Combination	12
Gesture Orientation	12
Pre-motion Hand Posture	11
Finger ⇔ Elbow	11
Finger ⇔ Combination	11
Elbow ⇔ Combination	10
Number of Times	8
Magnitude of the Elbow Movement	5
Magnitude of the Finger Movement	4
Gesture Duration	3
In-motion Hand Posture Change	2

Table 4.2: Differences participants used to distinguish their proposed gestures.

4.5 Parameters for Gesture Discrimination

Recall that, in our study design, participants were asked to avoid reusing a gesture within the same application context. As a result, one question we explore in our data is how participants discriminate one gesture from another. Table 4.2 shows the parameters participants used to differentiate the gestures. For example, all of the participants used opposite direction of movement (e.g. flick right versus flick left) to differentiate two or more gestures, fifteen of our sixteen participants used number of fingers (e.g. bend index finger versus bend index and middle finger), fourteen of sixteen switched joint combinations to discriminate gestures, and thirteen used post-gesture hand pose to discriminate between gestures (e.g. gestures were different depending on whether the hand faced up or down at the end of the gesture). In contrast, features such as timing (gesture duration) and in-motion hand pose changes only rarely occurred.

For scrolling tasks, participants prefer to use fingers to point towards the external large display and move wrist and elbow, while simply moving wrist up and down for the watch. This shows the participants' feeling about the ownership of the device. Since the watch is already on the users' wrist, the users know that the movement would be detected by the watch. However, since the external display is public, users are explicitly pointing to it to declare that their next action is meant for the external display.

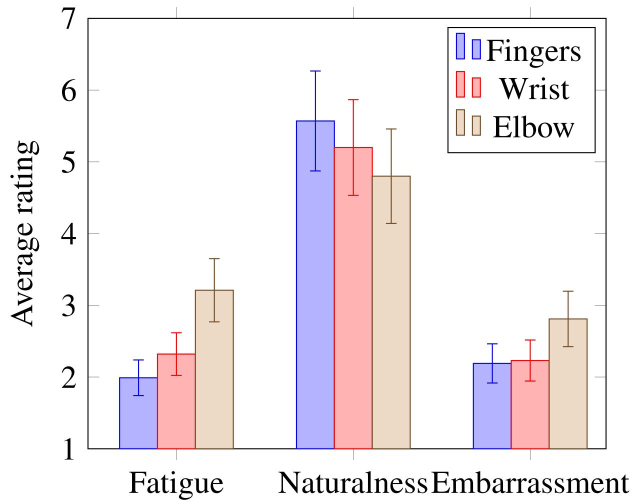

Figure 4.3: Average ratings for each joints with 95% CI. 1=Strongly disagree, 7=Strongly agree.

4.6 Subjective Rating of Gestures

Finally, the average rating for fatigue, naturalness, and embarrassment is shown in Figure 4.3. Of interest in this Figure is that finger and wrist are perceived of as significantly less fatiguing and embarrassing (low numbers are good for fatigue and embarrassment). While not statistically significant, qualitatively we note that fingers and wrist are also perceived to be more natural than elbow movements.

4.7 Consensus Gesture Set

Participants in our study prefer gestures that leverage finger or wrist manipulations over forearm movement. To generate a consensus gesture set, we chose the wrist and finger gestures with the highest agreement score for each task.

Table 4.3 shows our consensus wrist and finger gesture set. We found our consensus gestures can be classified into two categories: component gesture and compound gestures. Component

No.	Gesture Type	Components
1.	Component	b. Fist
2.	Component	d. Phone
3.	Component	e. Thumb
4.	Component	h. Flex
5.	Component	i. Extend
6.	Component	i. Close
7.	Compound	g. Close + e. Spread
8.	Compound	a. Point + m. Pro + h. Flex
9.	Compound	a. Point + l. Sup + h. Flex
10.	Compound	m. Pro + h. Flex
11.	Compound	l. Sup + l. Sup
12.	Compound	m. Pro + j. Adduct
13.	Compound	f. Pinch + c. Spread
14.	Compound	m. Pro + k. Abduct
15.	Compound	m. Pro + i. Extend

Table 4.3: Our set of 15 wrist and finger consensus gestures collected from elicitation study. The letters in front of the Components column are referring to the gestures labeled in Figure 4.4.

gestures are atomic gestures that involve only finger or only wrist movement. Component gestures are a hand pose, a single finger movement, or a single wrist movement. Compound gestures are those that consist of multiple component gestures. Figure 4.4 depicts all component gestures that comprise our consensus gesture set.

4.8 Discussion

Elbow movement is easy to capture using 9-axis IMUs available in commodity personal devices such as smartwatches [32]. Unfortunately, because elbow movement is the least common and least desired gesture type, it is clearly the case that finger, hand, and wrist manipulations must be sensed. This increases the complexity of the capture problem and validates the importance of on-going research in novel gesture sensing technologies to capture and interpret wrist and finger movement.

The next question is whether we should consider wrist, hand, or finger manipulations as targets for gesture sensing. Based upon our elicitation data, all manipulations were observed in

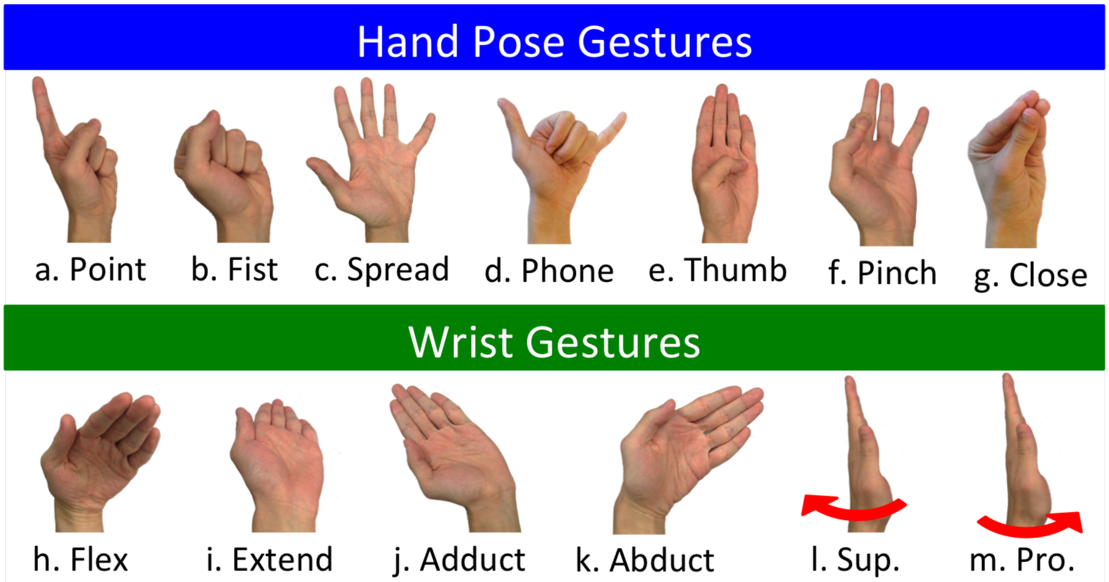

Figure 4.4: 13 component gestures determined from the consensus gesture set.

our data set. To identify a consensus set of finger and wrist gestures that could be used as a more appropriate referent than what has been used in past research, we focus on two data sources. First, from past work, we identify wrist, hand, and finger gestures that have been used in past evaluations of systems [20, 27, 43, 63, 68]. Second, we leverage gestures from our elicitation study to ascertain coverage of the elicited gestures.

