Advancing Multimodal Sensing and Robotic Interfaces for Chronic Care

Akhil Padmanabha

CMU-RI-TR-25-64

July 2025

The Robotics Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA

Thesis Committee:
Zackory Erickson, *co-chair*
Carmel Majidi, *co-chair*
Mayank Goel
Tapomayukh Bhattacharjee (*Cornell University*)
Arash Mostaghimi (*Brigham & Women's Hospital*)

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Robotics.

Copyright © 2025 Akhil Padmanabha. All rights reserved.

Abstract

The healthcare system prioritizes reactive care for acute illnesses, often overlooking the ongoing needs of individuals with chronic conditions that require long-term management and personalized care. Addressing this gap through technology can empower patients to better manage their conditions, greatly enhancing quality of life and independence. Multimodal sensing, incorporating inertial, acoustic, and vision-based sensors, within mobile form factors like wearables and probes, has the ability to enable continuous, quantitative monitoring of physiological and behavioral changes, while also serving as interfaces for individuals to manage and control aspects of their care. Building on this concept, this thesis introduces sensing technologies, devices, and algorithms aimed at improving the management of chronic conditions.

Focusing initially on dermatological conditions, particularly chronic itch diseases such as eczema, we present acousto-mechanic wearable sensing hardware and machine learning models for monitoring of scratching behavior. Next, we present work in 3D reconstruction of the skin surface using GelSight tactile sensing integrated into a probe, offering a validated tool for skin surface analysis and wrinkle depth estimation, with potential applications in diagnosis and treatment monitoring.

Extending beyond dermatology, we develop active, shared, and passive control interfaces for individuals with severe motor impairments to use caregiving robots. In the space of active control, we present HAT, a head-worn inertial device for robot teleoperation. For shared control, we refine HAT to blend autonomy with user intent, validated in a 7-day in-home deployment with a non-speaking individual with quadriplegia. We also introduce VoicePilot, a speech-based interface powered by large language models, supporting flexible and natural robot control and validated with older adults. Lastly, for passive control, we present WAFFLE, a wearable system that uses inertial and throat microphone data to estimate bite timing in robot-assisted feeding, generalizing across users, robot types, and dining contexts.

Together, these systems advance the state of the art in multimodal sensing and robotic interfaces for chronic care, contributing novel algorithms, validated hardware, and empirical insights that deepen our understanding of how portable, multimodal technologies can support long-term health management and promote human autonomy.

Acknowledgments

None of the work presented in this dissertation would have been possible without my advisors, colleagues, collaborators, mentors, friends, family, and the Pittsburgh community.

First, I would like to thank my co-advisors, Professor Zackory Erickson and Professor Carmel Majidi, who have been a huge part of my journey. Throughout my PhD, both of them have been very open to me pursuing any project ideas and this allowed me to work on problems that were so important to me and motivated by my own health challenges. Their willingness to collaborate additionally allowed me to form new collaborations and gain insight from numerous others including medical professionals, individuals with motor impairments, senior living facilities, and companies. I would also like to thank my three other committee members, Professor Tapomayukh Bhattacharjee and Dr. Arash Mostaghimi who I had the great opportunity to learn from and collaborate with on projects included in this thesis, and Professor Mayank Goel who provided me with feedback through the proposal and defense process and who I additionally took a class with.

Next, I would like to thank Dr. Sonal Choudhary and Dr. Arash Mostaghimi for their mentorship and guidance, leading to wonderful collaborations on two projects related to dermatological sensing. In conjunction with Zackory and Carmel, they trusted and encouraged me to bring my expertise with sensing systems in addition to my lived experience as a patient, into advancing dermatology. Their strong conviction for enabling better treatment outcomes through technology gives me huge faith in the progression of medical research and science. I would also like to acknowledge Professor Henny Admoni, Professor Taskin Padir, and Professor Wenzhen Yuan who I had the opportunity to collaborate with and get crucial advice from on various projects.

I would like to thank everyone from my two labs, the Robotic Caregiving and Human Interaction Lab (RCHI) and the Soft Machines Lab (SML). It has especially been an honor to play a role in building up the RCHI Lab with Zackory and other labmates and to see it grow and flourish over the last 4 years. Both labs have additionally been a huge support for me, with many selfless labmates being extremely willing to collaborate, teach, and train me with equipment and tools throughout the PhD.

Over the last 4 years, I have had the pleasure of mentoring and collaborating with many great undergraduate students: Jessie Yuan, Janavi Gupta, Qin (Alicia) Wang, Catherine Li, Tanisha Mehta, Chen Chen, Wesley Lewis, Alexandra Gillespie, Victoria De León, and Daphne Han. Seeing them grow as students and researchers has been one of the most heartwarming and exciting parts of my PhD journey. I've also had the privilege of collaborating with many talented and amazing masters, PhD students, and postdocs including Arpit Agarwal, Rajat Kumar Jenamani, Amal Nanavati, Jehan Yang, Kavya Puthuveetil, Nathaniel Hanson, Dinesh K. Patel, Anthony Wertz, Austin Williams, Fukang Liu, Jashkumar Diyora, Sankalp Chopkar, Achu Wilson, Yexin Hu, and Saksham Bhutani. I am very grateful for my collaboration with Hello Robot, specifically Vy Nguyen and Vinitha Ranganeni, and Henry and Jane Evans for warmly welcoming us into their home for the week-long human study. I would like to also thank all my study participants over the years.

I've been blessed with an amazing support system from afar during my PhD. To my parents and sister, thank you for your unwavering love and encouragement throughout this journey. To my friends from back home, from Berkeley, and from past internships, your support through calls, adventures and trips, and countless hangouts have been invaluable to me. I am deeply grateful to the teachers, mentors, and advisors from my high school years, during a time when I was extremely ill, who believed and saw potential in me, encouraged me, and helped set me on this journey. I am also extremely thankful to my undergraduate labs, ViaBot, SpaceX, and Meta teams who gave me opportunities to learn, grow, and explore robotics and sensing in so many forms, and who helped me build the confidence that continues to guide me today.

Finally, to the Pittsburgh community that made this city feel like home: my housemates and neighbors, friends from the Robotics Institute and across CMU, everyone who gathered and connected at our 621 Rager events, the Liberty basketball crew, and so many others, thank you for the laughter, adventures, and unforgettable memories. There are far too many of you to name, but the moments we shared and the community we built will stay with me forever.

Funding

This research was supported by the National Science Foundation Graduate Research Fellowship Program under Grant No.DGE1745016 and DGE2140739.

Contents

1 Introduction **1**
1.1 Thesis Overview . . . 2
1.2 Part 1 Overview . . . 4
1.3 Part 2 Overview . . . 5

I Sensing for Dermatological Conditions **7**

2 A Multimodal Sensing Ring for Quantification of Scratch Intensity **11**
2.1 Contributions . . . 11
2.2 Introduction . . . 11
2.3 Methods . . . 15
2.3.1 Device Design . . . 15
2.3.2 Scratch Intensity . . . 16
2.3.3 Scratch Detection . . . 21
2.3.4 Statistics and Reproducibility . . . 22
2.4 Results . . . 23
2.4.1 Scratch Intensity Dataset and Feature Extraction . . . 23
2.4.2 Scratch Intensity Model Performance . . . 24
2.4.3 Scratch Intensity Validation . . . 25
2.4.4 Scratch Detection Dataset and Feature Extraction . . . 27
2.4.5 Scratch Detection Model Performance . . . 28
2.5 Discussion . . . 28

3 In-Vivo Skin 3-D Surface Reconstruction and Wrinkle Depth Estimation using Handheld High Resolution Tactile Sensing **37**
3.1 Contributions . . . 37
3.2 Introduction . . . 38
3.3 Experimental Section . . . 41
3.3.1 Sensor . . . 41
3.3.2 Gel Design . . . 42
3.3.3 Gel Mechanical Characterization . . . 42
3.3.4 Calibration Indenting Rig . . . 43

3.3.5 Convolutional Neural Network (CNN) for estimation of surface normals . . . 45
3.3.6 Poisson Integration and Detrending . . . 46
3.3.7 Validation and Test Objects . . . 47
3.3.8 Wrinkle Detection and Height Estimation Algorithms . . . 48
3.3.9 Human Study Procedure . . . 49
3.4 Results . . . 50
3.4.1 Gel Elasticity . . . 51
3.4.2 Algorithm Performance . . . 52
3.4.3 Characterization of Skin Wrinkles . . . 54
3.4.4 Effect of Moisturizer on Wrinkle Depth . . . 56
3.5 Discussion . . . 58

II Interfaces for Control of Physically Assistive Robots 61

4 HAT: Head-Worn Assistive Teleoperation of Mobile Manipulators 65
4.1 Contributions . . . 65
4.2 Introduction . . . 65
4.3 Related Work . . . 68
4.4 Device Design . . . 69
4.4.1 Mode Switching . . . 69
4.4.2 Calibration and Robot Mapping . . . 69
4.5 Study Design . . . 72
4.5.1 Participants with Motor Impairments . . . 75
4.6 Results and Discussion . . . 75
4.6.1 Task Completion and Efficiency . . . 75
4.6.2 Errors and Recovery . . . 77
4.6.3 Ease of Use . . . 78
4.6.4 Learning Curve . . . 79
4.6.5 Workload . . . 81
4.6.6 Challenges and Future Directions . . . 81
4.7 Conclusions . . . 82

5 Independence in the Home: A Wearable Interface for a Person with Quadriplegia to Teleoperate a Mobile Manipulator 83
5.1 Contributions . . . 83
5.2 Introduction . . . 84
5.3 Related Work . . . 85
5.3.1 Teleoperation Interfaces . . . 85
5.3.2 IMU-based Interfaces . . . 86

5.3.3 Shared Control . . . 86
5.4 Methodology . . . 87
5.4.1 Device Design . . . 87
5.4.2 Mode Switching . . . 88
5.4.3 Robot Mapping and Thresholds . . . 88
5.4.4 Cursor Control . . . 90
5.4.5 Graphical User Interface . . . 90
5.4.6 Driver Assistance . . . 91
5.5 Study Design . . . 93
5.6 Results and Discussion . . . 94
5.6.1 Metrics . . . 95
5.6.2 Effect of Driver Assistance . . . 99
5.6.3 Out of Line-of-Sight Tasks . . . 100
5.6.4 Control of Multiple Devices . . . 101
5.6.5 Lessons Learned . . . 101

6 VoicePilot: Harnessing LLMs as Speech Interfaces for Physically Assistive Robots 105
6.1 Contributions . . . 105
6.2 Introduction . . . 105
6.3 Related Work . . . 108
6.3.1 LLMs for Robotics . . . 108
6.3.2 Robot-Assisted Feeding . . . 110
6.4 Version 1: Framework from Literature . . . 111
6.4.1 Framework Development . . . 111
6.4.2 Implementation . . . 113
6.4.3 Pilot Testing with Lab Members . . . 114
6.5 Version 2: Improved Accuracy and Customizability . . . 114
6.5.1 Framework Iteration . . . 114
6.5.2 Implementation . . . 115
6.5.3 Demonstration with Community Members . . . 116
6.6 Version 3: Increased Ease of Use . . . 117
6.6.1 Framework Iteration . . . 117
6.6.2 Implementation . . . 118
6.6.3 Testing with Older Adults . . . 119
6.7 Reflection . . . 122
6.7.1 Final Framework . . . 122
6.7.2 Analysis . . . 122
6.7.3 Design Guidelines . . . 123
6.7.4 Considerations and Future Work . . . 130

7 A Wearable Approach to Bite Timing Estimation in Robot-Assisted Feeding **133**
7.1 Contributions . . . 133
7.2 Introduction . . . 135
7.3 Related Work . . . 137
7.3.1 Robot-Assisted Feeding . . . 137
7.3.2 Wearable Sensing . . . 138
7.4 Methods . . . 138
7.4.1 Phase 1: Data Collection Study . . . 140
7.4.2 Bite Timing Model . . . 142
7.4.3 Assertiveness Thresholding . . . 144
7.4.4 Phase 2: Evaluation with Individuals without Impairments . . 146
7.4.5 Phase 3: Qualitative Evaluation with Individual with Motor Impairments . . . 149
7.5 Results and Discussion . . . 150
7.5.1 Model Performance using LOSO-CV . . . 150
7.5.2 Bite Timing Appropriateness . . . 151
7.5.3 Likert Items . . . 154
7.5.4 Preference Ranking . . . 158
7.5.5 Similarity to Caregiver . . . 158
7.5.6 Adoption of the Wearable Interface and Future Work . . . 160
7.6 Conclusion . . . 161

8 Conclusion **163**
8.1 Summary . . . 163
8.2 Future Work . . . 164

Bibliography **167**

When this dissertation is viewed as a PDF, the page header is a link to this Table of Contents.

List of Figures

1.1 Thesis Overview 2

2.1 Proposed multimodal wearable ring and raw data during scratching. . 13
2.2 Scratch Intensity and Detection Pipeline. 14
2.3 Scratch power labeling using a pressure sensitive tablet. 17
2.4 Scratch intensity dataset, features, and performance using leave one subject out cross validation. 33
2.5 Scratch intensity validation performance. 34
2.6 Scratch detection dataset and performance using leave one subject out cross validation. 35

3.1 Skin 3-D Reconstruction Pipeline Overview 40
3.2 Training and Deployment Pipelines 41
3.3 Gel Characterization and Calibration 41
3.4 Evaluation Pipeline 52
3.5 Sensor Accuracy and Precision on Validation and Test Objects 53
3.6 3-D Reconstruction, Wrinkle Detection, and Wrinkle Depth Estimation in Various Locations on the Human Body 55
3.7 Effect of Moisturizer on Wrinkle Depth 57

4.1 The head-worn interface (HAT) with integrated inertial measurement unit sensing. 67
4.2 Mapping from the head-worn interface to the robot. 70
4.3 Time lapse of a participant dragging a blanket off his legs with corresponding signals from the IMU along the pitch and roll axes. 71
4.4 Setup of the 4 study tasks: Cup, Trash, Blanket, and Cleaning. . . . 73
4.5 Average task times from 16 healthy participants, 2 participants with impairments, and 1 expert user. 76
4.6 Two participants with impairments completing study tasks. 77
4.7 Median results from two 7-point Likert items on task efficiency and errors. 79
4.8 Results from the 7-point Likert items presented to participants after completion of all tasks. 80
4.9 Median results from the NASA TLX survey. 80

5.1 System overview of HAT, a head-worn assistive interface. 84
5.2 Robot control modes and mapping. 89
5.3 A subset of the tasks conducted during the study. 89
5.4 Task completion times and resets for predefined tasks. 91
5.5 7-point Likert item responses for each day of the study. 96
5.6 7-point NASA TLX responses for each day of the study. 97
5.7 Driver Assistance Results. 98

6.1 Project overview: framework, speech interface, and design guidelines. 106
6.2 Obi robot and setup for study with 11 older adults at an independent living facility. 108
6.3 Final framework. 109
6.4 Participants from the Version 3 human study at an independent living facility. 116
6.5 Participant self-reported success per task for each attempt. 121
6.6 Responses from six 7-point Likert Items answered by the 11 participants at the end of the study. 122
6.7 Responses to a 0-100 NASA TLX Scale assessed at the end of the study with all 11 participants. 124
6.8 Design Guidelines. 124
6.9 A subset of quotes from our study involving 11 older adults. 127

7.1 WAFFLE Overview 134
7.2 Study Setup for the Three Phases 139
7.3 WAFFLE Deployment Pipeline 142
7.4 Phase 2 Participant Responses on Bite Timing 154
7.5 Phase 2 Individual Dining Mean Likert Ratings 155
7.6 Phase 2 Social Dining Mean Likert Ratings 157
7.7 Phase 2 Preference Rankings 159

Chapter 1

Introduction

The healthcare system is primarily structured to deliver reactive care, focusing on treating acute illnesses as they arise [209, 217, 227]. While this approach is effective for addressing immediate health concerns, it often neglects the complex, ongoing needs of individuals with chronic conditions [133, 228]. Chronic conditions differ from acute illnesses in that they lack a clear resolution and typically require long-term management. The unpredictable nature of chronic diseases adds to the difficulty of managing them, as symptoms can vary widely and change over time, complicating both diagnosis and treatment [231]. Many chronic conditions also impact the ability to perform activities of daily living (ADLs) such as mobility, self-care, and basic functioning, significantly reducing independence and quality of life [13, 57, 195]. For individuals facing these challenges, everyday tasks can become overwhelming or unachievable, adding to the emotional and physical burden of their condition. Despite the significant impact of chronic diseases, there remains a notable gap in the development of tools and technology for continuous monitoring and personalized care. Multimodal sensing, encompassing inertial, acoustic, and vision-based sensors, within compact, mobile devices like wearables and handheld probes, holds significant potential in healthcare [19, 81, 157]. This thesis presents multimodal sensing techniques and algorithms designed to manage chronic conditions, including dermatological conditions and motor impairments.

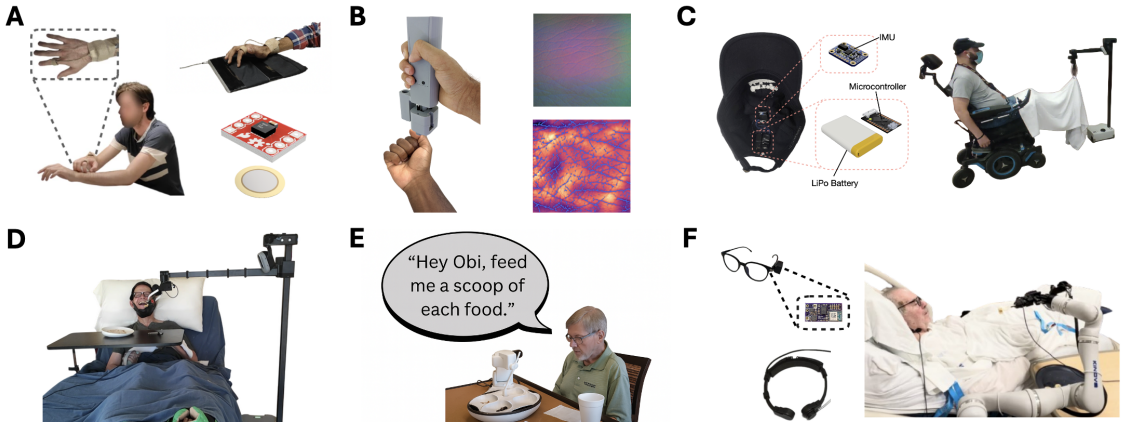


Figure 1.1: Thesis Overview. **A.** A multimodal ring for tracking scratching in patients in chronic itch, presented in Chapter 2. **B.** Our GelSight sensing probe for 3D skin surface reconstruction and wrinkle depth estimation, presented in Chapter 3. **C.** An introduction to HAT (Head-worn assistive teleoperation), a wearable interface for control of mobile manipulators for individuals with severe motor impairments, highlighted in Chapter 4. **D.** A 7-day in-home study using the HAT interface with Henry Evans, an individual with quadriplegia, presented in Chapter 5. **E.** Our VoicePilot framework for harnessing LLMs as speech interfaces for assistive robots with testing with older adults, detailed in Chapter 6. **F.** A passive sensing multimodal wearable interface for bite timing estimation in robot-assisted feeding generalizable across users and various settings, presented in Chapter 7.

1.1 Thesis Overview

This thesis consists of two parts. In Part I, we highlight two projects focused on using multimodal, compact sensing to monitor dermatological conditions, which are highly prevalent and often chronic. Among these conditions, itch, a debilitating symptom commonly linked to atopic dermatitis (eczema) and psoriasis, is particularly widespread with eczema alone affecting over 31 million people in the United States [84]. Itching frequently leads to scratching behaviors that disrupt sleep, reduce productivity, and negatively affect mood and overall well-being [34, 36, 117, 118, 159, 161, 181, 198, 204]. Despite its significant impact, there are currently no consumer devices that provide quantitative, continuous monitoring of scratching behaviors. We present advancements in developing such a system which could improve treatment evaluation, help identify triggers, and ultimately lead to better patient outcomes. Beyond

behavioral monitoring, tracking surface-level skin changes could also offer value to both patients and their clinicians. In this part, we additionally explore multimodal sensing for three-dimensional (3-D) reconstruction of the skin surface and estimation of wrinkle depth. This technology holds promise in dermatology for diagnosing and tracking textural changes in conditions such as eczema, psoriasis, and scleroderma, as well as assessing treatment effectiveness.