Leveraging gesture sets from the WristFlex system by Dementyev and Paradiso [20], the Empress system by McIntosh *et al.* [43], the Serendipity system by Wen *et al.* [63], the Tomo system by Zhang *et al.* [68], and the Myo gesture set from Thalmic Labs [4], the union of gesture sets produces a set of 27 gestures, shown in Figure 4.5. One thing that distinguishes these gestures from the consensus gestures depicted in Table 4.3 is the absence of compound gestures. The gestures in past research are typically either a single hand pose, hand movement, or wrist movement, whereas over half of the elicited gestures include some combination of movement and hand pose or some combination of movements.

If one contrasts the gesture set in Figure 4.5 to the component-based gesture set, Figure 4.4, we can see that, with two minor differences (the *d.Phone* gesture in Figure 4.4 is a slight variant on the *d2.Spiderman* gesture in Figure 4.5, the *g.Close* gesture in the Figure 4.4 is the combination of *Pinch* gestures in the Figure 4.5), the *union* of gestures in Figure 4.5 covers virtually all component gestures. However, no one gesture set covers all component gestures. In

Figure 4.5, we include the origin of gestures (Empress, Tomo, Serendipity, WristFlex, or Myo); the closest gesture set that fully covers all component gestures is the Empress gesture set, which covers 11 of the 13 component gestures. Balanced against this limited coverage, it is also the case that every gesture set that comprises the 27 gesture set provides some coverage of the component gesture set from the elicitation study.

4.9 Summary

In this chapter, we showed the result our elicitation study on hand, wrist, and elbow gestures that can be used under different contexts. From the result, we made two observations about gesture sets which have been used by previous research. First, in terms of strengths, past research has selected gesture sets that are natural products of our elicitation, an indication that past research has focused on a reasonable set of gestures. While some individual gestures (e.g. the d2.Spiderman gesture in Figure 4.5) are rare based upon elicitation, the overall movements examined by past capture systems are appropriate. We analyzed coverage of finger- and wrist-gestures as collected from our elication study. Neglecting orientation as a factor, we found that this 27-gesture vocabulary collected from the literature contains 72% of gestures obtained from the elicitation study. Furthermore, it covers all gestures of moderate or high agreement. Second, in terms of weaknesses, past research has focused primarily on simple component gestures, rather than on the combination of pose+wrist of multiple wrist gestures in sequence. In the 27 gesture set, all gestures are component gestures.

One question that our observation of this gap poses is how effectively past recognizers perform on compound gestures and also what is needed to recognize gestures like those elicited from the gesture set in Table 4.3. In the next chapter, we will tackle the problem by comparing the recognition accuracy of our self-built wrist-worn device WristRec with IMU based solution using the gesture set obtained from the elicitation study.

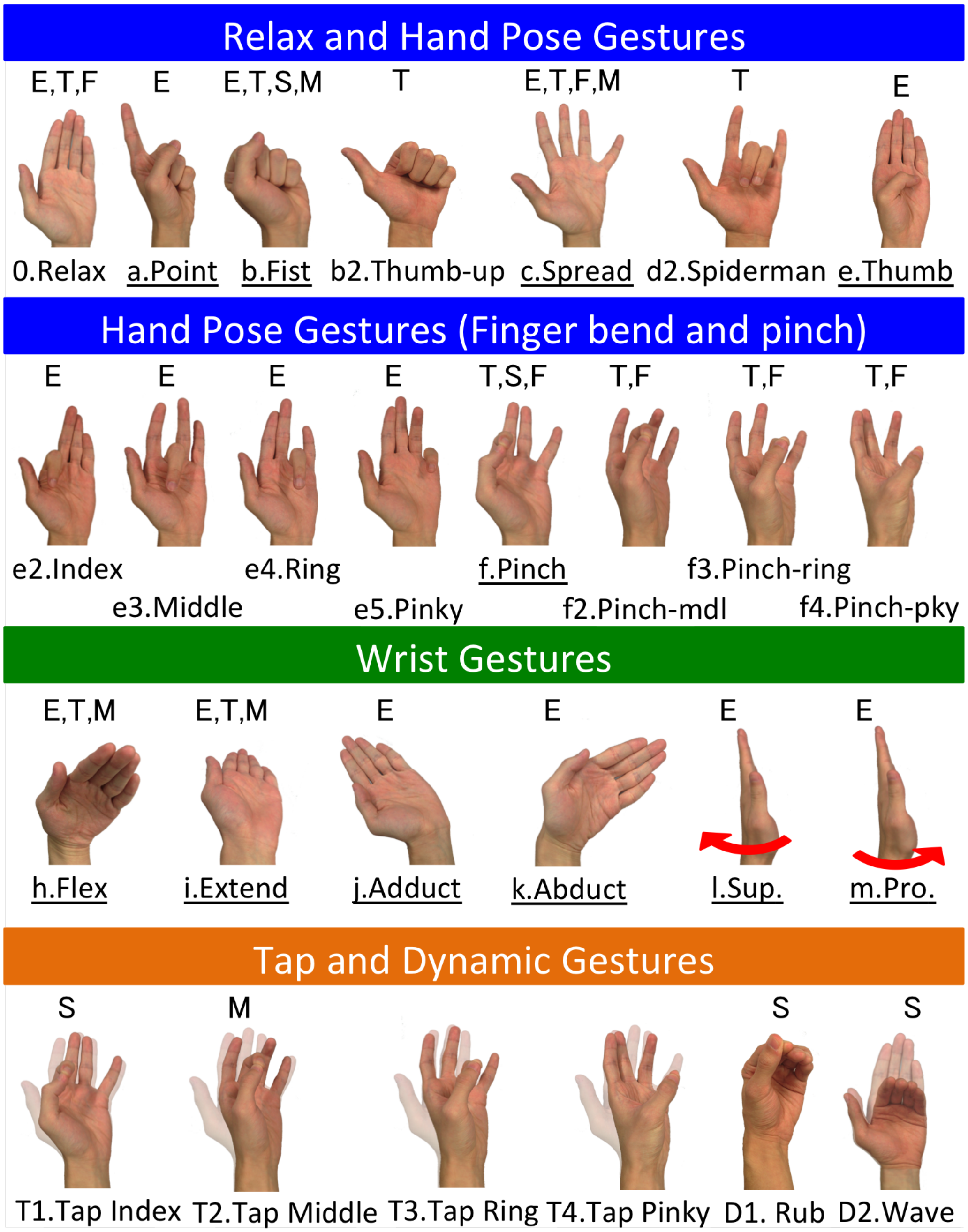

Figure 4.5: Our set of 27 gestures collected from related work and 13 component gesture. The alphabets represent the related work. E:Empress, T:Tomo, S:Serendipity, F:WristFlex and M:MYO armband.

Chapter 5

Is the IMU Enough?

Given our understanding of component and compound gestures, a related question regarding wrist-worn gesture sensing is the technical requirements for capturing input. In particular, the Serendipity system of Wen *et al.* [63] has achieved a relatively robust recognition on a set of dynamic hand gestures using only the IMU on an off-the-shelf smartwatch. Is an off-the-shelf smartwatch sufficient for hand pose and wrist gesture recognition? What is the difference between IMU-only and IMU+wrist-worn sensors in terms of recognition accuracy?

To address these questions, we did the following two studies:

- We evaluate bend sensor alone, IMU-alone, and bend+IMU recognition using WristRec described in Chapter 3 on the consensus gesture set from the elicitation study.
- We explore discrepancies between our implementation of IMU-only recognition and the Serendipity system, and verify the fidelity of our re-implementation of Serendipity's algorithm. We tested this using the 27 gesture set obtained from previous work.

5.1 Evaluation 1: Consensus Gesture Set Feasibility Study

In this section, we used both bend sensor, accelerometer and gyroscope data as captured by a smartwatch IMU, and the combination of data sources to evaluate the performance of our system using the consensus gesture set obtained from the elicitation study.

Figure 5.1: LG smartwach used in the experiment

5.1.1 Participants

12 paid participants were recruited, 9 male and 3 female. All participants used the hand they would typically wear a watch on to perform gestures. For 10 participants, this was their left hand; for 2 participants, this was their right hand.

5.1.2 IMU Based Recognition

To evaluate the IMU-based recognition, we replicated the recognition algorithm described in the Serendipity system [63] which uses only the IMU to recognize gestural input. To ensure a high fidelity re-implementation, we worked directly with the first author of the Serendipity system [63] to clarify ambiguities in implementation.

5.1.3 Apparatus and Data Collection

Using our hardware configuration described previously in Chapter 3, we captured data from both the bend sensors (via an Arduino board) and the IMU on the LG smartwatch (Figure 5.1). Participants wore the WristRec device first, then the smartwatch was worn on top of the WristRec. The Arduino board was connected to an external computer via USB cable and the smartwatch streamed IMU sensor data to the same computer via Bluetooth. The sampling rate for the bend sensors was 100Hz, and the sampling rate for the accelerometer and gyroscope exceeded 200Hz. Data were synchronized on the external computer. We used two 2-inch tilt sensors at 0°and 180°and two 1-inch bend sensors oriented tilt at 90° and 270° for this study. This setup ensures that the sensors cover as much area around the wrist as possible.

5.1.4 Method

Our experimental methodology in both number of participants and in procedure replicates the experimental methodology of Zhang and Harrison [68]. As well, gestures were presented in random orderings to limit carryover effects and muscle memory action similarity. For completeness, we provide a brief overview of our methodology.

Participant was welcomed to the lab, the purpose of the experiment was explained. Our recognizer was placed on participant's wrists and participant was seated at a table facing the computer used to coordinate experimental data.

We collected the 13 component gestures (Figure 4.4) and 9 compound gestures (the compound gestures from Table 4.3) in two separate sessions. First, participants completed a training block. A video of the gesture was shown before each gesture, and the participants performed all gestures in the gesture set once to familiarize the gesture set and system interface. Second, participants completed ten data collection blocks. For the component gestures, a picture of the gesture was shown in the computer display to cue the gesture. For the compound gestures, because the multi-component movement was difficult to understand using icons, we showed participants the same video we used for the training block.

During each trial, participants start from a resting position with their elbow on the armrest. A gesture is presented on a computer screen and participants press the space key on a keyboard with the other hand to start the gesture. The participants hear two auditory cues for each gesture: the first one is a signal at the time they press the key for them to start performing the gesture; the second one marks the end of the transitioning move of the gesture and participants were asked to finish their gesture before this second cue. Audio cues were spaced at 3.5 seconds apart for the compound gestures (due to combined movement requirements) and at 3 seconds apart for the component gesture set. Participants performed each gesture set one time each in a block and they repeated this for 10 blocks. As noted, the order of the gestures within the block was randomized. In total, we collected 220 gestures per participant.

5.1.5 Data Cleaning and Classification

As shown in Figure 3.4, magnitude, form, and dynamics of gesture input all vary across gestures. Therefor, we classified gestures by distinguishing the different data patterns.

To clean the bend sensor data, we first normalized the data by using the following formula:

$$z_i = \frac{x_i - min(x)}{max(x) - min(x)}$$

for each bend sensor. We then ran the data through a low pass filter with a cutoff frequency of 10Hz to eliminate noise. To reduce feature space, we down-sampled the data to 30 data points spaced equidistant over the input. From 3 seconds of component gestures and 3.5 seconds of compounds gestures, this process produced 10 Hz data for the component gesture and 11.7 Hz data for the compound gesture. Overall, our final data was a vector of 248 entries: the bend sensor data down sampled to 30 data points from each location (4x30), the deltas for each sensor between adjacent samples (4x30), and the overall minimum and maximum of each sensor input (8 values).

We perform recognition in three ways:

1. We use bend sensor data alone to perform recognition.
2. As noted earlier, (and with assistance from the first author [63]), we replicate the same data processing algorithm described in the Serendipity system [63] which uses only the accelerometer and gyroscope data from IMU to recognize gestural input. The raw sensor input contains data from three axes: x, y, and z. From these three data, we calculated magnitude of the combined axes ($\sqrt{x^2+y^2+z^2}$, $\sqrt{x^2+y^2}$, $\sqrt{y^2+z^2}$, $\sqrt{x^2+z^2}$). Therefore, we have 7 features for each sensor. Next, for each one of the features, we calculated the mean, standard deviation, max, min, 3 quantiles. Lastly, we took 30 power bands from a Fast Fourier Transform (FFT), yielding a total of (30 + 7) x 7 features x 2 sensors = 518 features.
3. We combine the bend sensor data and IMU watch sensor data to form an enlarged vector feature space for recognition. We used the expanded 766 feature vector combining 248 bend sensor features described above with 518 IMU vectors we used for Serendipity algorithm.

Given the above 248-point vector for bend sensor data and 518-point vector for IMU data, we analyzed data using both SVM with polynomial kernel in the same manner as Serendipity [63] and random forest algorithm. For the SVM, the data was fed into a multi-class classification SVM as an n-dimensional vector (time was a dimension) and cut with hyperplanes (Polynomial kernel).

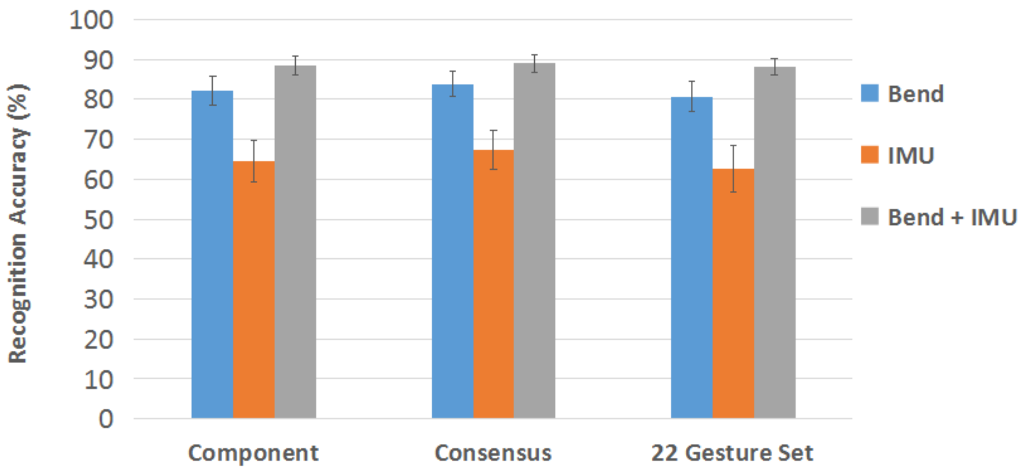

Figure 5.2: Random forest recognition accuracy for Bend-only, IMU-only, and Bend+IMU for different gesture sets. Error bars show 95% CI.