In Part II, we highlight four projects on using multimodal, compact sensing for developing interfaces for caregiving robots for individuals with severe motor impairments. Motor impairments can restrict individuals from performing ADLs such as eating, self-hygiene, and dressing, forcing many to rely on caregivers and limiting their independence. In the United States, approximately 1.7% of the population, over 5 million people, live with some form of paralysis, often resulting from neurodegenerative diseases, stroke, muscle atrophy, or spinal cord injury [18]. For individuals with tetraplegia or limited hand function, technologies that facilitate the performance of everyday tasks can have an immediate and profound impact, increasing independence and comfort while reducing caregiver workload [32, 41, 135, 164, 210]. Caregiving robots offer a potential solution, allowing these individuals to regain the ability to carry out physical tasks such as self-care and fetching objects, thereby improving their autonomy and quality of life [37, 40, 86, 116, 238]. Most interfaces have physical requirements, determining which subset of individuals with impairments will use them. The development and evaluation of novel assistive interfaces for mobile manipulators could lead to alternatives for individuals with impairments who may have a difficult time accessing traditional systems. We focus on the design and integration of wearable interfaces for active, shared, and passive control of assistive robotics. Active control refers to situations where the user's intentional actions or movements directly trigger robotic movements. This includes direct teleoperation interfaces, such as web-based controls or physical buttons that manually trigger sequences of robot actions like having a robot deliver a spoonful of food in robot-assisted feeding. We explore active control through an inertial-based interface named HAT and through a speech based interface named VoicePilot. While active control offers the user the highest degree of control, it results in high user workload. To address this, shared control can combine user input from direct teleoperation interfaces with the robot's autonomy, leveraging the robot's understanding of human intentions and the surrounding environment and

resulting in a reduction in cognitive load for the user. We explore incorporating a shared control algorithm named Driver Assistance to wearable interfaces like HAT for control of caregiving robots in the home. In comparison to active or shared control, passive control using wearables could enable robots to detect implicit signals and adjust their behavior accordingly, thereby reducing the need for frequent user input during repetitive tasks and lowering the user's overall workload. In the last project, we explore using a passive sensing multimodal wearable interface for bite timing estimation in robot-assisted feeding.

In summary, these projects advance chronic care by leveraging innovative sensing techniques and algorithms for dermatological monitoring and control of caregiving robotics. The technologies, methods, and insights discovered can be applied to various other acute and chronic conditions and fields including AR/VR applications.

1.2 Part 1 Overview

In Chapter 2, we present a framework for quantifying scratch intensity and detecting scratching, using a wearable ring equipped with a 3-axis accelerometer and a vibration contact microphone to capture acousto-mechanic features of scratching. We develop a machine learning pipeline processes sensor data in real-time, achieving low error in scratch intensity regression on a 0-600 mW scale and high performance in scratch detection with 89.98% accuracy. We further validate our algorithms with additional participants, demonstrating clinically relevant intensity discrimination on a 0-10 scale.

In Chapter 3, we further explore skin monitoring by introducing 3-D skin reconstruction methodology using GelSight tactile sensing. We design highly elastic gels and develop learning-based reconstruction algorithms that convert GelSight images into micron-level 3-D height maps, enabling precise wrinkle height estimation across the body. Our compact, handheld system is integrated with force sensing for consistent contact and validated across multiple participants and skin locations. We demonstrate its ability to detect treatment effects and provide the first quantitative wrinkle depth measurements in many body areas, showing promise for clinical and cosmetic dermatological monitoring.

1.3 Part 2 Overview

In this space of active and shared control, in Chapter 4, we introduce a wearable interface named HAT (Head-worn Assistive Teleoperation) for users with severe motor impairments to directly control a robot via head movements. Developing wearable interfaces presents unique challenges, including ensuring precise sensing and intuitive mapping of human movements to robot actions, minimizing latency, and accommodating the needs of users with varying levels of impairments. We initially evaluated HAT in 2 hour studies in a lab environment with both non-disabled and individuals with impairments and showed that it is both intuitive, and efficient, allowing users to perform a variety of physical self-care and household tasks with few errors. We subsequently iterated on the interface and system design using both quantitative and qualitative data from the initial study in conjunction with input from Henry Evans, a non-speaking individual with quadriplegia who has participated extensively in assistive robotics studies [8, 37, 40, 79, 86, 106, 158]. This process culminated in a 7-day evaluation study with Henry in his home setting, detailed in Chapter 5. As part of this work, we also introduced and evaluated a shared control algorithm named Driver Assistance for interfaces for control of caregiving robots in the home. Our results showed that HAT is a strong teleoperation interface for caregiving robots in the home across key metrics including efficiency, errors, learning curve, and workload.

In Chapter 6, we further explore active control interfaces by introducing a human-centered framework, VoicePilot, for harnessing large language models (LLMs) as speech interfaces for physically assistive robots. Our framework was iteratively constructed through three stages of testing with a feeding robot, ensuring adaptability and usability for individuals with physical limitations. The final evaluation, conducted with 11 older adults, combined quantitative and qualitative findings to validate the framework and establish design guidelines for using LLMs as speech interfaces in assistive robotics. Our framework and design guidelines will enable researchers, engineers, and designers to develop robust and human-centered LLM-based speech interfaces for assistive robots.

Lastly, in Chapter 7, we explore passive control interfaces through the challenge of predicting bite timing in robot-assisted feeding in both individual and social dining

scenarios. We introduce WAFFLE (Wearable Approach For Feeding with LEarned Bite Timing), a system that leverages a throat microphone and glasses-mounted IMU to predict bite timing using natural, implicit cues such as head movements, chewing, and talking. Unlike prior methods that rely on vision and microphone arrays placed on the table, WAFFLE is wearable, privacy-preserving, and generalizes across robot types, positions, and users. We develop a supervised regression algorithm trained on data from 14 individuals and evaluate it in a human study with 15 participants without motor impairments and one participant with impairments. WAFFLE is shown to outperform baseline approaches, achieving high perceived control and low workload, and is preferred by the majority of users.

Part I

Sensing for Dermatological Conditions

Pruritus, commonly known as itch, is a debilitating symptom caused by various conditions, including atopic dermatitis, psoriasis, and liver disease. Itch often leads to scratching behaviors, which negatively impact sleep, productivity, mood, and overall well-being of patients. In Chapter 2, we introduce a multimodal sensing ring and machine learning algorithms for detection of scratching and estimation of scratch intensity. Unlike prior wearable systems that focus primarily on binary scratch detection during sleep, our approach captures scratch intensity using accelerometer and contact microphone data in addition to scratch detection. By providing continuous and quantitative measurements of itch, this approach could enhance the evaluation of treatment efficacy and help identify itch triggers, ultimately improving patient outcomes.

Beyond monitoring scratching behaviors, tracking the skin's surface offers an additional, underutilized window into dermatological health. In Chapter 3, we introduce a novel system for in-vivo 3-D reconstruction and quantitative wrinkle depth estimation using a compact GelSight tactile sensor and custom-designed elastic gels. This system enables micron-level accuracy in detecting and quantifying fine skin textures across diverse body locations, overcoming key limitations of existing commercial and research-grade tools that are often bulky, rigid, or lack sufficient depth resolution. We present a learning-based 3-D reconstruction pipeline validated on wrinkle-like test pieces and evaluated on human participants. Our work is the first to report wrinkle depth metrics across a wide range of body sites. Additionally, we show statistically significant reductions in wrinkle depth following moisturizer application, suggesting that our probe can be used to evaluate the effectiveness of skincare interventions.

The work presented in the following chapters is derived from the following published journal paper and manuscript in preparation:

- Akhil Padmanabha, Sonal Choudhary, Carmel Majidi, and Zackory Erickson. A multimodal sensing ring for quantification of scratch intensity. *Communications Medicine,* 3(1):115, 2023. [166]
- Akhil Padmanabha, Arpit Agarwal, Catherine Li, Austin Williams, Dinesh K. Patel, Sankalp Chopkar, Achu Wilson, Wenzhen Yuan, Sonal Choudhary, Arash Mostaghimi, Zackory Erickson, Carmel Majidi. In-Vivo Skin 3-D Surface

Reconstruction and Wrinkle Depth Estimation using Handheld High Resolution Tactile Sensing. *Manuscript in Preparation*, 2025

Chapter 2

A Multimodal Sensing Ring for Quantification of Scratch Intensity

2.1 Contributions

For the following chapter, the contributions are as follows. I formulated the research idea and goals with my advisors, Zackory Erickson and Carmel Majidi. I designed, built, and programmed the wearable device and wrote the data collection scripts. I developed and validated the classification and regression algorithms. I wrote the IRB protocol, collected data during both human studies, and analyzed the data. I created the figures and wrote the paper draft. All authors edited the paper. Dr. Sonal Choudhary provided advice on the medical implications of the work and study design. Carmel Majidi and Zackory Erickson were responsible for advising on all aspects of the project.

2.2 Introduction

Pruritus, commonly known as itch, is a debilitating symptom that can be caused by several conditions including atopic dermatitis, psoriasis, and liver disease, to name a few. Atopic dermatitis alone affects over 31 million people in the United States [84]. Pruritus leads to scratching behaviors, affecting the sleep, productivity, mood, and

overall well-being of individuals [34, 36, 117, 118, 159, 161, 181, 198, 204]. The tracking of itch is important for determining treatment efficacy and enabling better patient outcomes. In practice, itch is quantified using uni-dimensional patient-reported scales such as the Visual Analogue Scale (VAS), a 0-10 continuous measure, and the Numerical Rating Scale (NRS), a 0-10 discrete measure. While simple and quick to assess, these tools are subjective in nature and affected by anchoring and context bias, recall bias, and response shifts [60, 150]. Moreover, they only provide a single data point at the time of measurement, and thus, are unable to provide a comprehensive understanding of the effect of the symptom on a patient. Multidimensional scales such as the 5-D Itch Scale and Itchy Quality of Life (ItchyQoL) provide more insight by tracking additional metrics such as degree (intensity of itching) and emotions, but suffer from the same downsides as the uni-dimensional scales and are additionally time consuming to assess [50, 58, 60, 123].

Objective measures of pruritus focus on binary classification of scratching behaviors, which indicate whether a patient is scratching or not [235]. Tracking of these scratching behaviors provides insight into the effect of itch on an individual over time. Video recording using cameras (sometimes infrared) is the current standard in many studies [183, 235]. However, this method is time intensive as it requires manual labeling of scratching events by researchers. This method can additionally suffer from blind spots, has privacy concerns, and is only possible while a patient is sleeping. Wearable devices have shown promise for automatic, objective classification of scratching at night [39, 67, 103, 125, 138, 149, 160, 235]. Yet, advancements are needed to develop devices that perform well in daytime when participants still itch and when there are many confounding movements.

Furthermore, while existing wearable devices have the ability to record when a patient is scratching, the vast majority are unable to track scratch intensity. Scratch intensity can be related to itch intensity, which is tracked by doctors and patients using scales and questionnaires such as the Peak Pruritus Numerical Rating Scale (PPNRS) and the 5-D Itch Scale [58, 240]. Scratch intensity is one of the factors that determines damage to skin and holds additional insight into the degree of disruption that itching has on everyday life and sleep. A wearable device, placed on the back of the hand, has previously been introduced to estimate scratching force during sleep [236]. In contrast to prior work, our system stands out due to its

implementation of a multimodal sensing strategy with accelerometer and contact microphone data and our use of a pressure-sensitive tablet for training and validation, enabling us to attain statistically significant differentiation among various scratch intensity levels.

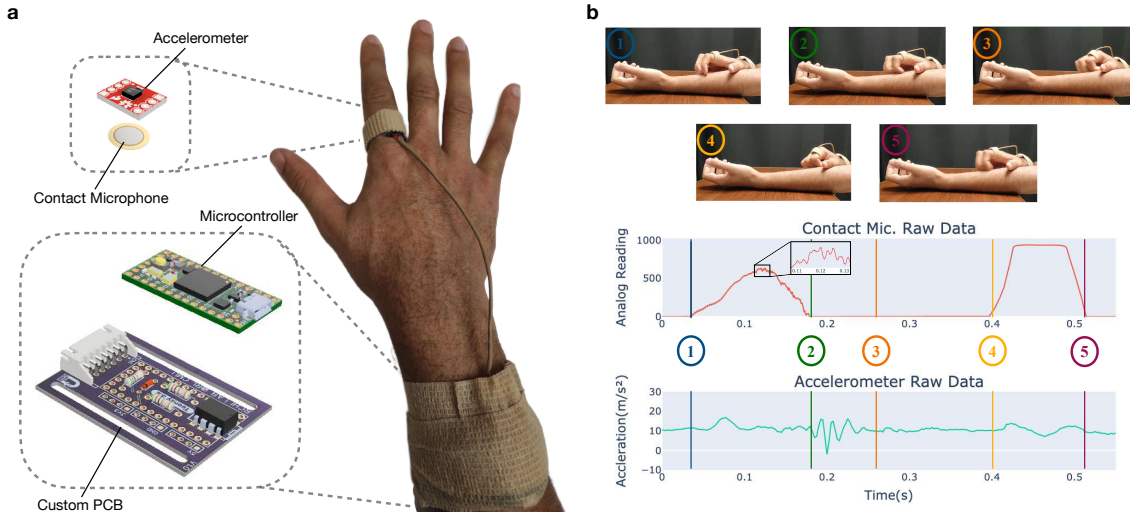


Figure 2.1: Proposed multimodal wearable ring and raw data during scratching. a. Components of the proposed multimodal wearable device. Right: human hand with wearable device donned. Top left: Rendering of the sensor suite inside of the ring consisting of an accelerometer and contact microphone. Bottom left: Rendering of the custom printed circuit board with an operational amplifier for the contact microphone, a microcontroller, and other supporting electronics. b. Clips from a finger movement only scratching motion are shown with corresponding raw data from the contact microphone and accelerometer (z-axis). 1-3: Fingers move up forearm. 4-5: Fingers move back down arm.

In this work, we introduce a framework for quantification of scratch intensity, a root cause of damage to skin due to chronic itch, and detection of scratching. Our framework consists of a wearable ring, shown in Fig. 2.1a, and multimodal algorithmic methods for real time analysis of scratch intensity and scratch detection. The multimodal ring consists of a 3-axis accelerometer and a vibration contact microphone, which provides complementary acousto-mechanic features on the scratching motion as shown in Fig. 2.1b. In conjunction, these sensors capture low frequency finger and arm

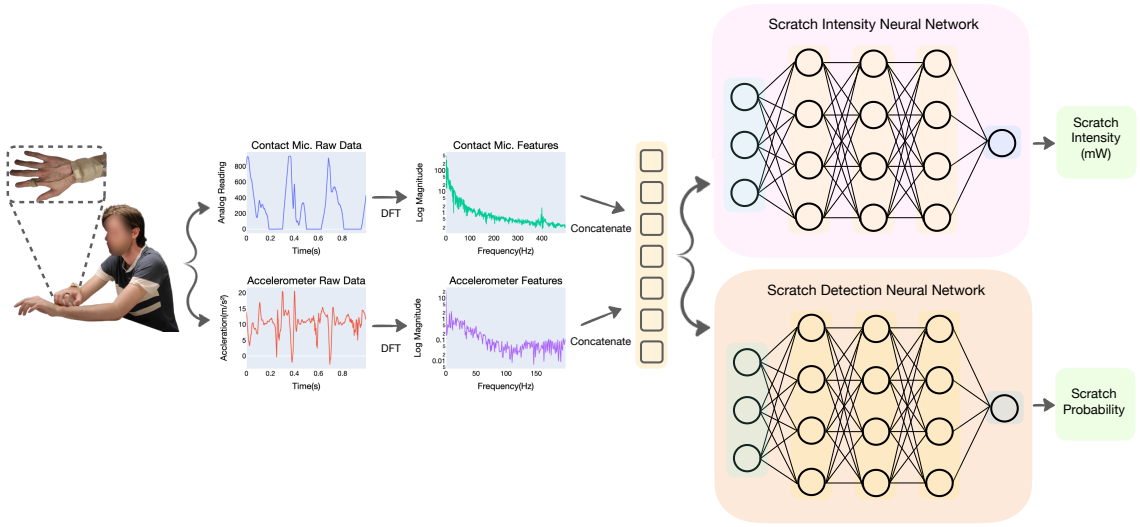


Figure 2.2: Scratch Intensity and Detection Pipeline. The Discrete Fourier Transform (DFT) is used to extract frequency features from the raw contact microphone and accelerometer data. The features are concatenated, normalized, and input into two neural networks, which independently regress to an estimate of scratch intensity and a scratch behavior probability. Permission was obtained from the participant for use of their image in publication.

motions, and high frequency vibrations propagating from the fingernail to the ring. For each second of raw sensor data, our processing and prediction pipeline, shown in Fig. 2.2, extracts frequency features and uses two machine learning algorithms for estimation of scratch intensity and detection in real-time. For training of our scratch intensity algorithm, we present methods for automatic generation of intensity labels using a touch sensitive surface embedded with an array of pressure sensors i.e. a pressure sensitive tablet. We assess the performance of our data-driven methods using Leave One Subject Out (LOSO) Cross Validation (CV) on data collected from 20 healthy participants, demonstrating the ability to accurately estimate scratch intensity on our proposed 0-600 mW power scale with a mean absolute error of 49.71 mW. Moreover, we highlight the advantages of multimodal sensing for scratch detection, achieving 89.98% accuracy using the combined accelerometer and contact microphone model in comparison to 86.24% and 79.98% accuracy for the accelerometer only and

contact microphone only models respectively. Finally, we extend the evaluation of our scratch intensity algorithm to include 14 new participants, achieving clinically relevant discrimination of intensity levels on a 0-10 scale with a mean absolute error of 1.37 units.

2.3 Methods

2.3.1 Device Design

The ring device consists of dual complementary sensing modalities as shown in Fig. 2.1a: accelerations from an accelerometer and finger vibrations from a contact microphone. A contact microphone is a piezo-electric element that senses vibrations through solid objects (i.e. they are unaffected by acoustic waves traveling through the air, hence eliminating audio and speech privacy concerns). The ring was fabricated using the Sparkfun ADXL362 3-axis accelerometer breakout board, which measures accelerations between ±2g, and PUI Audio Inc. AB1070B-LW100-R, which has a resonant frequency of 7kHz. A custom printed circuit board (PCB) holds the PJRC Teensy 4.0 microcontroller and additionally has a non-inverting operational amplifier with a gain of 11 to amplify the signal from the contact microphone, a 1 MOhm resistor to remove drift from the contact microphone signal, and a protection diode for the Teensy. The Teensy microcontroller reads the amplified contact microphone signal using an 8-bit analog to digital converter (ADC). As a result, the recorded values range from 0-1023 (corresponding to 0-3.3V).

For our experiments, the ring is worn on the dominant hand's index finger and secured using 3M Transpore medical grade tape and a compression finger sleeve. It is connected to the printed circuit board that can be worn elsewhere on the body, in this case, on the wrist using a velcro band and Coban self-adherent wrap. Data was sampled from the contact microphone and accelerometer at 8kHz and 400 Hz respectively.

2.3.2 Scratch Intensity

Labeling

As itch increases in severity, patients often scratch with a combination of increased normal force and velocity. We propose using mechanical power as a metric for scratch intensity, calculated by taking the product of force and velocity. Supervised machine learning algorithms can be used to obtain a mapping between wearable data and scratching power. It is non-trivial to obtain ground truth measurements of the power at which a person scratches (required for supervised learning) as force and velocity cannot be estimated accurately through human observation or techniques such as motion capture. As shown in Fig. 2.3, we instead generate power labels automatically by capturing scratching activity on a pressure sensitive tablet (Sensel Morph [3]). The Sensel Morph has a sensing area of 30,960 mm^2 (240 mm by 139 mm) consisting of 19,425 pressure sensors (185 columns x 105 rows) at a 1.25 mm pitch (separating distance between sensors), with 5g to 5kg sensing range per touch. The tablet's overlay is removed and the device is instead layered with a 100% cotton cloth which is fastened to the tablet to prevent creasing. The cloth is applied to the tablet as it has a texture more similar to human skin than the hard surface of the tablet and patients with pruritus may scratch over their clothing to satisfy an itch. For each 1 second window of measurements, we extract mean force and velocity estimates, allowing us to calculate the power label. Henceforth, we will refer to each 1 second window of data as a sample.

At each timestep, t, the Sensel Morph generates a force map using its array of pressure sensors as seen in Fig. 2.3. This image is automatically analyzed by the Sensel Morph's firmware to extract individual contact(s). For each contact, the Sensel Morph outputs the force, f_t, and the centroid of the contact, (x_t, y_t). The x and y axes of the Sensel Morph are shown in Fig. 2.3. Aside from the index finger where the ring was worn, participants were instructed to keep their palm and fingers from touching the tablet during data collection to ensure there is only a single point of contact at each timestep. We sampled data from the tablet using the Sensel Morph API at 150 Hz. The entire labeling process is visualized in Fig. 2.3.

We create a force algorithm, shown visually in Fig. 2.3, to estimate mean force, $\bar{F}^s$, for each sample. Note that when no contact is sensed by the Sensel Morph, it

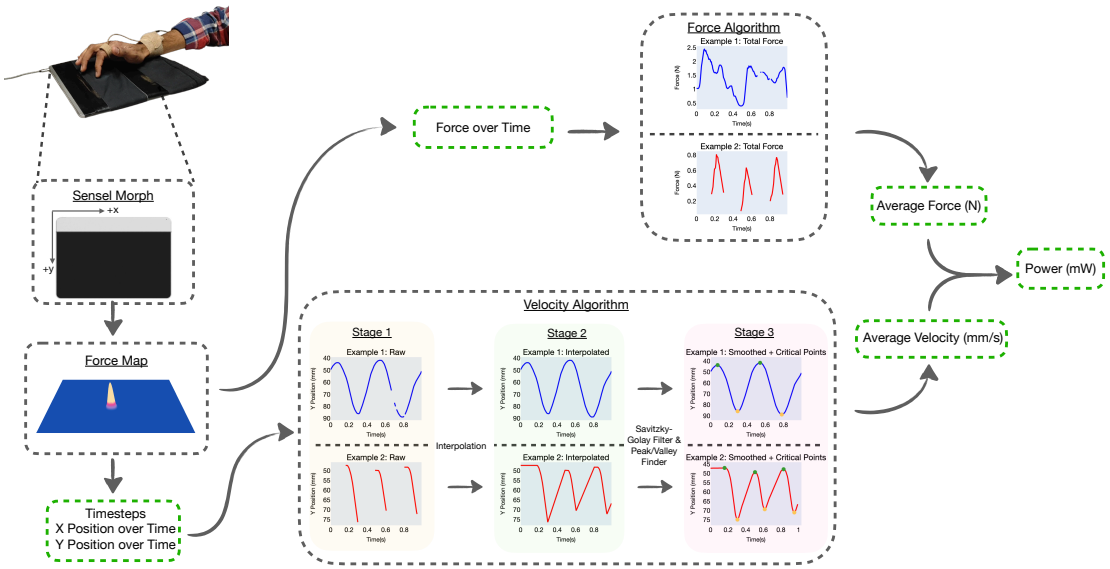


Figure 2.3: Scratch power labeling using a pressure sensitive tablet. The pressure sensitive tablet outputs force, x-position, and y-position over time for each contact. This data is fed through our force and velocity algorithms to calculate average force and average velocity for each sample of data. Power is subsequently calculated by multiplying average force and velocity. Permission was obtained from Sensel Inc. for use of the Sensel Morph image.

returns NaN; our method discards these values before calculating mean force using

$$\bar{F}^s = \frac{\sum_{j=1}^{n} F_j^s}{n} \tag{2.1}$$

where $F_j^s \in \boldsymbol{F^s}$ is the force value for a single timestep from the array of all force values $\boldsymbol{F^s}$ for a sample of data, and n is the number of values in the sample.

Calculating average velocity is more involved. As visualized in the velocity algorithm section of Fig. 2.3, the majority of scratching behavior occurs in one of two ways; either the individual maintains contact with the skin through the entire scratching motion as shown in Example 1 or they lift up their finger off of their skin in between each scratching motion as shown in Example 2. In our human studies, discussed in the Results, we had participants scratch on the Sensel Morph for z seconds at a time; as an example, in the first human study, we had participants scratch for $z = 10$ seconds. At a high level, velocity can be determined through the

segmentation of z seconds of data into distinct scratching motions. These motions are characterized by the finger moving up or down along the y-axis which are identified by detecting peaks and valleys (critical points) in the y-position signal from the Sensel Morph. By calculating the change in distance and the change in time during each of these individual scratching motions, we can obtain the average velocity across each 1-second sample. This process is explained mathematically as follows.

For every z seconds of data with n number of values, we have $\boldsymbol{x} = \Big[x_1, x_2, \ldots, x_n\Big]$, the array consisting of the x-axis values of the finger contact over time, $\boldsymbol{y} = \Big[y_1, y_2, \ldots, y_n\Big]$, the array consisting of the y-axis values of the finger contact over time, and $\boldsymbol{t} = \Big[t_1, t_2, \ldots, t_n\Big]$, the timesteps corresponding to $\boldsymbol{x}$ and $\boldsymbol{y}$. In Fig. 2.3, Stage 1 shows 1 second of $\boldsymbol{y}$ vs $\boldsymbol{t}$ data for Example 1 and Example 2. First, we interpolate both $\boldsymbol{x}$ and $\boldsymbol{y}$ to fill in any gaps where contact is broken (i.e. values that are NaN) between the finger and tablet. This yields $\boldsymbol{x}^I \in \mathbb{R}^n$ and $\boldsymbol{y}^I \in \mathbb{R}^n$. Stage 2 shows 1 second of $\boldsymbol{y}^I$ vs $\boldsymbol{t}$ data for Example 1 and Example 2. Next, as shown in Stage 3, a 5th order polynomial Savitzky-Golay filter with a filter window length of 0.21 seconds (31 points) is applied to smooth $\boldsymbol{y}^I$, yielding $\boldsymbol{y}^S$. These parameters were chosen through visual observation of the sensed contacts and generated peaks and valleys from a randomly selected subset of samples from each participant. For the jth point in $\boldsymbol{y}^S$, the formula for the Savitzky-Golay filtering is

$$y_j^S = \sum_{i=\frac{1-m}{2}}^{\frac{m-1}{2}} L_i y_{j+i}^I, \frac{m+1}{2} \leq j \leq n - \frac{m-1}{2}, \tag{2.2}$$

where $m = 31$ (the window size) and the convolution coefficients, L_i can be looked up in the Savitzky-Golay tables. Also shown in Stage 3, a peak and valley finding algorithm [55] is applied to $\boldsymbol{y}_S$, identifying peaks and valleys (critical points) i.e. each incident at which the finger's velocity changes sign along the trajectory of the finger's scratching motion. Peaks are shown in green while valleys are shown in yellow. The vector of critical points is defined as $\boldsymbol{y}^C \in \mathbb{R}^m$, where m is the number of critical points in the z seconds of data. We additionally have $\boldsymbol{x}^C \in \mathbb{R}^m$, the corresponding x positions for each critical point and $\boldsymbol{t}^C = \Big[t_1^C, t_2^C, \ldots, t_m^C\Big] \in \mathbb{R}^m$, the corresponding timesteps for each critical point.

Next, in order to calculate a velocity estimate, $\bar{v^s}$, for each 1 second sample, we

select a subset of the critical points $\boldsymbol{y}^{C,s} \in \mathbb{R}^q$, where q is the number of critical points which occur in the 1 second window of time. We additionally select the corresponding x positions, $\boldsymbol{x}^{C,s} \in \mathbb{R}^q$, and timesteps, $\boldsymbol{t}^{C,s} \in \mathbb{R}^q$. We calculate Euclidean distance (mm) between each adjacent pair of $\boldsymbol{x}^{C,s}$ and $\boldsymbol{y}^{C,s}$, forming an array of distances,

$$\boldsymbol{d}^{C,s} = \left[\sqrt{(y_2^{C,s} - y_1^{C,s})^2 + (x_2^{C,s} - x_1^{C,s})^2}, \ldots, \sqrt{(y_q^{C,s} - y_{q-1}^{C,s})^2 + (x_q^{C,s} - x_{q-1}^{C,s})^2}\right] \in \mathbb{R}^{q-1}. \tag{2.3}$$

We can also find the difference in timesteps,

$$\boldsymbol{\Delta t}^{C,s} = \left[t_2^{C,s} - t_1^{C,s}, t_3^{C,s} - t_2^{C,s}, \ldots, t_q^{C,s} - t_{q-1}^{C,s}\right] \in \mathbb{R}^{q-1}. \tag{2.4}$$

We then calculate an array of velocities using

$$\boldsymbol{v}^s = \boldsymbol{d}^{C,s} / \boldsymbol{\Delta t}^{C,s}. \tag{2.5}$$

Lastly, we find the mean velocity for the 1 second sample,

$$\bar{v}^s = \frac{\sum_{a=1}^{q-1} v_a^s}{q-1} \tag{2.6}$$

where $v_a^s \in \boldsymbol{v}^s$ is the velocity estimate between two adjacent critical points from the array of all velocities $\boldsymbol{v}^s$ for a sample of data. For each sample, the power label in mW, p^s, is found using

$$p^s = \bar{F}^s \bar{v}^s, \tag{2.7}$$

where $\bar{F}^s$ (N) is the calculated average force and $\bar{v}^s$ (mm/s) is the average velocity for the sample.

To mitigate errors in the power labeling procedure above, we remove outliers in our velocity algorithm. We filter using 3 conditions: 1. If there are less than 2 peaks/valleys in a 1 second window of data. 2. If there are consecutive peaks or consecutive valleys detected in a window of data. 3. If the change in contact x or y positions in consecutive timesteps is greater than 5 mm. Lastly, we remove any outliers greater than 600 mW as power values higher than this limit are often infeasible.

Dataset, Feature Extraction, and Model Design

For each 1-second sample of wearable data, we use linear interpolation to fill in any missing values in the contact microphone and accelerometer z-axis (normal to skin) signals, $\boldsymbol{c}$ and $\boldsymbol{a}$ respectively. Thus, we have $\boldsymbol{c^I} \in \mathbb{R}^{8000}$ and $\boldsymbol{a}^I \in \mathbb{R}^{400}$ as two interpolated signals in the time domain. Power information is captured in the frequency domain of the contact microphone and accelerometer data. We explore this idea further in the Results using participant data from a human study. The Discrete Fourier Transform (DFT), $\mathcal{F}$, can be applied to a discrete signal in the time domain to convert it to the frequency domain. For the kth point in the DFT result, $\boldsymbol{d}$, the DFT formula is

$$\boldsymbol{d}_k = \sum_{n=0}^{N-1} e^{-2\pi i \frac{kn}{N}} \boldsymbol{g}_n \tag{2.8}$$

where $\boldsymbol{g}$ is the time domain signal and N is the length of $\boldsymbol{g}$. We apply DFT and extract the single-sided frequency amplitude for the contact microphone signal,

$$\boldsymbol{c^A} = \frac{2}{B}|\mathcal{F}(\boldsymbol{c}^I)|, \tag{2.9}$$

where $B = 8000$. Similarly, for the accelerometer,

$$\boldsymbol{a}^A = \frac{2}{E}|\mathcal{F}(\boldsymbol{a}^I)|, \tag{2.10}$$

where $E = 400$.

From $\boldsymbol{c}^A$, we select the first 400 amplitudes, yielding contact microphone features, $\boldsymbol{c^G} \in \mathbb{R}^{400}$, which corresponds to frequencies from 0 to 399 Hz. From $\boldsymbol{a}^A$, we select the first 175 amplitudes yielding accelerometer features, $\boldsymbol{a}^G \in \mathbb{R}^{175}$, which corresponds to frequencies from 0 to 174 Hz. These features are concatenated, forming a feature vector, $\boldsymbol{G} \in \mathbb{R}^{575}$. Min-max normalization, fitted on the training data, was applied to $\boldsymbol{G}$, scaling all features to between 0 and 1. This processing pipeline is shown visually in Fig. 2.2.

Using our normalized wearable features and power labels from the Sensel Morph, we trained a neural network for regression of scratch intensity. To find the best model, we conducted a grid search over model hyperparameters. The hyperparameters that we searched over were the number of hidden layers, the number of nodes in the hidden

layers, dropout probability, and the number of accelerometer and contact microphone frequency features. For each model, we used leave one subject out cross validation (LOSO-CV) to calculate the mean absolute error (MAE) across all folds and the model with the lowest MAE was picked. A fully connected neural network with 2 hidden layers and hidden layer size of 1000 nodes with ReLU activation had the best performance. During training, we applied a dropout (p = 0.1) to each hidden layer to reduce overfitting. Lastly, we trained using the Adam optimizer with a learning rate of $5 * 10^{-6}$, MAE loss, batch sizes of 64, and 150 epochs.

2.3.3 Scratch Detection

Dataset, Feature Extraction, and Model Design

Supervised machine learning in conjunction with frequency features from the proposed multimodal wearable ring can additionally be used for detection of scratching behavior. Wearable data is divided into 1 second samples using a stride of 0.25 seconds and features were extracted using the DFT. From $\boldsymbol{c}^A$, the first 275 amplitudes are selected yielding contact microphone features, $\boldsymbol{c}^H \in \mathbb{R}^{275}$, which corresponds to frequencies from 0 to 274 Hz. From $\boldsymbol{a}^A$, the first 200 amplitudes are selected yielding accelerometer features, $\boldsymbol{a}^H \in \mathbb{R}^{200}$, which corresponds to frequencies from 0 to 199 Hz. These features are concatenated, forming a feature vector, $\boldsymbol{H} \in \mathbb{R}^{475}$. Min-max normalization, fitted on the training data, was applied to $\boldsymbol{H}$, scaling all features to between 0 and 1.

The normalized frequency features are used to train a fully connected neural network with 3 hidden layers with ReLU activations and a hidden layer size of 1200 nodes. A sigmoid activation was applied to the output layer. During training, we applied dropout (p=0.2) to each hidden layer. We used binary cross-entropy loss and the Adam optimizer with a learning rate of $1 * 10^{-5}$. The model was trained for 150 epochs with batch sizes of 64. Similarly to the intensity regression model, we found the best scratch detection model by conducting a grid search over model hyperparameters. The hyperparameters that we searched over were the number of hidden layers, the number of nodes in the hidden layers, dropout probability, and the number of accelerometer and contact microphone frequency features.

2.3.4 Statistics and Reproducibility

For evaluation of all of our models, we use Leave One Subject Out Cross Validation (LOSO-CV). In LOSO-CV, n iterations (folds) of cross validation are conducted where n is the number of participants. For each iteration, a single participant's data is held out as a test set while the model is trained on all other participant data. We utilize LOSO-CV as it mirrors the clinically relevant scenario of tracking scratching in new subjects whose data is not in the training set [196]. Metrics are averaged over all held-out participants (all test folds from LOSO-CV).

For intensity regression, we use mean absolute error (MAE) as a metric and as our loss function. For an array of ground truth labels, $\boldsymbol{y} \in \mathbb{R}^n$, and an array of predictions, $\hat{\boldsymbol{y}} \in \mathbb{R}^n$, MAE is calculated using the following formula:

$$\text{MAE(mW)} = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n}. \tag{2.11}$$

We additionally use mean absolute percentage error (MAPE) as a metric for better interpretability of our results. MAPE is calculated using the following equation:

$$\text{MAPE}(\%) = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{600n} \tag{2.12}$$

as 600 mW is the max power label on our 0-600 mW power scale. For detection of scratching behavior, we use mean classification accuracy as a metric and use binary cross entropy loss for our loss function.

For additional evaluation of our intensity regression results, we use the Wilcoxon signed-rank test and Spearman's rank correlation coefficient. Further details on these statistical tests are located in the Results section.

For reproducibility of our results, we release all data and code as detailed with the published work [166].

2.4 Results

2.4.1 Scratch Intensity Dataset and Feature Extraction

We conducted a human study (Carnegie Mellon University IRB Approval under 2021_00000480) with 24 healthy participants, of which data from 20 participants was used. Three participants' data was disregarded due to sensor failures during data collection while one participant was dis-enrolled prior to data collection due to not meeting inclusion criteria. Summarized demographics information can be found in Fig. 2.4a. Written informed consent was obtained from all participants. The participants were affixed with the wearable ring and instructed to scratch with just their index finger contacting the tablet, allowing our labeling algorithm to estimate power. They were led through the 9 combinations of low, medium, and high force, and low, medium, and high speed scratching on the Sensel Morph for 10 seconds each. The ordering of the combinations was low force low speed, low force medium speed, low force high speed, medium force low speed, medium force medium speed, medium force high speed, high force low speed, high force medium speed, and high force high speed. Participants have differing interpretations of low, medium, and high force and speed which allows us to collect a diverse dataset of varying scratching behavior with a large range of power labels.

Dataset information is shown in Fig. 2.4b. As specified in the Methods, we split the data into 1 second windows of sensor measurements with an associated power label automatically generated via the Sensel Morph pressure tablet. We capture a total of 30 minutes of scratching behavior across all 20 participants (approximately 1.5 minutes of scratching per participant) resulting in 6,600 data samples as seen in Fig. 2.4b. After cleaning erroneous labels, there are 4,227 samples. From this cleaned dataset, estimated force measurements are grouped for each instruction of low, medium, and high force and are shown on the left side of Fig. 2.4d. The mean and standard deviation for forces are 0.36 ± 0.23, 0.66 ± 0.29, and 1.56 ± 0.58 Newtons for instructed low, medium, and high force respectively. Similarly, estimated velocity measurements from the Sensel Morph are shown on the right side of Fig. 2.4d for each instruction. The mean and standard deviation for velocities are 111.97 ± 48.74, 136.36 ± 56.04, and 177.93 ± 62.67 mm/s^2 for instructed low, medium, and high

velocity respectively. The power labels are shown in Fig. 2.4e for each combination of low, medium, and high force, and low, medium, and high speed and range from 0 to 600 mW. The dataset is right skewed, with the vast majority of labels in the 0-200 mW range, seen by the histogram of all power labels in Fig. 2.4f.

We observe that power information is captured in the frequency domain of the contact microphone and accelerometer z-axis (normal to skin surface) data, which is transformed from the time domain using the Discrete Fourier Transform (DFT). As force and velocity increase, the amplitude of the vibrations between the fingernail and skin correspondingly rise. In Fig. 2.4c, each line corresponds to one of the nine combinations of low, medium, and high force, and low, medium, and high speed that participants were instructed to perform. For each of these combinations, we transform each sample into the frequency domain, min-max normalize to scale values to between 0 and 1, average all the signals across all the participants, and then apply smoothing using a Savitzky-Golay filter for visualization purposes. In Fig. 2.4c, note that as the instructed force changes from low, medium, to high, the magnitudes of the signals also increase for most frequencies. Similarly, as the instructed velocity changes from low, medium, to high, the magnitudes of the signals generally increase. Noticeably, in Fig. 2.4c, the vertical ordering of the signals in the higher frequency ranges ($\geq$ 150 Hz for the contact microphone and $\geq$ 70 Hz for the accelerometer and marked in red in the plots) directly matches the ordering of the power values recorded by the tablet, shown in Fig. 2.4e. These results suggest that frequency domain features from wearable data can be used for regression of scratching power. As discussed in the Methods, for the contact microphone data, we specifically use the first 400 amplitudes, which corresponds to frequencies from 0 to 399 Hz while for the accelerometer data, we use the first 175 amplitudes yielding accelerometer features, which corresponds to frequencies from 0 to 174 Hz.

2.4.2 Scratch Intensity Model Performance

As discussed in the Methods, normalized frequency features in conjunction with ground truth power labels from the pressure sensitive tablet are used to train a neural network for regression of scratch intensity. We utilize Leave One Subject Out Cross Validation (LOSO-CV) for evaluating model architectures and for tuning

model parameters. Performance metrics, specifically Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE), are then averaged across all folds of cross validation.