We set the parameters for the Random Forest algorithm to the following values: the number of features to consider in feature selection is calculated by 1+$\log_2$(total number of features), the max depth of a tree is 100, and the number of trees is 100. We analyzed data using 10-fold cross-validation.

5.1.6 Results

Overall, random forest algorithm outperformed SVM for each configuration. Figure 5.2 shows the random forest algorithm accuracy for the three gesture sets with different sensor readings. As a final test, we also combined all component and compound gestures into a single data set (yielding 22 gestures overall – all gestures in Figure 4.4 plus all compound gestures in Table 4.3). Our recognition accuracy on this 22-gesture gesture set with Bend sensor and IMU was 88.1%. Two-way repeated measure ANOVA (Sensor Type and Gesture Set as within-subject variables) finds a significant effect of Sensor Type ($F(2,22) = 44.3$, $p < 0.0001$) and Gesture Set ($F(2,22) = 23.226$, $p < 0.0001$) on recognition accuracy. Post-hoc Bonferroni shows Bend+IMU (88.6%) was significantly higher than both Bend only (82.5%) and IMU only (65.1%), and that Bend only is significantly higher than IMU only. For Gesture Set, Consensus (15 gestures in Table 3) is significantly higher than Component (13 gestures in Figure 3.) and 22 Gesture Set (13 Component gestures + 9 Compound gestures in table 3 used as a single set), and that Component is significantly higher than 22 Gesture Set. No interaction between Sensor Type and Gesture Set.

5.1.7 Discussion

From the perspective of gesture recognition technology, we believe that our results show that there exists complementary information in both wrist and IMU sensing that can serve to significantly enhance recognition of gestures of the form identified in our consensus gesture set. In analyzing the potential strengths of IMU-based recognition versus wrist-deformation-based recognition, it seems intuitive that the IMU would be best equipped to handle gestures where significant movement occurs at the wrist, whereas gestures that simply require finger movements might best be recognized by wrist sensing. To explore this theory, we examine, again, our component gesture set, Figure 4.4.

Within Figure 4.4 , we separate gestures into two categories: hand pose gestures where only finger movements are required; and wrist gestures where significant wrist flex or rotation occurs either because of movement at the wrist joint or at the elbow joint. Since IMU is good at measuring large movements, we further looked into the accuracies for Finger Pose Gestures and Wrist Gestures separately. From the calculation, we saw that IMU alone can detect Wrist Gestures with 78.6% accuracy and Hand Pose Gestures with 52.6%. This makes sense because Hand Pose Gestures cause little movements to the wrist, therefore it is hard to be detected by the IMU. After adding our WristRec data, the accuracies boost up to 81.7% and 96.53% respectively. This shows that even with Wrist Gestures, which would cause movements on the watch, additional sensors are still needed in order to detect gestures with a high accuracy. Table 5.1 summarizes these findings.

	Bend Only	IMU Only	Bend+IMU
Hand pose	75.83%	52.62%	81.67%
Wrist	89.72%	78.61%	96.53%

Table 5.1: Comparing Accuracies on Hand Pose versus Wrist gestures as defined in Figure 4.4.

Leveraging an experimental technique described by Zhang and Harrison [68, 69], we also explored reproducibility and generalizability. While our recognition accuracy was higher than Zhang and Harrison, it was primarily due to the addition of the IMU signal. However, our reproducibility and generalizability result were not good. We suspect, as per Zhang and Harrison [68], this is because biological internal wrist signal during movements are sufficiently different between users that wrist-worn recognition may be user-specific.

5.2 Further Explorations of Hand Pose Recognition

The results on our consensus gesture set indicate that Serendipity's IMU-based recognition algorithm struggles to provide high accuracy, particularly on hand pose data, and that adding wrist-worn sensing can improve recognition (i.e. the IMU-alone is less robust than IMU+wrist-worn sensing). However, in our interactions with researchers of the Serendipity system, one thing that became clear was that significant protocol differences exist between our approach to evaluation – inspired by Zhang and Harrison [68] – and the Serendipity approach. Zhang and Harrison [68] approach recognition accuracy conservatively, ensuring distance in time between one gesture and the next. Separating gestures in time means that the variation between how participants perform a gesture may be significantly greater, i.e. there is more variability in input.

In the original evaluation of the Serendipity system [63], to ensure kinematic consistency between movements, each gesture was performed multiple times during data collection (i.e. each gesture was collected in turn, performed multiple times, then the next gesture was collected). This approach can be valuable when one wants to assess expert-level recognition accuracy, where long term use and interaction with recognizer behavior [30] can be used to foster more consistent gesture performance. Note that the one constraint with this style of testing is that one must ensure that the sensors are normalized between subsequent gestures; Wen et al. were careful to do this in their data collection [63], so we suspect the increased accuracy may be due to more consistent kinematic performance of gestures.

In this section, to clarify distinctions due to evaluation strategy, we conduct a small-scale experiment. Because of the questionable nature of performance on finger-pose gestures, we expand our finger-pose gesture set significantly by using the 27-gestures identified in Figure 4.5. Our rationale for using the 27-gesture set was as follows: If we mimic Wen et al.'s evaluation strategy [63] but collect a larger gesture set gleaned from past research [20, 43, 63, 69], we can *both* benchmark our implementation of the Serendipity system and contrast IMU-only versus enhanced behavior on other gesture sets.

5.2.1 Participants

Six paid participants were recruited, 5 male and 1 female. All participants used the hand they would typically wear a watch on to perform the gestures. For five participants, this was their non-preferred hand. One participant used his or her preferred hand to perform the study.

5.2.2 Apparatus and Data Collection

As in the previous experiment (and in Zhang and Harrison [68] and Wen et al. [63]), participants were welcomed, seated, the purpose was explained, and the apparatus was placed on their wrist.

Using the same apparatus as in the previous experiment, we collected 27 gestures in Figure 4.5. In this study, we used four 1-inch bend sensors. Each participant performed each gesture in the gesture set ten times repeatedly for a total of 270 gestures per participant. The difference in gesture collection focused specifically on gesture ordering and on gesture timing, leveraging the experimental design of Wen et al. [63]. The participants were given three seconds to perform each gesture since all gestures are component gestures. The participants hear two auditory cues for each gesture: the first one is a signal for them to start performing the gesture and to continue performing the gesture until the second cue; the second cue marks the end of the gesture and participants were asked to move their fingers back to the start position after this second cue. The sound cues were spaced 1.5s (seconds) apart. While the cues were spaced 1.5s apart, we capture 3s of input.

5.2.3 Data Cleaning and Classification Strategy

We used the same Data Cleaning process and recognition algorithms as described earlier. Again using a SVM with polynomial kernel and Random forest, we performed 10-fold cross validation on the three different sensor configurations: bend sensor only, IMU only and combination of both the bend and IMU sensors (Bend + IMU).