Fig. 2.4g presents results for scratch intensity regression using LOSO-CV across the 20 healthy participants from the first human study. The model trained using both accelerometer and contact microphone features performs the best with a MAE of 49.71 mW and a MAPE of 8.29%. The minimum MAPE for 1 held-out participant (1 fold) is 2.33% while the maximum MAPE is 17.20%. As shown in Fig. 2.4g, we conduct an ablation study. For this ablation study, we evaluate our algorithm with just the contact microphone features and just the accelerometer features to compare the performance of single-sensor approaches with the multimodal model. The model trained using just accelerometer features performs similarly with a MAE of 50.58 mW and a MAPE of 8.43% while the model trained only with contact microphone features has a MAE of 60.66 mW and a MAPE of 10.11%. All models perform better than the naive predictor which uses the mean, 119.64 mW, of all the labels as the prediction. Fig. 2.4h shows the spread of test errors from LOSO-CV for different power label ranges.

2.4.3 Scratch Intensity Validation

We trained a scratch intensity model using all 20 participants' data from the first human study with optimal model hyperparameters found using a grid search. We then validated the performance of this trained model through a second human study (Carnegie Mellon University IRB Approval under 2021_00000480) in which 14 healthy participants were instructed to perform various intensities of scratching. Demographics information for this study can be found in Fig. 2.5a. Written informed consent was obtained from all participants.

Each participant participated in 4 sets of instructed scratching. Each set consists of scratching at intensities from 1-5 with 8 seconds of scratching at each intensity. Participants were told that 1 is the least intensely they would scratch on their skin to satisfy an itch while 5 is the most intensely they would scratch on their skin to satisfy an itch. Participants conducted the first two of these sets on the pressure sensitive tablet, scratching with just their index finger to allow us to record ground truth power

estimates using the Sensel Morph. Participants were then instructed to conduct two more sets on their skin, each at a location of their choosing and participants were allowed to scratch normally with no restrictions on the fingers used. All participants chose to scratch on either their medial forearm, lateral forearm, medial upper arm, or lateral upper arm. Sensel Morph data was labeled and wearable data was processed using the same pre-processing steps as used for the training data.

Using data from the two sets of scratching on the pressure sensitive tablet, Fig. 2.5b presents results for intensity regression for the 14 healthy participants from the second human study. Our multimodal wearable and trained intensity regression model achieves a MAE of 57.4 mW and a MAPE of 9.57%. The minimum MAPE for 1 held-out participant (1 fold) is 2.92% while the maximum MAPE is 24.15%.

Using data from the two sets of scratching on skin, we show that our learned model can differentiate between scratching intensity levels on skin. For all participants, for each instructed intensity, we calculate the mean regressed intensity value across the 8 seconds of scratching on skin. Fig. 2.5c shows that as the instructed intensity increases, the mean regressed intensity likewise increases from 96.52 mW to 129.17 mW to 170.92 mW to 203.78 mW to 262.03 mW. Using the Wilcoxon signed-rank test, we find statistical significance between each intensity level and especially strong significance between intensity levels 1 and 2 and intensity levels 4 and 5. For each intensity level, $n = 28$, as each of the 14 participants scratches twice at each intensity. The exact p values are 1.68e-5, 3.50e-4, 1.55e-3, and 2.79e-5. Fig. 2.5d additionally shows our device's ability to differentiate between scratching intensities on skin. Similarly to the last figure, for all participants, for each instructed intensity, we calculate the mean regressed intensity value across the 8 seconds of scratching on skin. To account for different understandings of the 1-5 scale by participants and variation in the two sets, we apply min-max normalization across the averaged regressed intensity values from each set. In Fig. 2.5d, we show the mean and standard deviation across all participants for the normalized sets and additionally overlay all normalized data points. The means for intensity levels 1-5 are 0.027, 0.218, 0.416, 0.572, and 0.948 respectively. The Spearman's rank correlation coefficient for this data is 0.83 with a p-value of 6.41e-37 implying a strong positive correlation in the data.

Our trained model predicts the power (in mW) associated with a scratching action. However, the 0-600 mW scale doesn't hold much interpretability to patients or their

doctors and thus, we present a method for converting this scale to a 0-10 continuous scale (similar to the Visual Analogue Scale (VAS) commonly used in dermatology assessment). We apply a simple linear scaling by dividing all labels and predictions by 60. Thus, the bounds in power units for the 0-10 scale are the following: 0, 60, 120, 180...540, 600 mW. In the new 0-10 scale, after the scaling, the MAE (from Fig. 2.5b) becomes 0.96 units with a standard deviation of 0.55 units. In the published work [166], we additionally present an alternate non-linear square root function which achieves a mean absolute error of 1.37 units.

Through capturing ground truth power labels with the pressure sensitive tablet, we observe that people have very different interpretations of the 0-10 continuous scale. For example, a scratch intensity of 6 for one person is an intensity of 1 for another person. To demonstrate this, using ground truth values from the two sets of scratching on the Sensel Morph, for each participant, for each instructed intensity from 1-5, we linearly scale the ground truth values to the 0-10 scale. For each participant, we fit a line to this data as shown in Fig. 2.5e. The lines have an average slope of 18.22 mW/unit (sd: 12.24) and an average y-intercept of 12.17 mW (sd: 32.93), showing a large difference in scales for each participant. The ideal line, shown in dark green in Fig. 2.5e, has a y-intercept of 0 as an intensity of 0 on the 0-10 continuous scale corresponds to no scratching (0 mW). It also has a slope of 60 as an intensity of 10 on the 0-10 continuous scale corresponds to 600 mW on the power scale. This ideal line is only possible with a "perfect" participant who understands the power scale and is able to control their scratching intensity exceptionally well. Fig. 2.5f further shows the difference in scales between 3 participants using a heatmap. Participant 3's scale ranges from 105.64 mW (corresponding to 1 on their subjective version of the 0-10 scale) to 354.00 mW (corresponding to 10 on their subjective version of 0-10 scale). Participant 7's scale ranges from 29.43 mW to 173.37 mW. Lastly, Participant 11's scale ranges from 23.01 mW to 81.95 mW.

2.4.4 Scratch Detection Dataset and Feature Extraction

To develop and validate our algorithms for scratch detection, we collected a dataset from 20 healthy participants. Demographics are shown in Fig. 2.6c. Participants were instructed to scratch on different locations of their body (top of hand/fingers,

forearm/wrist, inside elbow, neck, head, behind the knees, ankles) and to do a variety of non-scratching activities (hand waving, keyboard typing, texting/phone swiping, writing, table tapping, air scratching, clapping) for 30 seconds each. Raw data from the contact microphone and accelerometer for each scratching and non-scratching interaction made by a single representative participant can be seen in Fig. 2.6a and Fig. 2.6b. All data collected was divided into 1 second samples using a stride of 0.25 seconds. In total, there is 140 minutes of data with 32,760 samples. The dataset is balanced with equal number of scratching and non-scratching data. Dataset information can be found in Fig. 2.6b. As discussed in the Methods, we utilize the DFT to extract frequency features from wearable data and use these to train a neural network to detect scratching behavior.

2.4.5 Scratch Detection Model Performance

In Fig. 2.6d, we present results for scratch detection using LOSO-CV across the 20 healthy participants. The model using both accelerometer and contact microphone features achieves an accuracy of 89.98% on the balanced dataset. The minimum accuracy for 1 held-out participant (1 fold) is 80.22% while the maximum accuracy is 98.05%. Fig. 2.6d also provides results from an ablation study to quantify the benefits of the multimodal sensing ring; classification accuracy drops from 89.98% to 86.24% and 79.98% with the exclusion of either the accelerometer and contact microphone respectively.

Fig. 2.6e provides the classification accuracy for each interaction. For the non-scratching interactions, the model performs especially well for hand waving and clapping, classifying them with 97% and 98% accuracy respectively. It performs worst on the writing and tapping interactions, classifying them with 75% and 79% accuracy respectively. For the scratching interactions, the model correctly classifies scratching behavior at all scratching locations with over 89% accuracy.

2.5 Discussion

This work introduces a framework for the quantification of scratch intensity which included a multimodal wearable ring, methods for capturing ground truth intensity

of scratching interactions from a pressure sensitive tablet, and a supervised machine learning algorithm for regression of scratching intensity on a 0-600 mW power scale. We show how extracted frequency features from contact microphone and accelerometer raw data hold valuable information that allows for the regression of scratching power using a neural network. We additionally quantify the performance of this method using LOSO-CV and showed it performs well in both human studies, including the validation study with data from 14 new participants that our models were not trained on.

We showed that our scratch intensity model, trained on data collected from scratching on a tablet, accurately captures scratching on the skin. Although we have a pressure sensitive tablet for controlled validations, as a community, we lack the skin-tight pressure sensor array technology that would be necessary to quantify the error (in mW power units) that our method exhibits for on-skin scratching. Thus, to evaluate our algorithms, we instructed participants to scratch at intensities from 1-5 on their skin. Our model achieves statistical significance between the different intensities and additionally shows a strong Spearman's correlation after normalization, providing compelling evidence for the ability of a ring wearable and learned regression model to differentiate between scratching intensities on skin. The ability of our model to infer intensities on a 0-10 scale also maps to existing scales that are common in dermatology practice for evaluation of itch and scratching.

Through our human studies, we observed some outliers in performance; for example, there was low correlation for one participant between instructed intensities and regressed values for scratching on skin. Our analysis from data from both studies also found higher errors in predictions when scratching behaviors reached high power magnitudes. We suspect this is due to fewer samples in those higher bounds and due to some saturation in accelerometer and contact microphone signals.

Our method allows the research and dermatology communities to quantify for the first time the large differences in how people interpret the 0-10 scales that are commonly employed in clinical patient-reported assessments. The difference in participants' interpreted scratching behaviors in Fig. 2.5E and the 3 participants presented in Fig. 2.5F demonstrate the subjectivity of the 0-10 scale that is commonly employed in clinical practice. These results further motivate the need for a wearable device like the one presented that provides objective measures of scratch intensity.

We foresee intensity estimates from our device being utilized in various ways. The proposed power scale for scratch intensity holds physical meaning, allowing for future exploration of its relationship to skin damage. In clinical and drug trials, the regressed power values can be used directly by researchers for quantitative and statistical analysis of treatment efficacy. However, in outpatient clinical settings, this power scale holds little value and interpretability to practicing doctors and patients, who are more familiar with existing scales such as the 0-10 continuous Visual Analog Scale or 0-10 discrete Numerical Rating Scale. The data from the power scale can be converted to an objective 0-10 continuous scale, by applying a simple linear scaling or non-linear transformation on the output of the neural network, or a 0-10 discrete scale, by binning the labels and using classification algorithms. To make our results more understandable, we provide two examples of this. Using the 0-10 objective, continuous scale and a linear scaling, our method yields a MAE of 0.96 units and with a non-linear scaling, our method yields a MAE of 1.37 units. For our data, the linear model is superior to the non-linear model presented but for higher recorded scratching power, a non-linear model could fit the data better. For reference, in clinical trials, studies often use a change of 4 units on a subjective, patient reported version of the 0-10 scale as the minimal clinically important difference [115, 206]. This suggests that a linear mapping of our power measurements to a 0-10 scale has sufficiently low error and would allow for clinically-relevant discrimination of scratching intensity levels. Additionally, in contrast to existing patient-reported scales which only capture the symptom of itch at a snapshot in time, this technology has the potential to provide continuous monitoring without any added effort by doctors and patients.

Our results additionally show promise for improving binary classification for scratch detection. Using LOSO-CV, our model achieved 89.95% accuracy on the balanced dataset with equal numbers of scratching and non-scratching data. Our ablation results show that the proposed combination of two acousto-mechanic sensors yields superior performance over single sensor methods. This was especially the case for scratch detection, where the synthesis of features resulted in an increase of 3.74% and 10.00% in performance over the accelerometer and contact microphone single sensor approaches respectively. These results motivate further work in multimodal sensing wearables for detection of scratching behavior specifically for daytime use when there are confounding movements. While our study focused on differentiating scratching

behaviors from many common daily activities, future work could additionally consider behaviors that may be prevalent or similar to scratching, such as kneading, picking, and rubbing.

The form factor and location of the ring wearable also presents benefits to the long-term adoption of the technology, especially for daytime use. Conspicuous, non-aesthetic, assistive and medical devices with symbolism of impairment have shown to be less positively accepted due to social stigmas [53, 191]. While our prototype device is currently tethered and requires a battery and supporting electronics on the wrist in addition to the sensorized ring, there are existing, well-adopted commercial ring devices, such as the Oura Ring [2], which show that electronics with multiple sensors, including accelerometers, can be successfully embedded in a ring form factor with a long lasting battery life and aesthetic design [12, 142]. One limitation of this design choice is that the device only has the ability to track scratching motions by a sole finger on a single hand, which may result in undetected scratches. We foresee patients wearing the ring on the finger they scratch with the most regularly.

Another limitation of this work is the lack of evaluation with patients with pruritus. Future work will focus on development of a fully integrated, wireless ring for continuous scratch monitoring allowing long-term real-world studies that evaluate our algorithms in practice on representative patients, who may scratch differently than healthy participants. Additional testing and evaluation of our multimodal wearable ring for tracking scratching behaviors would be helpful before long-term deployment, including a signal strength and noise analysis for when the ring is placed at different locations along the index finger or moved to other fingers.

As proposed in past works, we envision this device automatically uploading this data to patient records in the cloud, where it could be subsequently aggregated, analyzed, and visualized in a mobile application [113, 124, 125]. This data can additionally be synced with other information such as sleep data, temperature/humidity, and patient reported details such as diet, allowing individuals to identify correlations between potential itch triggers and their subsequent itch symptoms.

Our work highlights a multimodal sensing ring for quantification of scratch intensity in addition to scratch detection, two essential metrics for the tracking of itch. Within clinical trials, this wearable and framework could provide researchers and clinicians with greater insight into treatment effectiveness via real time signals that directly

correspond to scratch behaviors. By providing direct objective quantification of scratching, our ring device has the potential to improve patient outcomes and aid with treatment discovery and validation.

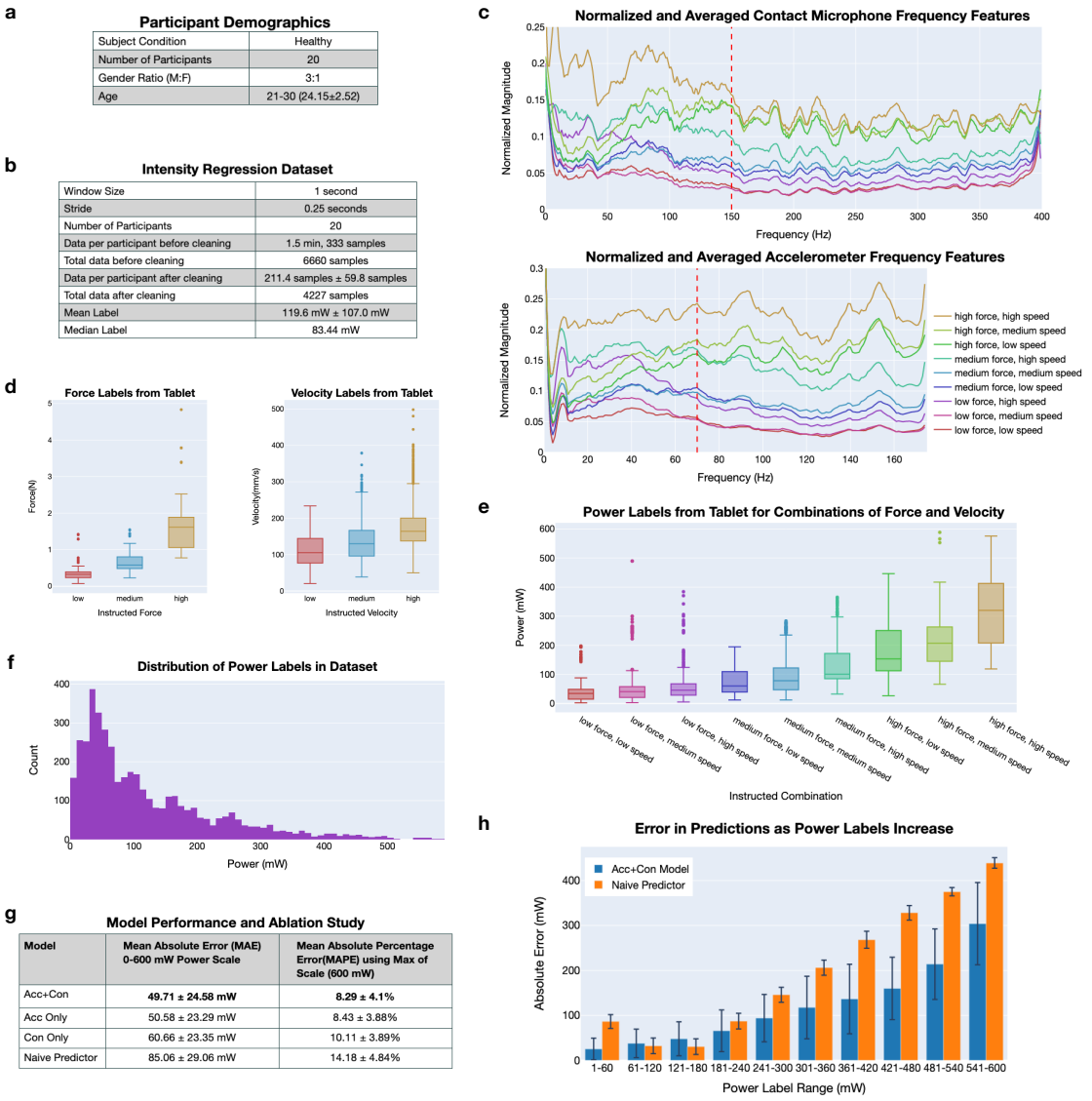


Figure 2.4: Scratch intensity dataset, features, and performance using LOSO-CV. a. Participant demographics from first human study. b. Details on the intensity regression dataset. c. Contact microphone and accelerometer DFT features for combinations of force and velocity. Vertical ordering of the DFT amplitudes greater than 150 Hz and 70 Hz (designated by the red dashed vertical lines) match the ordering of power labels in E. d. Box plots of the recorded forces and velocities from the Sensel Morph as participants scratched on the tablet at combinations of varying force and velocities. e. Box plot of power labels from all participant data for combinations of force and velocity. f. Histogram of all power labels in the dataset. g. Performance using LOSO-CV and ablation study over sensor features used. h. Box plot of errors in predictions on test data for various ranges of labels.

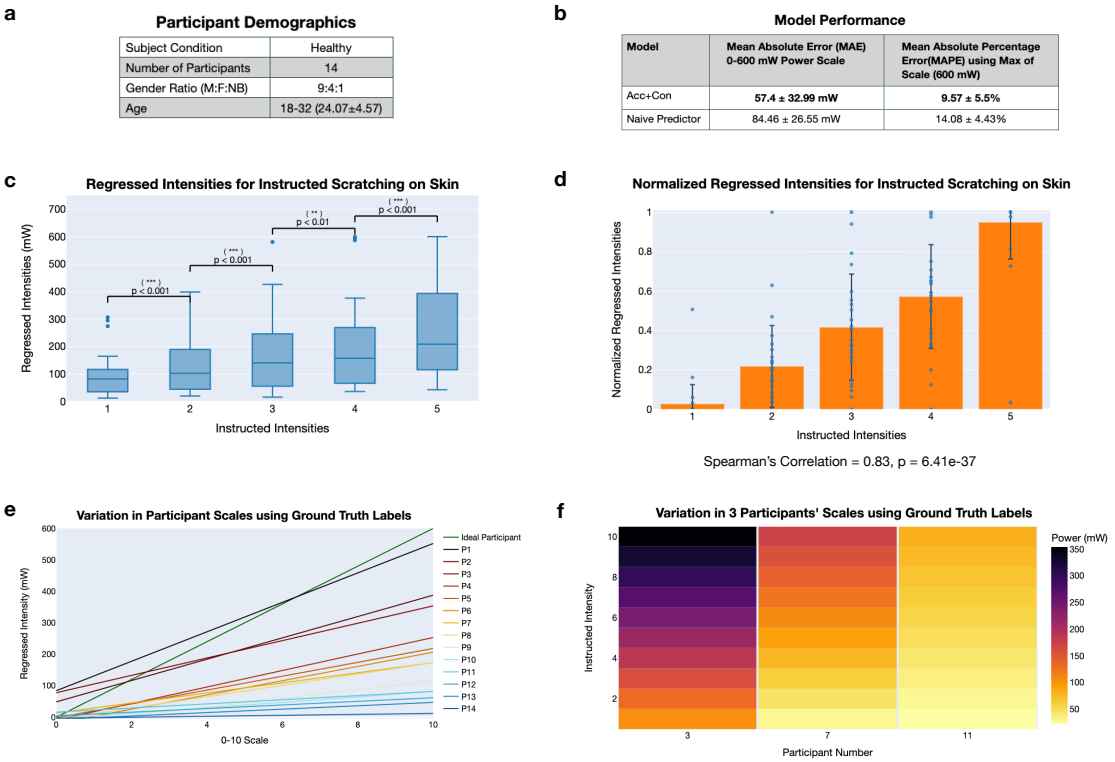


Figure 2.5: Scratch intensity validation performance. a. Participant demographics for second human study. b. Model performance on new scratch intensity data. ± denotes standard deviation. c. Regressed intensities on the 0-600 mW power scale for scratching on skin for instructed intensities from 1-5. Regressed intensities are grouped based on the instructed intensity. For each instructed intensity, $n = 28$. Statistically significant differences are found between all groups based on a Wilcoxon signed-rank test. d. Normalized regressed intensities for scratching on skin grouped by the instructed intensity value. A Spearman's correlation of 0.83 is calculated showing a strong positive correlation. Bars indicate the mean and error bars indicate 1 standard deviation from the mean. For each instruted intensity, $n = 28$. e. The variation in individually fit participant intensity scales using ground truth labels from the pressure sensitive tablet is shown. For each participant, a line is fit to the data from both sets of scratching on the Sensel Morph. f. Variation in three participants' scales using ground truth labels from the pressure sensitive tablet is shown.

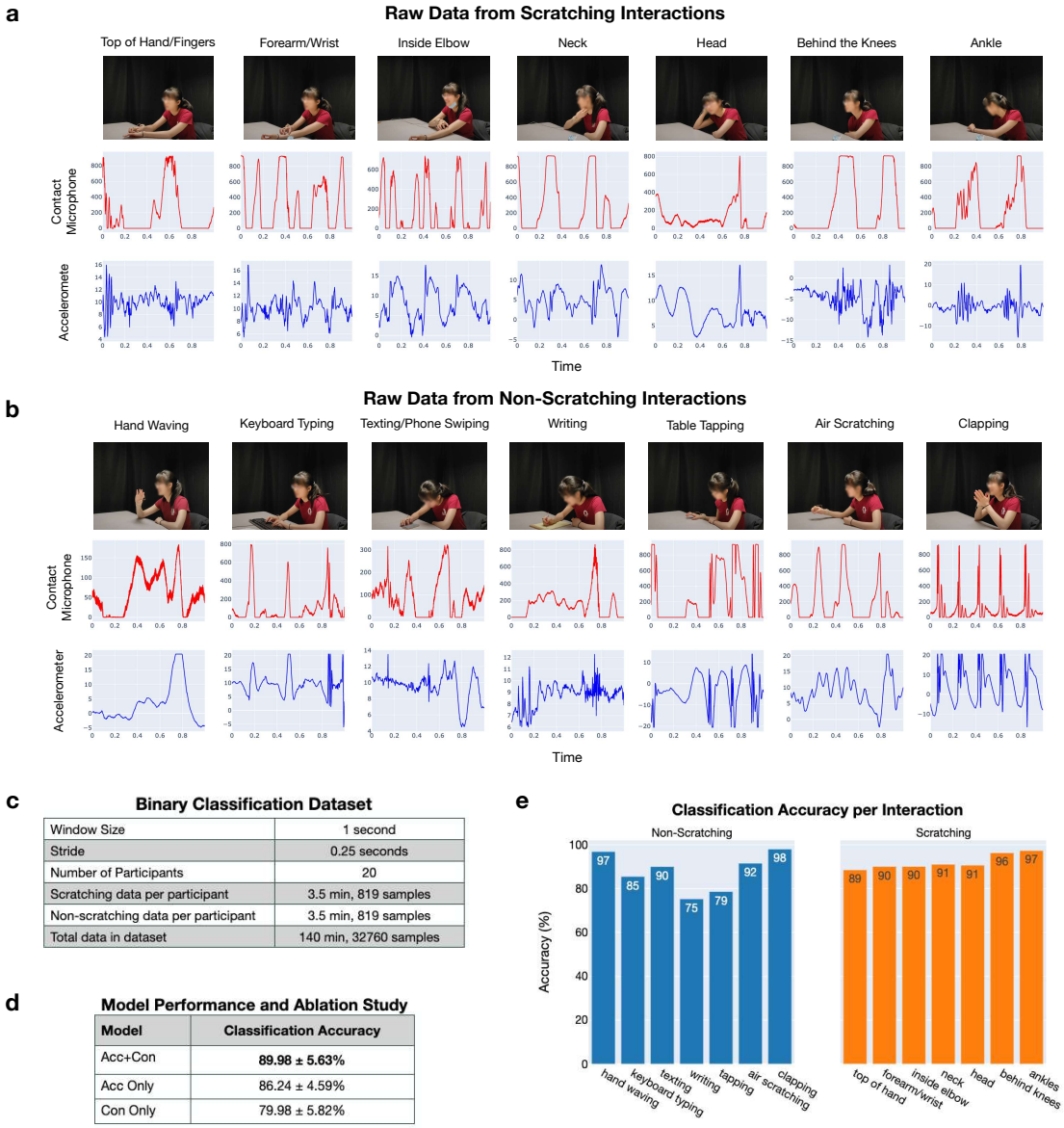


Figure 2.6: Scratch detection dataset and performance using leave one subject out cross validation. a. The 7 scratching interactions are shown with accompanying raw data from the contact microphone and accelerometer for a representative participant. b. The 7 non-scratching interactions are shown with accompanying raw data from the contact microphone and accelerometer. c. Details on the scratch detection dataset. d. Results from the ablation study conducted using LOSO-CV. Acc stands for Accelerometer and Con stands for Contact Microphone. ± denotes standard deviation. e. Classification test accuracy averaged across all folds for each interaction in the dataset using the best classification model (Acc+Con). Permission was obtained from the participant for use of their image in publication.

Chapter 3

In-Vivo Skin 3-D Surface Reconstruction and Wrinkle Depth Estimation using Handheld High Resolution Tactile Sensing

3.1 Contributions

For the following chapter, the contributions are as follows. I have led and contributed extensively to all aspects of this project. I formulated the core contributions with Zackory Erickson and Carmel Majidi with input from Arash Mostaghimi and Sonal Choudhary. Achu Wilson designed the initial gels and I iterated on this design. Austin Williams designed the final gels and manufactured them. I and Austin Williams characterized the mechanical properties of the gels. I designed and collected indentation data with the indenter rig. I implemented and tested various 3D reconstruction algorithms. I wrote the paper and all authors provided feedback. Arpit Agarwal provided code, guidance, and advising on calibration of GelSight sensors for 3D reconstruction. Catherine Li explored algorithms for the detection and estimation of skin wrinkles. Sankalp Chopkar designed the custom probe hardware with my guidance. Dinesh Patel assisted with characterization of the gels, printing of the test

pieces, and characterization of the test pieces. He also advised on figures. Wenzhen Yuan provided advice on the design of custom gels and 3D reconstruction. Dr. Arash Mostaghimi and Dr. Sonal Choudhary provided advice on the medical implications of the work. Arash also advised on study design and writing an IRB protocol. Carmel Majidi provided feedback on the characterization of the custom gels. Zackory Erickson and Carmel Majidi advised on all aspects of the projects.

3.2 Introduction

Three-dimensional (3-D) reconstruction of the human skin surface holds broad potential in dermatology. Clinically, it may aid in diagnosing and monitoring textural changes in conditions such as eczema, psoriasis, and scleroderma, as well as in evaluating treatment efficacy. In cosmetology, it can reveal insights into skin aging, sun damage, and the effects of skincare products. Traditionally, beyond visual inspection, dermatologists rely on manual palpation [46, 47, 186] and dermatoscopes [38, 61, 141] to assess skin texture. Portable 3-D reconstruction tools offer a promising path toward more objective and precise assessment, but no handheld probe with high spatial and depth resolution has been validated for use across body locations [29, 56, 119, 121, 216]. As a result, gaps remain in the literature, particularly a lack of quantitative data on wrinkle depth across different body locations with most prior studies having focused only on the wider, deeper wrinkles found on the face [35, 49, 155, 219].

Using photometric stereo techniques to convert tactile images into 3-D height maps [105], GelSight provides a compact and precise solution with micron-level accuracy. GelSight has already demonstrated utility in biomedical applications such as prostate stiffness estimation [52] and lump detection [104]. Previous work has also demonstrated GelSight's value in industrial applications, such as inspecting manufactured components [9, 130], robotic manipulation [7, 120, 165, 243], material hardness estimation [244], and forensics [21, 129]. Commercial GelSight systems developed by GelSight Inc. exist but require precisely molded elastomer gels and controlled illumination to achieve high accuracy, raising costs and subsequently limiting broader adoption. Additionally, as these probes are used only with rigid objects, their gels are designed to be less elastic to increase durability, which can cause significant deformation while being used on soft materials like human skin.

There also exist cheaper, compact GelSight sensors that rely on RGB lighting, but achieving high 3-D reconstruction accuracy is challenging. This is because of the shallow and uneven LED illumination angles, necessary for the compact form factor, leading to spatially varying illumination and artifacts that worsen with indentation.

Motivated by the limitations of commercially available devices and existing compact GelSight technologies for skin surface height reconstruction and wrinkle depth estimation, we propose 3-D reconstruction, skin wrinkle detection, and skin wrinkle height estimation methodology and algorithms using compact GelSight sensors. We first design highly elastic gels, seen in Fig. 3.1A, to reduce skin distortion during contact, enabling accurate capture of skin features. We then introduce a learning-based reconstruction algorithm that achieves a mean absolute error (MAE) of 12.55 microns on wrinkle-like test objects, allowing precise measurement of fine skin textures in any skin location on the body, exceeding the capabilities of current commercially available devices. From the resulting GelSight-generated height maps, we develop methods to detect and quantify skin wrinkles. We integrate the compact sensor and custom gel into a probe, shown in Fig. 3.1B, with an embedded load cell to control the applied normal force, and evaluate its performance in measuring skin wrinkle depth.

In a human study of 15 participants without skin conditions, we use our device and methods to quantify wrinkle depth across multiple body locations, providing the first set of metrics for many of these areas in the literature. We further demonstrate the ability to detect reductions in wrinkle depth at three body locations following the application of an over-the-counter moisturizer, with statistically significant differences observed before and after moisturization for all locations. We show our full pipeline in Fig. 3.1, including GelSight images from the wrist before and after moisturizer, corresponding reconstructed 3-D height maps, wrinkle detection, and estimation of wrinkle height metrics from these height maps. Unlike existing portable dermatological probes, which are typically single-purpose and lack micron-level precision, our device and methodology offer the potential for a wide range of future applications both in clinical and home settings. These include skin condition diagnosis, treatment monitoring, moisture estimation, hair thickness and density estimation, and tracking subtle micro-textural changes due to aging or environmental factors.

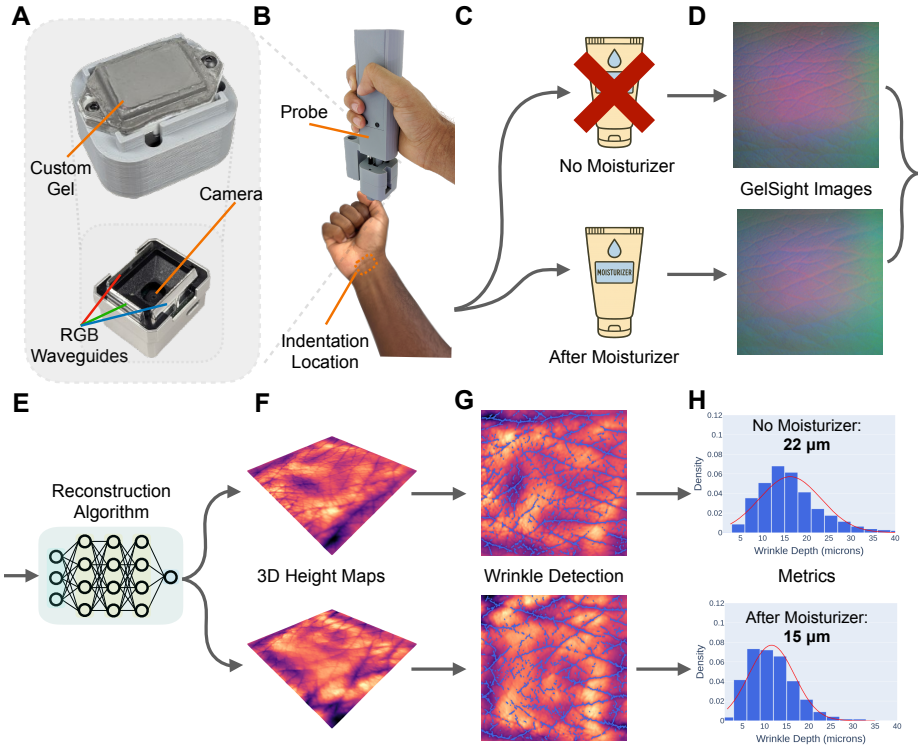


Figure 3.1: **A.** Bottom: The GelSight Mini sensor is shown without the gel cartridge. The sensor has a high-resolution camera and three waveguides that inject red, green, and blue (RGB) light into the gel. Top: Our custom elastic gel is shown attached to the GelSight Mini sensor. **B.** Our custom 3-D printed probe is shown integrated with the GelSight sensor, custom gel, and a load cell for measuring normal force. **C.** We collect GelSight readings before and after moisturizer application. **D.** GelSight images from the wrist are shown before and after moisturizer application. **E.** Our reconstruction algorithm converts GelSight images into 3-D height maps. **F.**3D Height Maps are shown for before and after moisturizer application for the wrist. **G.** Our wrinkle detection algorithm identifies wrinkle valleys, shown in blue overlaid on the height map. **H.** Normalized histograms of wrinkle depth are shown for before and after moisturizer application. Metrics on wrinkle depth can be automatically calculated, with the 80th percentile wrinkle depth shown. In this case, moisturizer application results in a 7 micron reduction in wrinkle height from 22 micron to 15 micron.

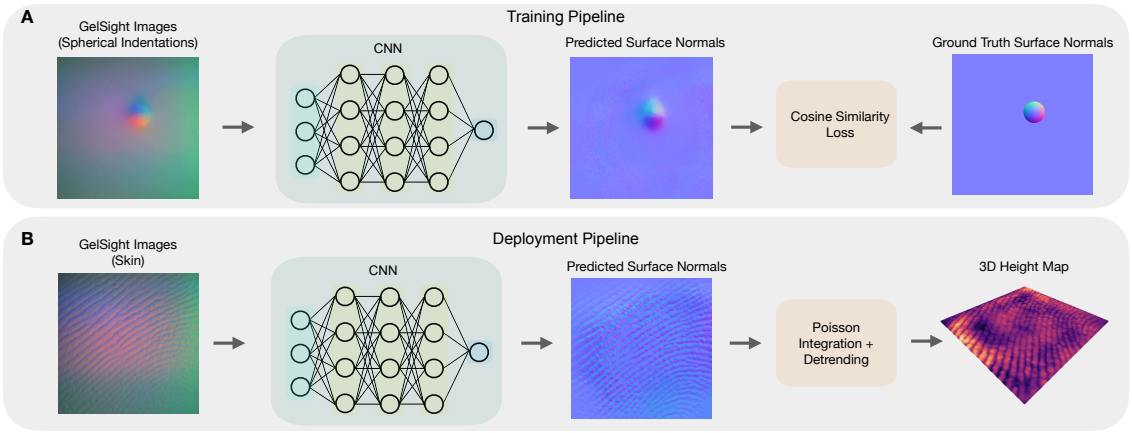


Figure 3.2: Training and Deployment Pipelines. **A.** Training Pipeline. GelSight images of spherical indentations with corresponding ground truth surface normals are used to train a convolutional neural network (CNN) with cosine similarity loss for prediction of surface normals. **B.** Deployment Pipeline. The trained CNN inputs GelSight skin images and outputs predicted surface normals. These surface normals are converted to 3-D height maps using poisson integration and are detrended using a 2-D high pass filter. The skin image displayed is a fingerprint.

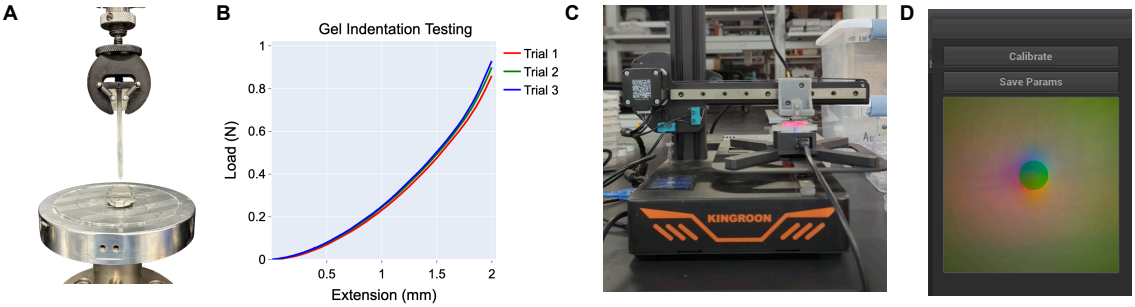


Figure 3.3: Gel Characterization and Calibration. **A.** Instron indentation testing setup used to estimate the elasticity of the custom gels. **B.** Load (N) versus extension (mm) curves from three indentation test trials. **C.** Our calibration indentation rig is shown with the indenter attached and the GelSight sensor, featuring the custom gel, mounted on the stage. **D.** Our indentation labeling graphical user interface (GUI) is shown which allows us to label GelSight images with corresponding ground truth surface normals.

3.3 Experimental Section

3.3.1 Sensor

We utilized the GelSight Mini Tactile Sensor from GelSight Inc. The GelSight Mini consists of a high-resolution 8 megapixel camera providing images of 2448 by 3264

pixels. Red, green, and blue light from LEDS from three of the four sides of the rectangular sensor are directed into an acrylic piece attached to the gel cartridge providing uniform illumination of the gel surface. The GelSight Mini sensor without the gel cartridge is shown in Fig. 3.1A. Additional information on the sensor can be found on the GelSight Inc. website [1].

3.3.2 Gel Design

To minimize deformation of the soft material as the GelSight is pressed against its surface, we designed and manufactured custom gels for the GelSight Mini sensor. Our custom gel attached to the sensor is shown in Fig. 3.1A. The cartridge hardware consists of a 3-D printed mount made of PLA (polylactic acid) and a clear acrylic piece (0.7 inches by 0.95 inches by 0.11 inches). First, the acrylic is press fit and glued into the 3-D-printed mount. A mold is constructed using another 3-D printed piece. The gel consists of three layers: base elastomer, powder coating, and the skin-safe elastomer. The base elastomer is made with Silicones Inc. XP-565 at a 1:22 ratio of the activator to the base. The two parts of the elastomer are mixed, degassed for 12 minutes, and poured into the mold. The elastomer is cured for 24 hrs in an oven at $55°C$ and removed from the mold. Next, to add the opaque diffuse coating for the gel, we used solid aluminum spherical powder with particles of diameter 1 µm from US Research Nanomaterials, Inc. The aluminum powder was first sifted using a sieve to remove powder clumps and then applied using a makeup brush to the surface and sides of the gel. Next, we mix a certified skin-safe elastomer, Smooth-On SORTA-Clear™ 18, with a solvent, Smooth-On NOVOCS™ Matte, in a ratio of 1:16.5:33 for the activator to the base to the solvent. We pour the mixture over the coated gel and leave it tilted for 5 min for the excess elastomer to flow off. Lastly, the gel is cured for 24 hrs in an oven at $55°C$.