5.2.4 Results

One advantage of our 27-gesture set is that it emcompasses the gesture sets of many competing systems including EMPress's 15-gesture gesture set [43] (EMP15), Tomo (both the 8 wrist and 5 finger gesture sets) [68, 69] (Tomo8, Tomo5), Serendipity's five-gesture set [63] (IMU5), and WristFlex's six finger gestures [20] (WFlex6). We also present comparison of performance with component gestures from Figure 4.4 (Comp) which also exist in the 27-gesture set. In this section, we perform a final evaluation of IMU only, bend sensor only, and Bend+IMU recognition to competing systems in the literature.

With this dataset, there was no significant difference between recognition accuracy using SVM and the Random Forest Algorithm. In Figure 5.3, we show a comparison between our implementations and competing systems evaluated on their respective gesture sets. In this figure,

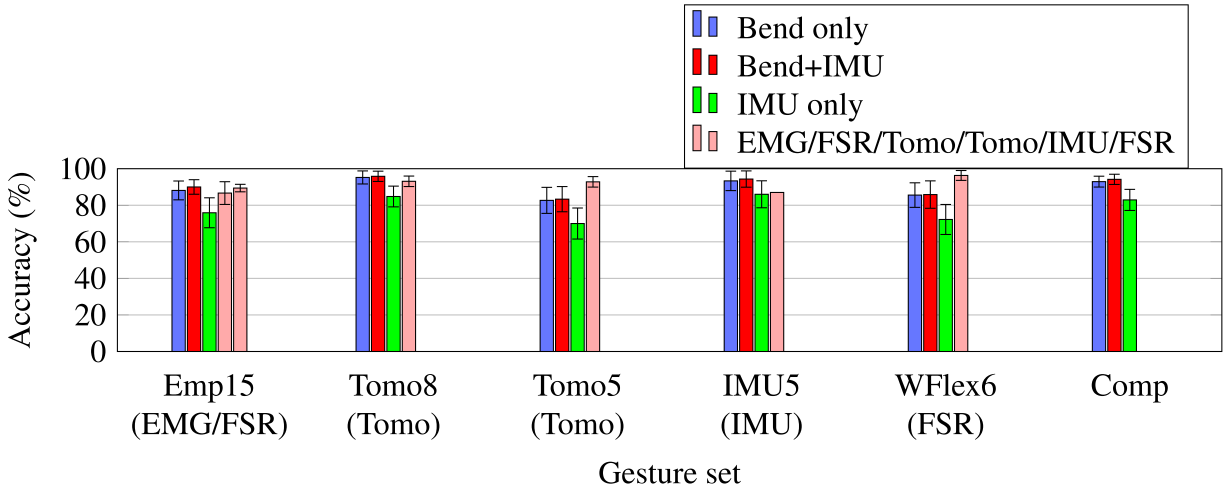

Figure 5.3: Recognition accuracy for each gesture set. Error bars show std deviation.

Bend-only (dark blue bar) is our bend sensor configuration, Bend+IMU (red bar) is our bend sensor augmented with IMU data. Finally, IMU only (green bar) is our reimplementation of Wen *et al.*'s Serendipity algorithm [63]. For each system, we report recognition rates as provided by the authors of the papers. For Empress, two recognition rates are reported, one using sensor fusion and the second only FSR. For Tomo8 and Tomo5, we use Zhang *et al.* (2016) [69] figures for 16-electrode, 4-pole configuration. We believe that this configuration would be most likely to perform reliably on our extended gesture set of 27 gestures because, for tap gestures, dynamic information is needed, and the 3Hz sampling rate of the 32-electrode, 4-pole configuration might be too low to capture tap data.

First, analyzing the Serendipity system, labeled IMU5 in Figure 5.3, our reimplementation (86.0%) is consistent with Serendipity's accuracy on the same gesture set (87%). We believe (given recognition accuracy and standard deviation across data folds) that this indicates that we have accurately re-implemented Serendipity. Furthermore, despite some significant exploration of competing algorithms and hardware enhancements [36], we have been unable to improve significantly on this accuracy for the Serendipity gesture set.

Second, consider changes in the accuracy of recognition when wrist-based sensing is added. Examining the Serendipity gesture set using Bend only data or Bend+IMU, we see a significant boost in recognition – to over 94%. Looking at other competing systems, IMU-based recognition provides surprisingly high accuracy, but enhancing IMU-based recognition with additional wrist-worn sensing consistently enhances recognition.

Given the availability of a 27-gesture set, we provide a final analysis of IMU versus IMU+wrist-

	Bend only	IMU only	Bend+IMU
Accuracy	80.8%	66.0%	83.7%
SD	5.8%	6.1%	5.3%

Table 5.2: Recognition accuracy for 27 gestures using our WristRec with bend sensor only (Bend only), with smartwatch IMU (IMU only) and combination of bend sensor and IMU (Bend + IMU).

worn behavior. Again using 10-fold cross validation with a SVM with polynomial kernel, we used the three different sensor configurations described above: our wrist band with only bend sensors, Serendipity's IMU-based algorithms as reimplemented (IMU Only) and a combination of both the bend sensor and accelerometer-based sensors (Bend + IMU) on the entire 27-gesture set. Recognition results are shown in Table 5.2. Figure 5.4 presents the confusion matrix for our 27 gestures using IMU+Bend data together (overall recognition rate of 83.7%). Note the three areas of higher confusion on individual finger bends, taps, and pinches. In particular, we see higher levels of confusion between index-middle and ring-pinky fingers for both bends, taps, and pinches.

5.3 Summary

In this chapter, we used two studies to show that there is a lot of potential in gesture recognition using wrist-worn devices. Even though researchers have shown that the IMU in modern smartwatches is capable of recognizing gestures with high accuracy, IMU based solutions are mainly good at recognizing gestures that require large movements. To detect a wide range of gestures including gestures that require small movement (e.g. finger tap), we do need the help of other sensors. What kind of, or what combination of, sensors are needed to recognize a wide range of gestures remains an open question for future research.

	1. Bend Thumb	2. Bend Index	3. Bend Middle	4. Bend Ring	5. Bend Pinky	6. Flex	7. Extend	8. Adduct	9. Abduct	10. Sup	11. Pro.	12. Relax	13. Spread	14. Fist	15. Point	16. Thumb-up	17. Spiderman	18. Tap Index	19. Tap Middle	20. Tap Ring	21. Tap Pinky	22. Rub	23. Wave	24. Pinch Index	25. Pinch Middle	26. Pinch Ring	27. Pinch Pinky
1	55	1	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
2	3	45	7	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0
3	0	4	49	4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
4	2	3	2	41	5	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	2	1	0	2
5	1	0	4	6	47	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
6	0	0	0	1	2	55	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	58	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	58	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	4	55	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	1	0	2	0	0	0	0	0	0	0	0	56	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	1	0	0	0	0	0	0	58	0	0	0	0	0	0	0	0	0	0	1	0	0	0
14	0	0	0	0	0	0	0	0	0	1	0	1	0	54	1	0	0	0	0	0	0	0	0	1	1	0	1
15	2	1	0	1	0	0	0	0	0	0	0	0	0	1	47	0	0	2	0	1	0	0	0	0	3	1	1
16	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	53	0	0	0	0	1	1	0	0	2	0	0
17	0	0	0	1	2	0	0	0	0	0	0	1	0	1	1	0	51	0	0	0	0	1	0	0	0	2	0
18	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	38	10	8	2	0	0	0	0	0	0
19	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	9	39	6	4	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	5	6	41	6	0	0	0	0	1	0
21	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	5	4	47	1	0	0	0	0	0
22	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	55	1	0	0	0	0
23	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	55	0	1	0	0
24	2	0	0	1	0	0	0	0	0	0	0	1	2	1	1	0	0	0	0	0	0	1	0	43	4	4	0
25	0	0	2	0	0	0	0	0	0	0	0	0	0	1	1	1	3	0	0	0	1	0	1	5	42	2	1
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	1	4	44	8
27	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	6	50

Figure 5.4: Confusion matrix for 27 gestures.

Chapter 6

Conclusion

6.1 Discussion

The goal of this thesis is to further the understanding in the space of wrist- and hand-gesture input. To accomplish this goal, we:

- build a low-cost wrist-worn device with bend sensors that can recognize gestures from tendon movement.
- present an elicitation study exploring wrist and hand gestures. Leveraging data from this elicitation study and from past work in finger and wrist gesture recognition, we synthesize a set of 27 referent gestures that provide full coverage of gesture sets used in selected recent research systems and provide coverage of medium and high consensus gestures from our elicitation study.
- examine signals from IMUs and additional wrist-worn hardware, and show that adding wrist-worn sensing to IMU-only recognition continues to boost performance.

6.1.1 Examining Gesture Sets

One comment we received during the earlier stage of our research highlights the potential pitfalls of consensus gesture sets: "Though a uniformed gesture set can be helpful, it can also be harmful for future gesture-recognition techniques if not positioned properly. Unlike machine learning or computer vision problems where different methods utilize on the same data (e.g., images), hand gesture recognition systems leverage different signals (e.g., current, sound, light, pressure, etc.). These signals behave differently and therefore it is understandable that previous approaches prioritize gestures that work the best with the signal" [1].

We wholeheartedly agree that consensus gesture sets should never be viewed as constraining. Rather than representing a definitive gesture set, they provide insight into how a group of participants conceptualize interaction. Even within this elicited conceptualization, there are confounds involving legacy biases [44, 53] and pragmatic constraints of recognition feasibility [31]. For us, the more important feedback from elicitation involves an analysis of joint use and preference (Figure 4.2) and how participants distinguish between different gestures in the elicited set (i.e. what parameters they manipulate to turn one parameter into another as in Table 4.2).

While there may be concerns regarding consensus sets, we also believe that an over-reliance on what works well given a signal can be problematic if what works well does not coincide with user perception of gestures that make sense. When one breaks the elicited consensus set down into component gestures (Figure 4.4), one thing we see is that gestures used in past work (Figure 4.5) comprises the building blocks participants used to compose gestures in the consensus set. In this way, consensus sets can also add value to research, by validating the past work that has been done on gesture sensing.

We believe that there are two primary implications of this work for the design of gesture systems. First, from the perspective of comparison, we believe that there exists a benefit to consistency in evaluation. Zhang and Harrison [68] note the challenges associated with finding benchmarked values for comparison with their system. We suggest that the 27 gestures identified in Figure 4.5 can serve as a useful benchmark for recognition systems. Furthermore, we believe that the four categories of gestures identified in Figure 4.5, finger bends, hand gestures, tap and dynamic gestures, and pinch gestures, can serve as a useful starting point for developing a richer taxonomy of gestures destined for recognition by any one system.

6.1.2 IMU-Based Recognition and Beyond

While some commentary on work in gesture sets questions the form that gestures take, a second commentary examines need for any specific gesture sensing technology. The query is often phrased as: "Systems like Serendipity already obtain high accuracy. Do we really need other sensors when the IMU alone already has high accuracy?" While one can take issue with phrasing IMU-based accuracy as a proof of sufficiency, the more generalizable question is whether additional wrist-based sensing can enhance the signal from the IMU. This is not immediately clear: What systems such as Tomo, WristFlex, and EmPress do is sense subtle changes in wrist geometry of internal wrist biophysics. These changes occur in the wrist because of movement, and a high-frequency IMU may be able to discern these changes to a similar level to wrist-worn sensors.

From the perspective of performance, given that our system achieves accuracies above 80% on a 27-gesture candidate set, more complex systems like Tomo [69] should be able to obtain significantly higher recognition accuracies. One of the primary goals of Zhang et al. (2016) [69] was to explore issues of accuracy in recognition; we encourage researchers to expand the number of gestures in their gesture set and to push recognizers toward failure cases. Given the scalability we observe, there is every reason to believe that their sensor configurations could perform with reasonable discriminatory power on larger gesture sets.

One obvious problem with larger gesture sets is a reduction in recognizer accuracy values. For example, our reimplementation of Serendipity's algorithm achieves identical accuracy on their set of five gestures, but then falls to 66% accuracy on the set of 27 gestures, as shown in Table 5.2. Given that a gesture set of cardinality 27 is significantly larger than gesture sets of a handful of gestures, one would expect performance to degrade. Gesture recognition using an SVM assumes that hyperplane division of recognition space produces a set of n-dimensional Voronoi regions; therefore, the smaller the overall data set, the larger the target volumes and the less likely we are to observe gesture confusion during recognition, all other factors being equal.

However, we also believe that the presence of confusion in recognition presents an opportunity for insight and improved design. As a specific example of this, consider the confusion matrix shown in Figure 5.4. Our configuration struggles (slightly) with individual finger identification during the bend, tap, and pinch gestures *when analyzing all 27 gestures*. In particular, index-middle and ring-pinky confusion is observed. This confusion is a result of the close proximity of the carpel and ulnar tunnel on the palmar face of the wrist. Tomo5 and WristFlex, with their higher sensor density, can leverage carpel nerve/ulnar nerve cross-talk to better discriminate movement (i.e. ring finger movement, primarily ulnar, results in more carpel ligament activation than does pinky movement; likewise middle finger movement, primarily carpal, results in more ulnar ligament activation than does index finger movement). The implication: for finger identification, denser sensing on the palmar face of the wrist is necessary.

6.2 Future work

6.2.1 Classification Algorithm

During the post-analysis, we have tried to use SVM and Random-Forrest algorithms to classify gestures. They are used mainly because we are are following previous papers' way of processing data. In the future, we can try more sophisticated machine learning algorithms or neural networks to do classification. With the more sophisticated learning algorithms, it is possible that we can achieve even higher recognition accuracies.

6.2.2 More Realistic Testing Case

Due to the length of the study, we only tested accuracy without taking the device off. For a more realistic setting, we would test the accuracy after participants taking the device off and putting it back on. We would also ask the participants to come back after a few days, using the first day's data as training data and test the accuracy on the second day's data. We can also test cross user accuracy by using one participant's data to classify other participants' data.

6.2.3 Different Gesture Set

We have tested with the gesture set we obtained from the elicitation study with our device. However, it would also be interesting to see how the system would work with other existing gesture sets (e.g. American Sign Gestures).

6.3 Conclusion

In this work, we describe an exploration of the space of wrist and hand gestures. To accomplish this goal, we present results of an elicitation study and a synthesis of past work in wrist and hand gesture input. We also explore questions around IMU-based recognition and its sufficiency within the space of hand and wrist gestures. We re-implemented a state-of-the-art, IMU-based recognition strategy [63], and we verified that our implementation performed similarly (Figure 5.3). We built a wrist movement sensor out of $15 worth of materials, and we showed that there is complementary information available from wrist-movement. Observed recognition accuracies based upon our hardware argues for on-going research into additional configurations of wrist-worn sensors for hand and wrist gesture capture and should, we hope, encourage more extensive and comparative evaluation of wrist-worn gesture recognition systems.