3.3.3 Gel Mechanical Characterization

We characterized the mechanical properties of our soft, skin-safe gels, described in Section 3.3.2, using a spherical indenter with radius, $R = 1.5$ mm, and an Instron

[1] GelSight Inc. https://www.gelsight.com/

tensile testing machine. The spherical indenter is 3-D printed using a Formlabs Form 3 SLA printer with Gray Pro Resin. Our setup is shown in Fig. 3.3A. Using Hertzian contact theory, we can estimate the Young's modulus of the gel, E_{gel}.

For a sphere indented into a half-space, using Hertzian contact theory, we can use the following equations:

$$F = \frac{4}{3} E^* R^{\frac{1}{2}} d^{\frac{3}{2}} \tag{3.1}$$

$$\frac{1}{E^*} = \frac{1 - v_1^2}{E_1} + \frac{1 - v_2^2}{E_2} \tag{3.2}$$

where F is the applied force, R is the radius of the spherical indenter, d is the displacement, E_1 and E_2 are the Young's moduli of the indenter and gel respectively, and v_1 and v_2 are the Poisson ratio's of the indenter and gel respectively. To simplify Eq. 3.2, we can assume E_1 is infinity as the indenter is orders of magnitude more stiff than the gel. Thus,

$$\frac{1}{E^*} = \frac{1 - v_2^2}{E_2} \tag{3.3}$$

We assume $v_2 = 0.49$ as silicone elastomers can be approximated as nearly incompressible materials. Substituting Eq. 3.3 into Eq. 3.1, we get the following:

$$F = \frac{4}{3} \left(\frac{E_2}{1 - v_2^2}\right) R^{\frac{1}{2}} d^{\frac{3}{2}} \tag{3.4}$$

Lastly, using the raw data from the force-displacement curves, shown in Fig. 3.3B, from the Instron machine, we can solve for E_2 in Eq. 3.4, the Young's modulus of the gel, by using non-linear least squares regression. We conduct 3 total indentation tests on Gel 1, indenting 2 mm into the gel. Results from these tests are presented in Section 3.4.1.

3.3.4 Calibration Indenting Rig

In order to collect calibration data for each cartridge, we used a modified Kingroon KP3 FDM 3-D printer as a 3-axis indenting rig. As seen in Fig. 3.3C, the sensor is mounted to the center of the 3-D printer bed using a 3-D printed mount. The printer nozzle is removed and an indenter is attached in place of the nozzle using a

3-D printed mount. We 3-D printed 3 spherical indenters with radii of 0.5, 0.875, and 1.25 mm using the Formlabs Form 3 stereolithography (SLA) printer out of Clear Resin with a layer thickness of 25 micron.

We utilized a Python script to concurrently send commands to the calibration indenting rig and to collect data from the mounted GelSight sensor. For each indenter, 49 total indentations are collected in a grid pattern. Specifically, from the center (0, 0) x,y location on the gel, we indent at various x and y coordinates. We define two sets, $X = (-4.5, -3, -1.5, 0, 1.5, 3, 4.5)$ and $Y = (-4.5, -3, -1.5, 0, 1.5, 3, 4.5)$ and use the 49 unique combinations of these X and Y values.

Because the gel surface may not be perfectly flat, we use a simple computer vision algorithm to align the indenter with the surface of the gel before indenting by the radius of the spherical indenter. The algorithm automatically detects the moment when the indenter is perfectly aligned with the surface of the gel by moving down at larger increments and then detecting when the indenter untouches the gel by moving up at smaller increments. This can be defined mathematically as follows:

For each unique position, (x,y) with $x \in X$ and $y \in Y$, the printer first moves in the positive z-axis to a clearance height and then to the (x,y) position. We define $I_{untouched} \in R^{300 \times 300 \times 3}$, which is the untouched median image of 5 total images taken concurrently with a crop of 300 by 300 around the indentation location. We also define $e_{untouched}$, the median of the pixel intensity mean squared error (MSE) between all unique pairwise combinations of the 5 images with a crop of 300 by 300 around the indentation location. Lastly, we define $t_t = 1.1 * e_{untouched}$, the touch threshold. We use 5 images with medians to account for the measurement noise in the images. Next, we move the indenter down by 50 micron incrementally. Each time the indenter is moved, we collect 5 new images, crop them around the indentation location, and find $E_{curr} \in R^5$ which is the MSE between each image and $I_{untouched}$. Next, e_{curr} is calculated using the median of E_{curr} and once $e_{curr} > t_t$, we know that the indenter is touching the gel. At this point, we set $e_{final} = e_{curr}$ and we define t_u, the untouched threshold which is defined as $t_u = 0.7 * e_{final}$. We follow a similar process to detect when the indenter untouches the gel, but using positive z-axis movements of 15 micron. When $e_{curr} < t_u$, we know the indenter is no longer touching. Finally, we indent the gel by the radius of the indenter. Using this process, we can ensure that our indentations have a depth error of at most 15 micron. The thresholds were selected

by the research team through testing and visual inspection of the GelSight images.

Using a set of digital calipers lightly pressed into the gel, we estimate the millimeter to pixel conversion factor as 0.0077 mm/pixel. Next, for each indenter, we visually examine the GelSight images from the indentation at (x,y) = (0,0). As seen in Fig. 3.3D, we visually label the center of each indentation and we use these values to automatically label the centers of the other 48 indentations. Using these center labels and the radius of the indenters converted to pixel space using the mm to pixel conversion factor, we generate ground truth surface normal images to pair with each GelSight image. An example ground truth surface normal image is shown in Fig. 3.2.

3.3.5 Convolutional Neural Network (CNN) for estimation of surface normals

Our training pipeline is illustrated in Fig. 3.2. Each GelSight image is cropped from its original size of 2448 by 3264 pixels to a 1500 by 1500 pixel region. We center the crop at (1224, 1482), which corresponds to a 150-pixel shift to the left along the width dimension. This cropping removes border regions with uneven RGB illumination.

Using the 144 images, 48 from each size of indenter, and the corresponding ground truth normals, we train a convolutional neural network (CNN) to estimate surface normals from the RGB GelSight images. We use a train-test split of 90-10 to partition our data into training and testing sets. The CNN input, $f \in R^{1500\times1500\times8}$, consists of the GelSight image, $I \in R^{1500\times1500\times3}$, the unindented (untouched) GelSight image, $B \in R^{1500\times1500\times3}$, and two positional encodings for the image, $u \in R^{1500\times1500\times1}$, and $v \in R^{1500\times1500\times1}$. The untouched GelSight image helps the CNN identify the changes in lighting during indentation while the positional encodings help the CNN learn the variation in lighting across the entire surface of the gel. We pass our input through two convolutional blocks, each consisting of a 2D convolution layer with 64 output channels, a kernel size of 3, stride of 1, and "same" padding to preserve spatial dimensions. Each convolutional layer is followed by a batch normalization layer to stabilize training and for better generalization, and a ReLU activation function to introduce non-linearity. A final convolutional layer maps the output, $O \in R^{1500\times1500\times3}$, to 3 channels using the same kernel size, stride, and padding. We use cosine similarity

loss defined as

$$\mathcal{L}_{\cos} = \frac{1}{\sum_{i=1}^{H}\sum_{j=1}^{W} M^{(i,j)}} \sum_{i=1}^{H}\sum_{j=1}^{W} M^{(i,j)} \left(1 - \frac{\langle \mathbf{n}_{\text{pred}}^{(i,j)}, \mathbf{n}_{\text{gt}}^{(i,j)} \rangle}{\|\mathbf{n}_{\text{pred}}^{(i,j)}\| \cdot \|\mathbf{n}_{\text{gt}}^{(i,j)}\|}\right)$$

where $M \in \{0,1\}^{H \times W}$ is a binary mask indicating the loss regions, H and W are the height and width of the image (e.g., 1500×1500), and $\mathbf{n}_{\text{pred}}^{(i,j)}$ and $\mathbf{n}_{\text{gt}}^{(i,j)}$ denote the predicted and ground truth normal vectors at pixel (i, j).

The mask M is constructed as a binary matrix with 1s in a square region surrounding the indentation and an additional random subset of pixels outside this region, selected to cover a fixed percentage $\gamma = 5\%$ of the total image area. This supervision allows the model to learn the lighting changes associated with deformations while also encouraging accurate estimation in unindented regions, which should be predicted as flat.

We implement our CNN architecture in PyTorch and train it using one NVIDIA 4090 and one 4070 GPU in parallel. Due to the high dimensionality of the input data, we use a batch size of 1 with 32 gradient accumulation steps. Training is performed using the ADAM optimizer with a weight decay of 1×10^{-5} and an initial learning rate of 1×10^{-3}, which decays exponentially with a gamma of 0.95. We train the model for a total of 30 epochs. We perform several hyperparameter sweeps over parameters such as kernel size (1, 3, 5, 11), number of convolutional blocks (1, 2, 3), and number of output channels per block (16, 32, 64, 128). The final model configuration corresponds to the best-performing setup on the validation objects using Gel 1. We then train a separate model for Gel 2 using the same parameters.

3.3.6 Poisson Integration and Detrending

As seen in Fig. 3.2, during deployment, we use Poisson integration to convert surface normals into a 3-D height map. Prior works commonly use the discrete sine transform (DST)-based solvers with Dirichlet boundary conditions, which assumes the height is zero at the image boundaries [10, 243]. However, this assumption does not hold for our application, as the gel surface often deforms at the edges while pressing the gel against skin. Instead, we use an FFT-based Poisson solver with periodic boundary conditions. While periodicity is not physically accurate, it avoids imposing artificial

constraints at the borders and we find it works well in practice since our primary interest lies in local variations in surface height rather than absolute global structure.

After Poisson integration, we additionally detrend the height map to highlight higher frequency features by applying a 2-D high-pass filter with a cutoff frequency of 0.002, effectively removing low-frequency components such as large-scale surface curvature or tilt that is larger than 500 pixels, approximately 3.85 mm. This allows us to focus on finer geometric details and local variations in the surface, which are more informative for wrinkle detection. We crop the boundaries of our reconstructed height maps by 100 pixels on each side to remove the effects of boundary artifacts introduced by the Poisson integration solver. Our final height maps have dimensions of 1300 by 1300, corresponding to a physical area of approximately 1 cm by 1 cm.

3.3.7 Validation and Test Objects

In machine learning, it is essential to have separate validation and test sets to ensure robust model development and evaluation. We use the validation set to tune hyperparameters and select the best-performing model. The test set is then used to evaluate how well the final model generalizes to unseen objects with different height profiles.

For both our validation and test objects (rigid and soft), we 3-D print test pieces using a digital light processing (DLP) based 3-D printer (Pico2 HD@27, Asiga). This printer was operated by a top-down DLP system with a digital mirror device (DMD) and a UV–LED light source operating at 385 nm. The rigid resin used in this work is GR-1 (Pro3-Dure). The soft resin used for 3-D printing comprises of epoxy aliphatic acrylate (EAA, Ebecryl 113, Allnex USA) and aliphatic urethane acrylate (AUD, Ebecryl 8411, Allnex, USA) in the weight ratio of 1:1. 2 wt. % TPO (diphenyl(2,4,6-trimethylbenzoyl)phosphine oxide, Genocure TPO, RAHN USA Corp.) as photo-initiator and 0.15 wt. % of Rhodamine dye add to the acrylate mixture in a hot water bath at 86 °C with string at 300 rpm [114, 127, 177, 222]. Each layer was irradiated for 0.2 s, and layer thickness was 12 μm. The printed structures were sonicated with isopropyl alcohol (IPA) for 3 min to remove the uncured resin and post cured in UV oven (UVP CL-1000 UV Oven) for 6 mim.

For our validation set, we use 4 3-D-printed objects with straight channels as

seen in Fig. 3.4. The channels are 500 micron wide with 500 micron between each channel. The 4 objects have channels with 24 micron, 48 micron, 72 micron, and 96 micron depths respectively. We collect a GelSight reading in both the x and y directions as reconstruction performance may differ in the two axes due to uenven surface illumination; this is specifically exaggerated in the GelSight Mini sensor we utilized due to the lack of even angle and spacing between the three waveguides as seen in Fig. 3.1A.

For the 6 validation images collected, for each of the 3 test pieces in the 2 orientations (vertical and horizontal), we assess the height reconstruction performance as follows. We take cross sections of the reconstructed height map at 100 pixel increments. For each cross section, we apply a peak detection algorithm [55]. We invert the cross section by multiplying all heights by -1 and then run our peak detection algorithm again to find the valleys. Lastly, we take the absolute value of all differences across all cross sections. Using these height differences and comparing to the ground truth channel depths, we can find the errors in height reconstruction across the surface of the entire gel and at various depths.

For our test set, we use 4 3-D-printed objects with circular channels as seen in Fig. 3.4. The channels are 350 micron wide with 350 micron between each channel. The 4 test objects have channels with 12 micron, 36 micron, 60 micron, and 84 micron respectively. We collect a single GelSight reading with the test piece approximately centered in relation to the sensor. Next, we visually label the center of the test object in our GelSight image. We then follow a similar procedure to the validation test pieces, except that we use cross sections in a circular pattern at angle increments of 15 degrees.

3.3.8 Wrinkle Detection and Height Estimation Algorithms

We detect wrinkles in the 3-D height maps by identifying valleys (wrinkle troughs). To do this, we iterate through all $(x, y) \in R$, a 1300×1300 grid of locations on the height map. For each location, we perform angled cuts passing through that location at angles of $-60°$, $-30°$, $0°$ (vertical line), $30°$, $60°$, and $90°$ (horizontal line). We extract the corresponding 1D height cross-sections from the height map.

For each of these cross-sections, to identify valleys, we invert the cross section

by multiplying all height values by -1 and apply peak detection [55]. The resulting valleys correspond to wrinkle locations. Lastly, we skeletonize all valley candidate points using a 2-D skeletonization algorithm [247].

Next, to estimate wrinkle depth, from our skeletonized valleys, we randomly sample 10,000 valley points for time and compute efficiency. For each valley point, we search for the highest peak within a 30 pixel radius, approximately 0.23 mm. This value serves as a reasonable approximation informed by typical ridge-to-ridge distances observed in human fingerprints [148]. The height difference between each valley and its corresponding highest peak within this radius is computed as the wrinkle depth. Of the 10,000 total wrinkle height estimates, we use the 80th percentile value to represent wrinkle depth. This choice reflects the observation that many body locations contain numerous micro-wrinkles, which dominate the lowest points in the data, while deeper, more prominent wrinkles are less common. The 80th percentile provides a rough but more representative estimate of overall wrinkle depth.

3.3.9 Human Study Procedure

We conduct a human study with 15 participants, approved by Carnegie Mellon University's Institutional Review Board. The objective is to collect GelSight images to characterize wrinkle depths at various body locations and to examine the effects of an over-the-counter moisturizer on skin wrinkle depth.

We use a custom-designed 3-D-printed probe (Fig. 3.1B) with an integrated load cell to collect GelSight readings with consistent force during skin contact. The load cell is a TE Connectivity FX1901-0001-0025-L, with a maximum operating force of approximately 111.21 N and an accuracy of $\pm 1\%$. Force signals from the load cell are amplified using the SparkFun Load Cell Amplifier Breakout Board with an Avia Semiconductor HX711 Analog to Digital Converter, and read by an Arduino Nano microcontroller. The Arduino transmits the force measurements to a companion laptop, which displays both live GelSight images and force data. A button embedded in the probe handle allows the researcher to tare (zero) the load cell, accounting for gravitational effects from probe orientation. The researcher tares before each indentation on the skin and monitors the live force and image streams to ensure proper contact. A GelSight image is captured by pressing the spacebar on the companion

laptop. All readings are taken at a load cell reading of approximately 19.62 N (2000 grams).

The study proceeds as follows. After obtaining written informed consent, we begin by collecting demographic information. Three body locations (palm, wrist, elbow) are marked with a Sharpie marker. Participants rub the areas using a nail polish removing wipe containing acetone on each location in order to remove moisture from the skin for initial readings. This is due to the studies being conducted in June in a geographical location (Pittsburgh, Pennsylvania, USA) with a relative humidity that regularly exceeds 75%. Next, two GelSight readings are captured at each site with the sensor aligned to the Sharpie mark. We apply 0.1 milliliters (mL) of Vanicream Moisturizing Cream for Sensitive Skin using a syringe to each location. Participants wear a glove to apply the cream to the three marked locations, rubbing it in until no visible residue remains. A timer is set for 8 minutes to allow the skin to absorb the moisturizer. Meanwhile, we collect a single GelSight reading at an additional 7 locations: forehead, inside elbow, inside upper arm, top of the middle finger, knuckle, and top of the hand. Once the 8 minutes have elapsed, the participant uses a disinfecting hand wipe to wipe any cream residue from the 3 moisturized areas. After the locations dry out, we realign the GelSight sensor with the Sharpie marks and capture 1 post-treatment image at each of the three locations.

All measurement locations are illustrated in Fig.3.6E and Fig.3.7E. For all hand and wrist readings, participants rest their left hand in a relaxed position on a table. We sample the knuckle of the middle finger and the fingerprint region of the pinky finger. For the top of the finger location, we target the middle phalanx of the middle finger. The top of the hand is sampled in an area with minimal body hair. For the inside elbow and upper arm, participants fully extend their arm. For the outside elbow, participants bend their elbow and rest it on a table; the measurement is taken just above the joint, where the skin surface is flatter than directly over the joint.

3.4 Results

In order to minimize deformation of the skin during contact with the GelSight sensor, we design custom elastic gels as detailed in Section 3.3.2. Our gel is attached on the GelSight Mini sensor as seen in Fig. 3.1A. In Section 3.4.1, we detail results from

mechanical characterization of our custom gels. We use two gels produced in separate manufacturing batches, referred to as Gel 1 and Gel 2.

Our goal is to reconstruct a 3-D height map from GelSight RGB images. As seen in Fig. 3.2B, our pipeline consists of two steps: 1. A convolutional neural network (CNN) to estimate surface normals given the GelSight image input and 2. Poisson integration to estimate height from the predicted surface normals and a 2-D high-pass filter to detrend the height map. Our outputted height maps have dimensions of 1300 by 1300, corresponding to a physical area of approximately 1 cm by 1 cm.

For training a supervised learning CNN to estimate surface normals, for each gel, we gather an indentation dataset using spherical indenters of 3 diameters and a 3-axis linear stage, shown in Fig. 3.3C. This process is detailed in Section 3.3.4 and the training pipeline is shown visually in Fig. 3.2A. More information on the CNN algorithm and Poisson reconstruction methods are provided in Section 3.3.5 and Section 3.3.6 respectively. In Section 3.4.2, we present results on the performance of our method on 3-D printed validation and test pieces.

Lastly, we develop skin wrinkle detection and height estimation algorithms, detailed in Section 3.3.8, that can be applied to our reconstructed height maps. We run a human study with 15 participants, approved by Carnegie Mellon University's Institutional Review Board (Protocol Number: 2023_00000269), to characterize wrinkle depths at various locations on the human body and to additionally show the impact of moisturization using an over the counter cream on skin wrinkle depth. The procedure used for this human study is detailed in Section 3.3.9 and we present the results from this study in Section 3.4.3.

3.4.1 Gel Elasticity

We use an Instron tensile testing machine, shown in Fig. 3.3A, equipped with a 50 N load cell (1% resolution) and a spherical indenter to estimate the Young's modulus of one of our two custom manufactured gel cartridges, Gel 1, using Hertzian contact theory. Additional methodology details can be found in Section 3.3.3.