References

[1] Anonymous. Personal Communication from Anonymous Reviewer, 2017.

[2] Inertial measurement unit. http://www.ssl.umd.edu/Projects/RangerNBV/thesis/2-4-1.htm, 1998. Online; accessed 5 November 2017.

[3] Flicktek. http://www.flicktek.com/, 2016. Online; accessed 5 March 2016.

[4] Myo. https://www.myo.com/, 2016. Online; accessed 29 March 2016.

[5] Android sensors. https://developer.android.com/guide/topics/sensors/sensors_motion.html, 2017. Online; accessed 5 November 2017.

[6] Electromyography wikipedia. https://en.wikipedia.org/wiki/Electromyography, 2017. Online; accessed 15 October 2017.

[7] Fitbit one sleep band. https://www.fitbit.com/en-ca/shop/accessories/one-sleep-band, 2017. Online; accessed 12 February 2017.

[8] Flexpoint bend sensors. http://www.flexpoint.com/, 2017. Online; accessed 22 October 2017.

[9] Leap motion. https://www.leapmotion.com/, 2017. Online; accessed 20 October 2017.

[10] Tests and procedures: Electromyography. https://www.mayoclinic.org/tests-procedures/emg/basics/definition/prc-20014183/, 2017. Online; accessed 15 October 2017.

[11] Proximity sensor wikipedia. https://en.wikipedia.org/wiki/Proximity_sensor, 2018. Online; accessed 15 Jan 2018.

[12] Roland Aigner, Daniel Wigdor, Hrvoje Benko, Michael Haller, David Lindbauer, Alexandra Ion, Shengdong Zhao, and Jeffrey Tzu Kwan Valino Koh. Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for hci. Technical report, November 2012.

[13] Daniel C Alexander and James C Gee. Elastic matching of diffusion tensor images. *Computer Vision and Image Understanding*, 77(2):233–250, 2000.

[14] Lisa Anthony and Jacob O. Wobbrock. $n-protractor: A fast and accurate multistroke recognizer. In *Proceedings of Graphics Interface 2012*, GI '12, pages 117–120, Toronto, Ont., Canada, Canada, 2012. Canadian Information Processing Society.

[15] Shaikh Shawon Arefin Shimon, Courtney Lutton, Zichun Xu, Sarah Morrison-Smith, Christina Boucher, and Jaime Ruiz. Exploring non-touchscreen gestures for smartwatches. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 3822–3833, New York, NY, USA, 2016. ACM.

[16] Gilles Bailly, Jörg Müller, Michael Rohs, Daniel Wigdor, and Sven Kratz. Shoesense: A new perspective on gestural interaction and wearable applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1239–1248, New York, NY, USA, 2012. ACM.

[17] Asker M Bazen and Sabih H Gerez. Fingerprint matching by thin-plate spline modelling of elastic deformations. *Pattern Recognition*, 36(8):1859–1867, 2003.

[18] Donald J Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, pages 359–370. Seattle, WA, 1994.

[19] Edwin Chan, Teddy Seyed, Wolfgang Stuerzlinger, Xing-Dong Yang, and Frank Maurer. User elicitation on single-hand microgestures. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 3403–3414, New York, NY, USA, 2016. ACM.

[20] Artem Dementyev and Joseph 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*, UIST '14, pages 161–166, New York, NY, USA, 2014. ACM.

[21] Haiwei Dong, Ali Danesh, Nadia Figueroa, and Abdulmotaleb El Saddik. An elicitation study on gesture preferences and memorability toward a practical hand-gesture vocabulary for smart televisions. *IEEE Access*, 3:543–555, 2015.

[22] A. Ferrone, X. Jiang, L. Maiolo, A. Pecora, L. Colace, and C. Menon. A fabric-based wearable band for hand gesture recognition based on filament strain sensors: A preliminary investigation. In *2016 IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT)*, pages 113–116, Nov 2016.

[23] Rui Fukui, Masahiko Watanabe, Tomoaki Gyota, Masamichi Shimosaka, and Tomomasa Sato. Hand shape classification with a wrist contour sensor: Development of a prototype device. In *Proceedings of the 13th International Conference on Ubiquitous Computing*, UbiComp '11, pages 311–314, New York, NY, USA, 2011. ACM.

[24] Jun Gong, Xing-Dong Yang, and Pourang Irani. Wristwhirl: One-handed continuous smartwatch input using wrist gestures. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, pages 861–872, New York, NY, USA, 2016. ACM.

[25] Saul Greenberg and Chester Fitchett. Phidgets: easy development of physical interfaces through physical widgets. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 209–218. ACM, 2001.

[26] Faizan Haque, Mathieu Nancel, and Daniel Vogel. Myopoint: Pointing and clicking using forearm mounted electromyography and inertial motion sensors. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 3653–3656, New York, NY, USA, 2015. ACM.

[27] Chris Harrison, Desney Tan, and Dan Morris. Skinput: Appropriating the body as an input surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 453–462, New York, NY, USA, 2010. ACM.

[28] Luke Hespanhol, Martin Tomitsch, Kazjon Grace, Anthony Collins, and Judy Kay. Investigating intuitiveness and effectiveness of gestures for free spatial interaction with large displays. In *Proceedings of the 2012 International Symposium on Pervasive Displays*, PerDis '12, pages 6:1–6:6, New York, NY, USA, 2012. ACM.

[29] P. G. Jung, G. Lim, S. Kim, and K. Kong. A wearable gesture recognition device for detecting muscular activities based on air-pressure sensors. *IEEE Transactions on Industrial Informatics*, 11(2):485–494, April 2015.

[30] Ankit Kamal, Yang Li, and Edward Lank. Teaching motion gestures via recognizer feedback. In *Proceedings of the 19th International Conference on Intelligent User Interfaces*, IUI '14, pages 73–82, New York, NY, USA, 2014. ACM.

[31] Keiko Katsuragawa, Ankit Kamal, and Edward Lank. Effect of motion-gesture recognizer error pattern on user workload and behavior. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, IUI '17, pages 439–449, New York, NY, USA, 2017. ACM.

[32] Keiko Katsuragawa, Krzysztof Pietroszek, James R. Wallace, and Edward Lank. Watchpoint: Freehand pointing with a smartwatch in a ubiquitous display environment. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '16, pages 128–135, New York, NY, USA, 2016. ACM.

[33] David Kim, Otmar Hilliges, Shahram Izadi, Alex D. Butler, Jiawen Chen, Iason Oikonomidis, and Patrick 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*, UIST '12, pages 167–176, New York, NY, USA, 2012. ACM.

[34] Jonghwa Kim, Stephan Mastnik, and Elisabeth André. Emg-based hand gesture recognition for realtime biosignal interfacing. In *Proceedings of the 13th International Conference on Intelligent User Interfaces*, IUI '08, pages 30–39, New York, NY, USA, 2008. ACM.

[35] Sven Kratz and Michael Rohs. A $3 gesture recognizer: simple gesture recognition for devices equipped with 3d acceleration sensors. In *Proceedings of the 15th international conference on Intelligent user interfaces*, pages 341–344. ACM, 2010.