Force-displacement curves from three indentation trials, performed using a 1.5 mm radius indenter indenting 2 mm into Gel 1, are shown in Fig. 3.3B. The estimated Young's moduli from non-linear least squares regression for Trials 1, 2, and 3 are

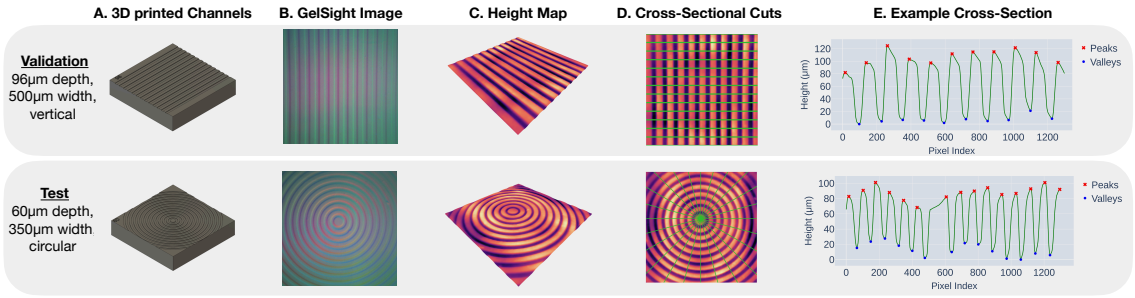


Figure 3.4: Evaluation Pipeline. **A.** Renderings of a representative straight channel validation object and a circular channel test object. **B.** Corresponding GelSight images. **C.** Reconstructed height maps of the two objects. **D.** The height maps are shown with green lines representing the cross-sectional cuts used to extract metrics. **E.** Two example cross-sections are shown with detected peaks and valleys used to automatically extract channel heights.

124.44 kPa, 129.55 kPa, and 132.74 kPa, respectively. The average Young's modulus across trials is 128.91 kPa.

3.4.2 Algorithm Performance

As detailed in Section 3.3.7, we 3-D print 4 rigid validation and 4 rigid test pieces that consist of channels in straight line and circular configurations, respectively. We use channels specifically as they resemble wrinkle ridges and valleys on the skin. The 4 rigid validation pieces have channels of depth 24, 48, 72, and 96 micron with channel width of 500 micron. The 4 rigid test pieces have channels of depth 12, 36, 60, and 84 micron with channel width of 350 micron. We note that further work is needed on validating the accuracy of these 3-D printed validation and test pieces.

In Fig. 3.4, we visually show the process of extracting channel depth estimates from the validation and test pieces. This involves taking cross-sectional cuts across the reconstructed and detrended height map and automatically detecting the peaks and valleys in the cross sections. We then find the difference in height between adjacent peaks and valleys to estimate the height of each channel. By doing this across the entire height map, we can get the average error and standard deviation of the error for each of the validation and test pieces.

We run a series of hyperparameter sweeps using only Gel 1. We select the algorithm

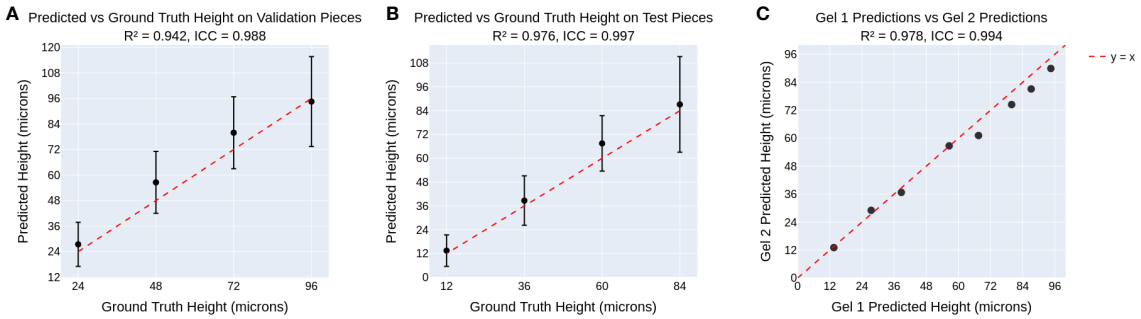


Figure 3.5: Sensor Accuracy and Precision on Validation and Test Objects. **A.** Mean predicted height vs. ground truth height is shown for the four straight channel validation objects. Results are averaged across Gel 1 and Gel 2. Error bars represent 1 standard deviation from the mean. **B.** Mean predicted height vs. ground truth height is shown for the four circular channel test objects. Results are averaged across Gel 1 and Gel 2. Error bars represent 1 standard deviation from the mean. **C.** Gel 1 mean predicted height vs. Gel 2 mean predicted height is plotted for all test and validation objects to assess precision of the sensors and algorithms. We include R^2 and ICC values for all plots, and show the $y = x$ line as a reference to indicate perfect agreement between the values on the two axes.

which achieves the highest performance (lowest mean absolute error) on the validation pieces. Our best performing algorithm is a Convolutional Neural Network (CNN); more information on this algorithm is provided in Section 3.3.5. We use the test pieces to show generalization of our method to different channel shapes, widths, and depths. We summarize algorithm performance on our validation and test sets in Fig. 3.5A and B. For these plots, we average the results across both Gel 1 and Gel 2. Both Fig. 3.5A and B show strong correlation between ground truth and predicted height for all objects. For the validation pieces, our algorithms achieve a coefficient of determination (R^2) of 0.942 and an Intraclass Correlation Coefficient (ICC) of 0.988, while for the test pieces, we get an R^2 of 0.977 and an ICC of 0.997. On the validation pieces, averaged across both gels and all 4 objects, our algorithm achieves a validation MAE of 14.01 ± 11.41 microns. On the test pieces, the average MAE across both gels is 12.55 ± 11.35 microns. Gel 1 achieves a validation MAE of 13.83 ± 11.15 microns and a test MAE of 14.67 ± 13.49 microns, while Gel 2 achieves a validation MAE of 14.19 ± 11.67 microns and a test MAE of 10.42 ± 9.21 microns.

In Fig. 3.5C, to analyze repeatability of our sensor with different gels, we plot

the mean measurement for each of the 4 validation objects and the 4 test objects for Gel 1 versus Gel 2. We obtain an R^2 of 0.974 and ICC of 0.993, showing strong repeatability of our sensor and algorithm, even with different gels.

3.4.3 Characterization of Skin Wrinkles

As detailed in Section 3.3.8, we introduce methods for detection of skin wrinkles from 3-D height maps and estimation of wrinkle height. These methods input in the 3-D height maps and output valleys that correspond to where wrinkles are located. Using these detected valleys, we can estimate wrinkle depths throughout the height map using surrounding peaks.

To further validate our probe, algorithms, and wrinkle detection algorithms, we conducted a human study with participants with no skin conditions. We detail our human study procedure in Section 3.3.9. Fifteen participants (9 male, 6 female; sex and gender by self-report) were recruited for the study, with a mean age of 23.20 years (SD = 3.91, range = 18–31). Fourteen participants identified as Asian and one as African American. No participants reported having excessive wrinkles at any of the measured body locations. Fitzpatrick skin types were distributed as follows: Type III (2 participants), Type IV (9 participants), Type V (2 participants), and Type VI (2 participants). The majority of participants did not apply any creams, lotions, or moisturizers to their skin within four hours prior to the study. One participant reported applying a product to their elbow and palm, another to their elbow and wrist, and a third to their palm. Gel 1 was used for 6 participants while Gel 2 was used for the remaining 9 participants.

Wrinkle depth can vary depending on whether the skin is relaxed or stretched, particularly around joints such as the elbow. To ensure consistency across participants, we standardize joint positioning during all measurements as described in Section 3.3.9. The specific body locations are illustrated in Fig.3.6E and Fig.3.7E.

In Fig. 3.6A and B, we present a subset of GelSight images collected in our human study and corresponding reconstructed 3-D height maps from 7 skin locations: forehead, upper arm, inside elbow, top of hand, knuckle, top of finger, and fingerprint. In Fig. 3.6C and D, we visually illustrate the results of our wrinkle detection algorithm alongside histograms of the estimated wrinkle depths across the reconstructed surfaces.

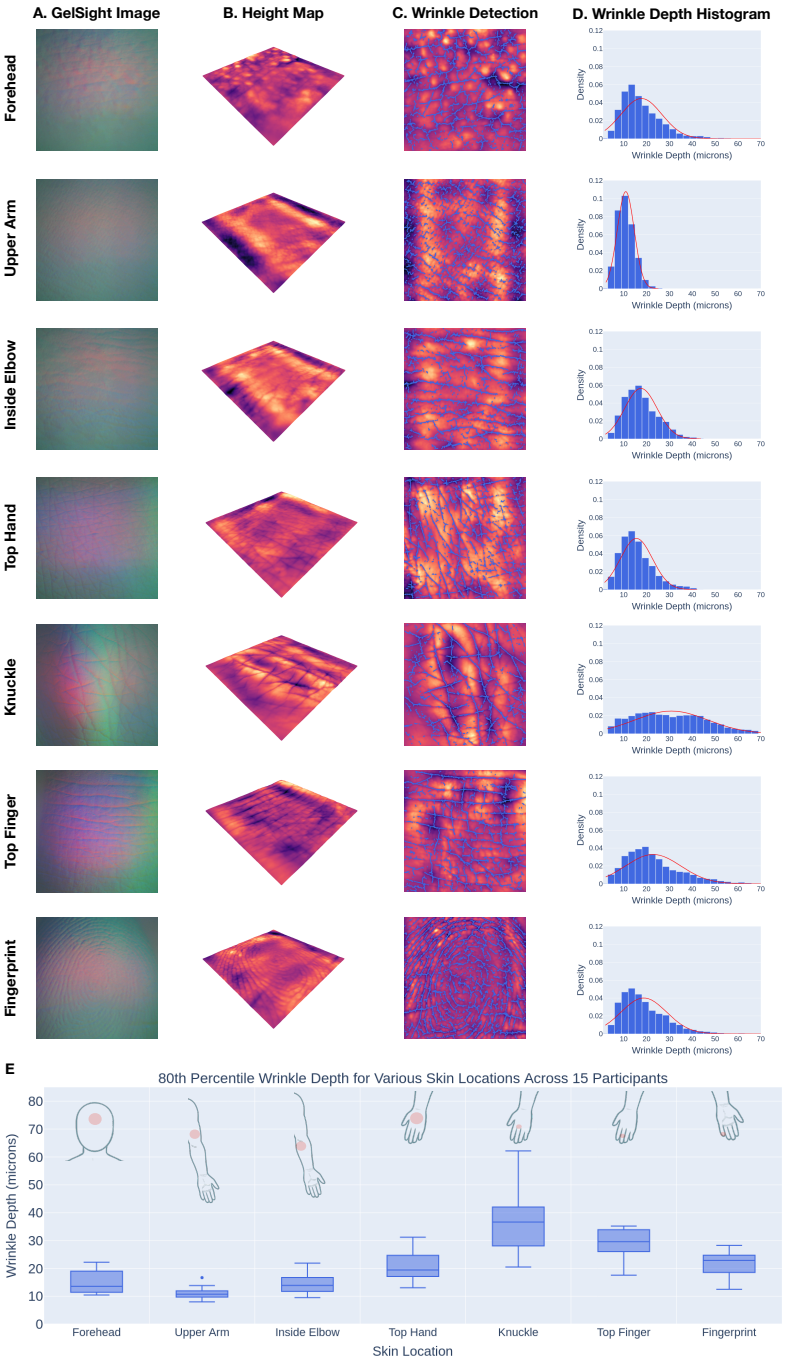


Figure 3.6: 3-D reconstruction, Wrinkle Detection, and Wrinkle Depth Estimation in Various Locations on the Human Body. **A.** GelSight images from seven skin locations: forehead, upper arm, inside elbow, top of hand, knuckle, top of finger, fingerprint. **B.** Corresponding 3-D height maps are shown. **C.** The detected wrinkle valleys (troughs) are shown in blue overlaid over the height map. **D.** Normalized histograms of wrinkle depth are shown for each location, with Gaussian fits overlaid in red. **E.** Boxplots of the 80th percentile wrinkle depths across all 15 participants are shown with clear differences seen across the various locations.

In Fig. 3.6E, we show boxplots of 80th percentile wrinkle depth for each location for all participants.

Across all participants, the knuckle region exhibited the highest mean 80th percentile wrinkle depth at 35.90 micron, followed by the top of the finger (28.46 micron) and the fingerprint region (21.54 micron). The top of the hand also showed relatively high values with a mean of 21.12 micron. In contrast, lower mean wrinkle depths were observed in the upper arm (11.09 micron), inside elbow (14.35 micron), and forehead (14.88 micron).

3.4.4 Effect of Moisturizer on Wrinkle Depth

In Fig.3.7A, we present a subset of GelSight images and corresponding 3-D height maps captured before and after moisturizer application in our human study. We also include the outputs of our wrinkle detection algorithm and histograms of estimated wrinkle depths across the reconstructed surfaces. Fig.3.7B displays boxplots of the 80th percentile wrinkle depth for each location (palm, wrist, and elbow) across all participants, including two pre-moisturizer readings and one post-moisturizer reading.

Across all three locations, we observe consistent reductions in 80th percentile wrinkle depth following moisturizer application. On average, the palm shows a decrease from 19.10 micron (Pre 1) and 18.77 micron (Pre 2) to 13.90 micron (Post). The wrist exhibits a reduction from 16.73 micron and 16.34 micron to 11.33 micron, and the elbow from 29.88 micron and 29.05 micron to 18.57 micron.

To evaluate statistical significance, we apply the non-parametric Friedman test to compare the three time points at each location. The test reveals significant differences for the palm ($\chi^2 = 19.73$, $p = .0001$), wrist ($\chi^2 = 22.53$, $p < .0001$), and elbow ($\chi^2 = 21.73$, $p < .0001$). For locations with significant overall effects, we perform post-hoc Wilcoxon signed-rank tests with a Bonferroni correction factor of 3. These tests indicate no significant difference between the two pre-moisturizer readings and statistically significant reductions in wrinkle depth between both pre readings and the post reading at all locations. These results demonstrate both the repeatability of our pipeline and the measurable effect of moisturizer application on reducing wrinkle depth.

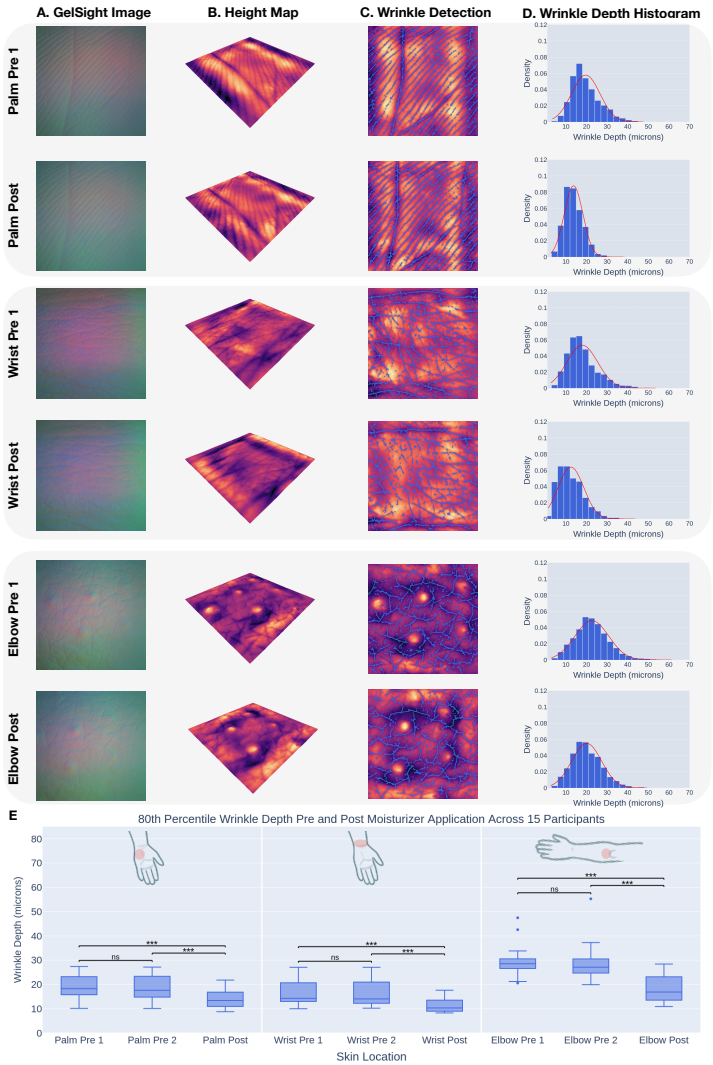


Figure 3.7: Effect of Moisturizer on Wrinkle Depth. **A.** GelSight images from three skin locations: palm, wrist, and elbow. Two readings (Pre 1 and Pre 2) are taken pre-moisturizer application and 1 reading (Post) is taken after application. **B.** Corresponding 3-D height maps are shown. **C.** The detected wrinkle valleys (troughs) are shown in blue overlaid over the height map. **D.** Normalized histograms of wrinkle depth are shown for each location, with Gaussian fits overlaid in red. The post-application Gaussian fits are shifted left relative to the pre-application curves, indicating reduced wrinkle depth. **E.** Boxplots of the 80th percentile wrinkle depths across all 15 participants are shown for the two pre moisturizer readings and one post moisturizer reading. Asterisks denote statistical significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. No statistical difference was found between the two pre-application readings, indicating sensor repeatability. Both pre-application readings showed strong statistical differences from the post-application reading (Wilcoxon signed-rank test with Bonferroni correction for 3 comparisons), demonstrating the sensor's ability to detect changes in wrinkle depth due to moisturizer across participants.

3.5 Discussion

We present a compact, high-resolution 3D skin reconstruction probe and learning-based algorithm capable of estimating micron-level wrinkle depth using GelSight tactile sensing. Our system is accurate, repeatable across different gels, generalizes to a wide range of skin topographies, and reliably detects changes in surface morphology following topical treatments. To the best of our knowledge, this is the first handheld in-vivo 3D skin reconstruction method validated on test pieces and through a human study. Existing skin measurement methods fall into direct and indirect categories [69]. Indirect methods, such as silicone replica-based profilometry using optical, confocal, or laser systems [42, 45], offer high accuracy but require bulky lab setups and are unsuitable for in-home use. In contrast, direct methods—typically integrated into handheld probes lack accurate depth (z-axis) estimation, limiting their utility for applications requiring precise surface assessment [121]. The most common direct approach, surface evaluation of living skin (SELS), uses integrated RGB cameras to assess roughness and wrinkling [44, 69]. However, SELS is not validated for 3D reconstruction and has limited depth resolution (e.g., 50 µm for the Visioscan VC 20plus), which is insufficient for detecting fine wrinkles. Optical stereo and fringe projection methods, like the PRIMOS system[193], offer 3D measurements but require mounts to limit motion in subjects, restricting their use mostly to facial wrinkles [29, 218]. Additionally, SELS and PRIMOS rely on optical imaging of the skin and have not been validated for diverse skin tones, raising concerns about generalizability. In contrast, while GelSight is an optical method, it uses opaque gel coatings and thus is unaffected by skin pigmentation.

Our custom-designed gels, characterized using indentation testing and Hertzian contact theory, achieved Young's modulus values around 129 kPa, helping minimize deformation during measurement. For comparison, skin elasticity reported in the literature varies widely from tens of kPa to over 3 MPa depending on factors such as skin location, hydration, measurement method, and the specific skin layer being assessed [172, 221]. While our gel ensures minimal deformation for less elastic skin, it may be less effective on highly elastic regions. To address this limitation, future work can explore the design of more elastic gels to further reduce skin deformation during contact. In parallel, efforts can focus on directly accounting for skin deformation in

wrinkle depth estimation. For instance, placing a spherical object between the gel and the skin and applying Hertzian contact theory could allow for estimation of local skin elasticity. Alternatively, the skin location could be identified either manually or via a trained classifier using GelSight images or reconstructed height maps and combined with participant-specific data (e.g., age) in a lookup table to estimate elasticity and adjust measurements accordingly.

The core of our reconstruction pipeline combines a CNN-based surface normal estimation model with Poisson integration and high-pass filtering to generate detrended height maps of the skin surface. Our methods were rigorously evaluated using both straight and curved channel objects 3-D printed at micron-scale depths, mimicking the geometry of natural skin wrinkles. On these test structures, our method achieves 12.55 micron mean absolute error (MAE) with high R^2 and ICC values, indicating both accuracy and repeatability. Importantly, our model generalizes well across different geometries and different gels, highlighting the robustness of our approach.