[36] Gierad Laput, Robert Xiao, and Chris Harrison. Viband: High-fidelity bio-acoustic sensing using commodity smartwatch accelerometers. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, pages 321–333, New York, NY, USA, 2016. ACM.

[37] Yang Li. Protractor: A fast and accurate gesture recognizer. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 2169–2172, New York, NY, USA, 2010. ACM.

[38] Jhe-Wei Lin, Chiuan Wang, Yi Yao Huang, Kuan-Ting Chou, Hsuan-Yu Chen, Wei-Luan Tseng, and Mike Y. Chen. Backhand: Sensing hand gestures via back of the hand. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology*, UIST '15, pages 557–564, New York, NY, USA, 2015. ACM.

[39] Mingyu Liu, Mathieu Nancel, and Daniel Vogel. Gunslinger: Subtle arms-down mid-air interaction. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology*, UIST '15, pages 63–71, New York, NY, USA, 2015. ACM.

[40] Z. Lu, X. Chen, Q. Li, X. Zhang, and P. Zhou. A hand gesture recognition framework and wearable gesture-based interaction prototype for mobile devices. *IEEE Transactions on Human-Machine Systems*, 44(2):293–299, April 2014.

[41] Julio C. Mateo, Javier San Agustin, and John Paulin Hansen. Gaze beats mouse: Hands-free selection by combining gaze and emg. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '08, pages 3039–3044, New York, NY, USA, 2008. ACM.

[42] Jess McIntosh, Asier Marzo, Mike Fraser, and Carol Phillips. Echoflex: Hand gesture recognition using ultrasound imaging. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 1923–1934, New York, NY, USA, 2017. ACM.

[43] Jess McIntosh, Charlie McNeill, Mike Fraser, Frederic Kerber, Markus Löchtefeld, and Antonio 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*, CHI '16, pages 2332–2342, New York, NY, USA, 2016. ACM.

[44] Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, m. c. schraefel, and Jacob O. Wobbrock. Reducing legacy bias in gesture elicitation studies. *interactions*, 21(3):40–45, May 2014.

[45] Mehran Moshfeghi. Elastic matching of multimodality medical images. *CVGIP: Graphical Models and Image Processing*, 53(3):271–282, 1991.

[46] Prajwal Paudyal, Junghyo Lee, Ayan Banerjee, and Sandeep K.S. Gupta. Dyfav: Dynamic feature selection and voting for real-time recognition of fingerspelled alphabet using wearables. In *Proceedings of the 22Nd International Conference on Intelligent User Interfaces*, IUI '17, pages 457–467, New York, NY, USA, 2017. ACM.

[47] Krzysztof Pietroszek, Anastasia Kuzminykh, James R. Wallace, and Edward Lank. Smartcasting: A discount 3d interaction technique for public displays. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design*, OzCHI '14, pages 119–128, New York, NY, USA, 2014. ACM.

[48] Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. User-defined gestures for augmented reality. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 955–960, New York, NY, USA, 2013. ACM.

[49] Zhou Ren, Jingjing Meng, Junsong Yuan, and Zhengyou Zhang. Robust hand gesture recognition with kinect sensor. In *Proceedings of the 19th ACM International Conference on Multimedia*, MM '11, pages 759–760, New York, NY, USA, 2011. ACM.

[50] Julie Rico and Stephen Brewster. Gestures all around us: User differences in social acceptability perceptions of gesture based interfaces. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '09, pages 64:1–64:2, New York, NY, USA, 2009. ACM.

[51] Dean Rubine. Specifying gestures by example. In *Proceedings of the 18th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '91, pages 329–337, New York, NY, USA, 1991. ACM.

[52] Jaime Ruiz, Yang Li, and Edward Lank. User-defined motion gestures for mobile interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 197–206, New York, NY, USA, 2011. ACM.

[53] Jaime Ruiz and Daniel Vogel. Soft-constraints to reduce legacy and performance bias to elicit whole-body gestures with low arm fatigue. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3347–3350. ACM, 2015.

[54] T. Scott Saponas, Desney S. Tan, Dan Morris, Ravin Balakrishnan, Jim Turner, and James 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, pages 167–176, New York, NY, USA, 2009. ACM.

[55] T. Scott Saponas, Desney S. Tan, Dan Morris, Jim Turner, and James A. Landay. Making muscle-computer interfaces more practical. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 851–854, New York, NY, USA, 2010. ACM.

[56] Charles C. Tappert. Cursive script recognition by elastic matching. *IBM Journal of Research and development*, 26(6):765–771, 1982.

[57] Gineke A. ten Holt, Marcel J.T. Reinders, and Emile A. Hendriks. Multi-dimensional dynamic time warping for gesture recognition. In *Thirteenth annual conference of the Advanced School for Computing and Imaging*, June 13-15 2007.

[58] Robert Tidwell, Robert Akl, Sarath Akumalla, Sarada Karlaputi, David Struble, and Krishna Kavi. Evaluating the feasibility of emg and bend sensors for classifying hand gestures. In *Proceedings of the International Conference on Multimedia and Human Computer Interaction*, MHCI '13, pages 63:1–8, Toronto, ON, Canada, 2013. ASET.

[59] Radu-Daniel Vatavu and Jacob O. Wobbrock. Formalizing agreement analysis for elicitation studies: New measures, significance test, and toolkit. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1325–1334, New York, NY, USA, 2015. ACM.

[60] Radu-Daniel Vatavu and Ionut-Alexandru Zaiti. Leap gestures for tv: Insights from an elicitation study. In *Proceedings of the 2014 ACM International Conference on Interactive Experiences for TV and Online Video*, TVX '14, pages 131–138, New York, NY, USA, 2014. ACM.

[61] Daniel Vogel and Ravin Balakrishnan. Distant freehand pointing and clicking on very large, high resolution displays. In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology*, UIST '05, pages 33–42, New York, NY, USA, 2005. ACM.

[62] David Way and Joseph Paradiso. A usability user study concerning free-hand microgesture and wrist-worn sensors. In *Proceedings of the 2014 11th International Conference on Wearable and Implantable Body Sensor Networks*, BSN '14, pages 138–142, Washington, DC, USA, 2014. IEEE Computer Society.

[63] Hongyi Wen, Julian Ramos Rojas, and Anind 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*, CHI '16, pages 3847–3851, New York, NY, USA, 2016. ACM.

[64] Daniel Wigdor and Dennis Wixon. *Brave NUI world: designing natural user interfaces for touch and gesture*. Morgan Kaufmann, 2011.

[65] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1083–1092, New York, NY, USA, 2009. ACM.

[66] Chao Xu, Parth H. Pathak, and Prasant Mohapatra. Finger-writing with smartwatch: A case for finger and hand gesture recognition using smartwatch. In *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*, HotMobile '15, pages 9–14, New York, NY, USA, 2015. ACM.

[67] Yoonsik Yang, Seungho Chae, Jinwook Shim, and Tack-Don Han. Emg sensor-based two-hand smart watch interaction. In *Adjunct Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, UIST '15 Adjunct, pages 73–74, New York, NY, USA, 2015. ACM.

[68] Yang Zhang and Chris Harrison. Tomo: Wearable, low-cost electrical impedance tomography for hand gesture recognition. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology*, UIST '15, pages 167–173, New York, NY, USA, 2015. ACM.

[69] Yang Zhang, Robert Xiao, and Chris 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, pages 843–850, New York, NY, USA, 2016. ACM.