We show the utility of our methods to skin surfaces by contributing wrinkle detection algorithms. We qualitatively displayed the performance of these algorithms for detecting wrinkles in 10 locations on the human body: forehead, upper arm, inside elbow, top of hand, knuckle, top of finger, fingerprint, palm, wrist, and outside elbow. We also presented methods on using this pipeline to automatically extract histograms of wrinkle depth and other wrinkle metrics. In our human study across 15 participants and seven body locations, we observe distinct variations in wrinkle depth based on the location on the human body. As anticipated, regions exposed to frequent movement and mechanical stress, such as the knuckles and fingers, exhibit higher wrinkle depths, while smoother areas like the upper arm show lower values. We also observe that joint orientation affects measured wrinkle depths; for instance, the outside elbow yields lower than expected values due to the skin being stretched when the elbow is bent.

Our work is the first to enable quantitative measurement of wrinkle depth in physical units (microns) across a wide range of body locations and participants. Prior studies have primarily focused on prominent facial wrinkles, such as frontal furrows (horizontal lines), glabellar lines (frown lines), and lateral canthal lines (crow's feet)[35, 49, 155, 219]. While some research has examined fingerprint wrinkle depth, most do not report results in micron-scale units [94, 230]. One study reports wrinkle

depth around 30 microns [82], but only for a single participant. In contrast, our system generalizes to any location where the probe can be applied and is capable of detecting and estimating the depth of all wrinkles, including fine micro-wrinkles.

Lastly, we demonstrate the ability of our probe to detect changes in skin wrinkles due to topical treatments. Across palm, wrist, and elbow sites, wrinkle depth significantly decreases following moisturizer application, with statistically significant differences confirmed using non-parametric testing. The minimal variation between the two pre-moisturizer measurements on the extracted 80th percentile wrinkle depth further underscores the repeatability of our probe and algorithms. These findings validate the potential of our system to detect subtle changes in skin hydration and elasticity and may have applications in dermatological research, cosmetics testing, and longitudinal skin health monitoring.

Overall, our results show that combining tactile imaging with learning-based surface reconstruction and domain-specific analysis pipelines offers a strong approach to quantifying fine-grained skin surface properties. Future work should examine performance of our probe and algorithms in older populations or those with clinically relevant skin conditions (e.g., eczema, psoriasis).

Part II

Interfaces for Control of Physically Assistive Robots

The sensing hardware, methods, and algorithms discussed in Part I support scratch and wrinkle tracking for dermatological conditions, offering valuable insights to inform treatment and promote self-management strategies. However, there is a pressing need for research into systems that deliver actionable physical care, particularly for patients with limited mobility due to their chronic condition. In the following part, we explore how similar sensing technologies, such as accelerometers, can be utilized as interfaces for controlling mobile manipulators to aid in chronic patient care.

The development of novel interfaces has the potential to transform care delivery by enabling individuals to more effectively control assistive robots, increasing patient autonomy while reducing caregiver burden. In the chapters that follow, I present four projects that contribute to the design and evaluation of interfaces for controlling physically assistive robots. These projects span active, shared, and passive modes of interaction, offering a comprehensive exploration of how multimodal interfaces can support individuals with motor impairments in real-world settings.

Chapters 4 and 5 introduce HAT, a head-worn interface based on inertial measurement units (IMUs) that enables users to control mobile manipulator caregiving robots through head movements. Chapter 4 details the design and early evaluation of HAT, including a controlled study with 16 participants without impairments and 2 participants with motor impairments. Chapter 5 builds on this work, introducing a shared control algorithm called Driver Assistance and reporting findings from a week-long in-home deployment with Henry Evans, an individual with severe motor impairments.

In Chapter 6, we present VoicePilot, a large language model (LLM)-integrated speech interface that enables intuitive, flexible voice control of assistive robots. This work highlights how natural language interfaces can lower the cognitive and physical barriers to robot operation.

Finally, Chapter 7 investigates a passive sensing approach to robot control. By using wearable sensors to estimate bite timing during robot assisted feeding, our system enables the robot to infer appropriate moments to act without requiring explicit commands, reducing workload while maintaining the feeling of control.

Together, these four projects demonstrate the value of multimodal interface design in assistive robotics. The work presented in the following chapters is drawn from the following published conference papers, demo paper, and manuscript in preparation:

- Akhil Padmanabha, Qin Wang, Daphne Han, Jashkumar Diyora, Kriti Kacker, Hamza Khalid, Liang-Jung Chen, Carmel Majidi, and Zackory Erickson. Hat: Head-worn assistive teleoperation of mobile manipulators. In *2023 IEEE International Conference on Robotics and Automation (ICRA),* pages 12542–12548. IEEE, 2023. [167]
- Akhil Padmanabha, Janavi Gupta, Chen Chen, Jehan Yang, Vy Nguyen, Douglas J Weber, Carmel Majidi, and Zackory Erickson. Independence in the home: A wearable interface for a person with quadriplegia to teleoperate a mobile manipulator. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction,* pages 542–551, 2024. [168]
- Akhil Padmanabha, Jessie Yuan, Janavi Gupta, Zulekha Karachiwalla, Carmel Majidi, Henny Admoni, and Zackory Erickson. VoicePilot: Harnessing LLMs as Speech Interfaces for Physically Assistive Robots. In *The 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24),* October 13–16, 2024, Pittsburgh, PA, USA. ACM, New York, NY, USA, 18 pages. [169]
- Jessie Yuan, Janavi Gupta, Akhil Padmanabha, Zulekha Karachiwalla, Carmel Majidi, Henny Admoni, and Zackory Erickson. Towards an LLM-Based Speech Interface for Robot-Assisted Feeding. In *Adjunct Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (UIST '24), October 13–16, 2024, Pittsburgh, PA, USA. ACM, New York, NY, USA.* [242]
- Akhil Padmanabha, Jessie Yuan, Tanisha Mehta, Rajat Kumar Jenamani, Victoria De León, Anthony Wertz, Janavi Gupta, Ben Dodson, Carmel Majidi, Tapomayukh Bhattacharjee, and Zackory Erickson. A Wearable Approach to Bite Timing Estimation in Robot-Assisted Feeding. *Manuscript in Preparation,* 2025.

Chapter 4

HAT: Head-Worn Assistive Teleoperation of Mobile Manipulators

4.1 Contributions

For the following chapter, the contributions are as follows. Jashkumar Diyora formulated the preliminary ideas for a head-worn device and created an early prototype with Kriti Kacker, Hamza Khalid, and Liang-Jung Chen with advising from Zackory Erickson. I designed and built the device highlighted in the following work. Qin Wang and I collaborated on building the final system and running the human studies. Daphne Han assisted with testing the final system and study design. I wrote the IRB protocol and the paper draft. Zackory Erickson and Carmel Majidi provided guidance and advising on all aspects of the project.

4.2 Introduction

Motor impairments and loss of hand function can restrict an individual from performing activities of daily living (ADLs), such as eating, self hygiene, and dressing. Many individuals with lost motor function rely on assistance from caregivers, limiting their self-sufficiency, independence, and control in self-care and household tasks [65, 143, 197]. Motor impairments affect a significant subset of the United

States with approximately 1.7% of the population, over 5 million people, reportedly living with some form of paralysis [18]. Loss of motor function can be due to a range of neurodegenerative diseases, stroke, muscle atrophy, and spinal cord injury (SCI). As an example, more than 100,000 individuals in the United States currently live without the ability to use their hands because of a cervical SCI [15, 18, 51]. Many individuals with SCI lose some or all ability to use their upper and/or lower extremities, but often retain motion in their head and neck [163]. For many with tetraplegia or limited hand function, technologies that restore the ability to perform everyday tasks can have an immediate and profound impact [32, 41, 135, 164, 210]. Teleoperation of mobile manipulator robots can enable individuals with motor impairments to once again perform physical tasks including self-care, fetching objects, etc. [37, 40, 86, 116].

While developing teleoperation interfaces for individuals with impairments, it is important to consider ease of use, efficiency (performing a task with a minimal number of steps in a short amount of time), and comfortability. Most interfaces have physical requirements, determining which subset of individuals with impairments will use them; for example, conventional hand-operated joysticks provide precise, continuous control but require users to have motor function in their hand. Meanwhile, web interfaces, the current convention for teleoperation, rely on the use of a device with a screen, requiring the ability of an individual to move a cursor and click on buttons through an assistive device, often requiring fine motor control. The development and evaluation of novel assistive interfaces for mobile manipulators could lead to alternatives for individuals with impairments who may have a difficult time accessing traditional systems.

In this work, we present a novel inertial-based assistive interface for individuals with motor impairments to directly control a lightweight mobile manipulator using residual head motion. The inertial-based interface, embedded in a baseball hat as shown in Fig. 4.1, sends absolute orientation angles of the user's head to a mobile manipulator. The angles are mapped to velocity commands for actuation of the robot's base, arm, and end effector. We evaluated the interface using both quantitative and qualitative metrics through a human study with 16 able-bodied participants and 2 participants with motor impairments. Participants use this interface to control a mobile manipulator and complete four self-care and household tasks. Our results show that the head-worn device is an efficient and intuitive control interface for a mobile

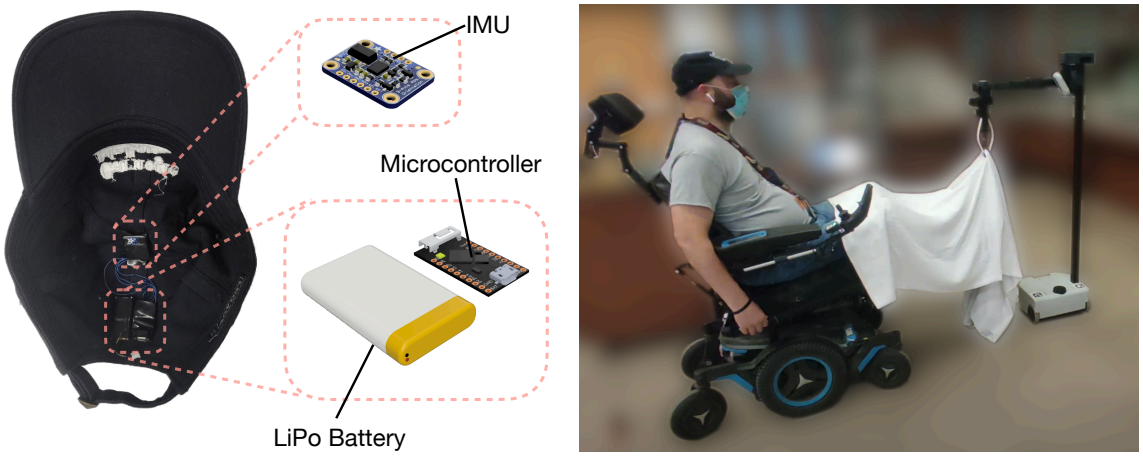


Figure 4.1: Left: The head-worn interface with integrated inertial measurement unit (IMU) sensing to control an assistive robotic manipulator. The hat consists of an absolute orientation IMU connected to a microcontroller with builtin Bluetooth capabilities and powered by a small lithium polymer (LiPo) battery. A thin layer of Neoprene foam is added over the electronics for comfort. Right: A participant (I1) with spinal cord injury completing a blanket removal task using the head-worn interface.

manipulator, presenting individuals with motor impairments a strong alternative to conventional teleoperation platforms. The contributions of this work is as follows:

- We introduce a novel assistive interface with inertial measurements integrated into an everyday clothing article.
- We present how this inertial head-worn interface can enable teleoperation of a high degree-of-freedom mobile manipulator using only residual head motion retained after cervical spinal cord injury (SCI) or tetraplegia.
- We evaluated this wearable interface with 16 able-bodied participants and 2 participants with motor impairments and show that the head-worn device enabled participants to complete self-care and household tasks with low error rates, high perceived ease of use, and low workload measures.

Build instructions and code for this interface are open-sourced on the project website[1].

[1] https://sites.google.com/view/hat-teleop/home

4.3 Related Work

Current assistive interfaces for wheelchairs and mobile robots include hand and mouth operated joysticks [63], and web-based visual interfaces [33, 37, 79] which can be used in conjunction with other assistive devices such as head tracking [24, 180, 214], eye tracking [95, 107], etc. These web interfaces enable control of robot motions and often incorporate sub-modules that allow users to select autonomous routines such as grasping of an object. Researchers have evaluated several iterations of these teleoperation platforms with both healthy participants and one participant with tetraplegia using a PR2 robot [37, 40, 86]. This interface enabled control of the robot for various tasks including grasping in cluttered environments, handing out candy, opening of a drawer and extracting an object, etc. Efforts have been made to make these interfaces more intuitive, by overlaying movement buttons directly over camera feeds and providing additional views of the environment, and to evaluate them with more individuals with motor impairments [33, 79].

Researchers have previously explored teleoperation of robots using an IMU-based device that captures hand and arm movements [62, 126]. While highly efficient, these interfaces cannot be used by individuals without intact limb movement. Prior work also includes head-worn IMU-based interfaces for control of electric wheelchairs [31, 73, 140], but has not shown generalization to mobile manipulators that have many additional degrees of freedom.

Lastly, brain computer interfaces (BCI) show promise for efficient control of physical devices for individuals with motor impairments. Researchers have demonstrated control of a robotic arm for tetraplegics using embedded neural interfaces [90]. Even though potentially applicable to a large population, these BCI techniques are often invasive and need to be extensively calibrated with each new user. As an alternative to BCI, EEG and EMG signals from skin-mounted electrodes are non-invasive and have been used to teleoperate simple robot motions based on residual myoelectric signals [66, 237]. In contrast, a head-worn noninvasive assistive interface may be generally applicable to a wide population of individuals who have lost significant motor function below the neck, as is the case for many individuals with cervical SCI or tetraplegia due to neurodegenerative disease, stroke, or injury.

4.4 Device Design

In Fig. 4.2, signals from the head-worn interface are communicated to the mobile manipulator via Bluetooth and mapped to velocity commands for the robot's actuators. The interface is used for direct control of the robot's motion. Speech recognition, using audio captured by a wireless microphone worn by the participant, is used for selection of four robot modes: drive, arm, wrist, and gripper, as shown in Fig. 4.2A. The head movements are mapped to robot motions based on the mode the user is in; for example, the arm mode mapping is shown in Fig. 4.2B.

The inertial head-worn interface, shown in Fig. 4.1, consists of an Adafruit BNO055 absolute orientation inertial measurement unit (IMU) which fuses accelerometer, gyroscope, and magnetometer data and outputs absolute roll, pitch, and yaw orientation angles. A Tiny PICO microcontroller is used to sample the IMU at 20 Hz and to send this data to the mobile manipulator over Bluetooth.

4.4.1 Mode Switching

For the human study, participants use speech recognition to enable control of the robot by the head-worn interface and to switch between the four robot modes. While in each mode, users can command specific robot motions by tilting their head along the roll or pitch axis. The speech commands are "start", "switch to drive", "switch to arm", "switch to wrist", and "switch to gripper". To trigger the speech recognition, participants shake their head lightly to the right and left, along the Z-axis, shown in Fig. 4.2B. The participants receive audible confirmation, either the recognized command or "repeat" if the phrase is unidentifiable. We foresee mode switching to be one of many plug-and-play options including a clicker, sip/puff device, or even head motions along the Z-axis based on a user's ability and preference.

4.4.2 Calibration and Robot Mapping

The robot receives mode-switching commands and interface orientation angles over Bluetooth. The robot pushes commands to its actuators at 10 Hz. Before controlling the robot, users must calibrate the initial IMU orientation by triggering voice recognition and saying "start". During calibration, the orientation of the head-worn interface

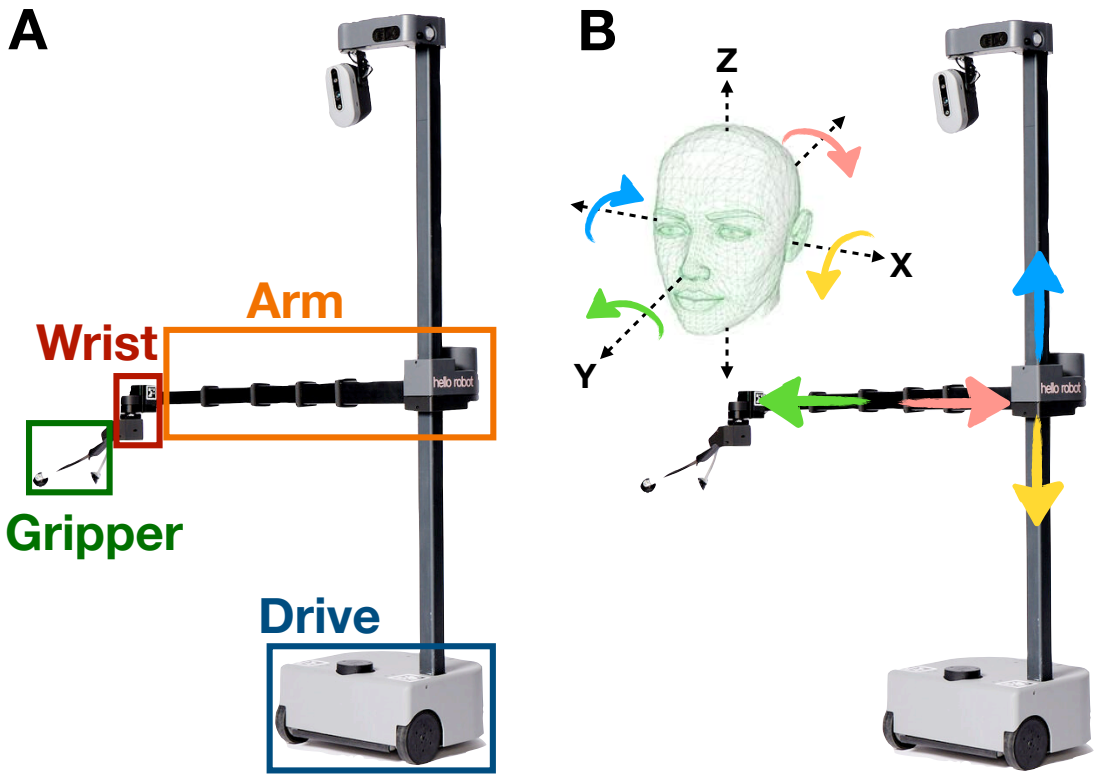


Figure 4.2: Mapping from the head-worn interface to the robot. (A) The 4 separate modes of the robot controlled by the interface. (B) Visual representation of the mapping from head movements to robot motion in the arm mode.

is saved and used for setting four thresholds. The minimum motion thresholds, $t_{l,a}$ and $t_{l,b}$, are set at 15° and −15° respectively from the calibrated position, θ_c. The maximum thresholds, $t_{h,a}$ and $t_{h,b}$, are set at 45° and −45° respectively from the calibrated position, θ_c. If the user's head is tilted less than $t_{l,a}$ and greater than $t_{l,b}$, in both the X and Y axes, the robot stops all motion. For a single axis (either X or Y), the angle measurement is proportionally scaled to velocity commands for the

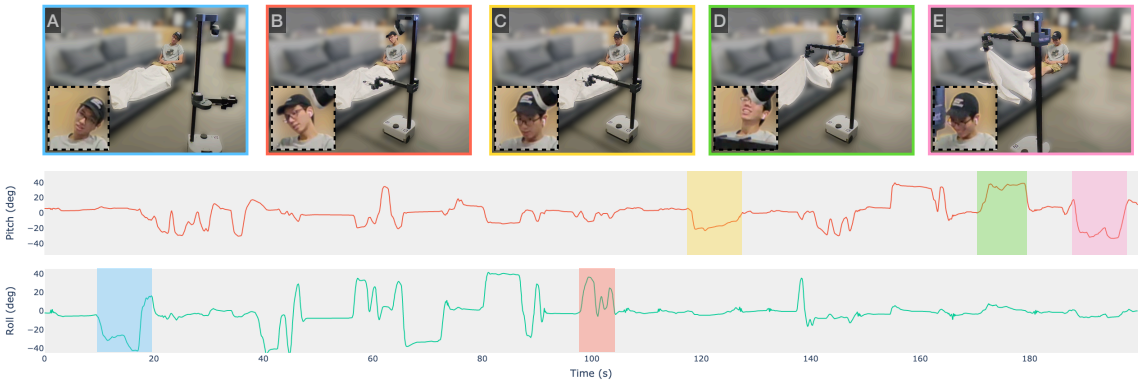


Figure 4.3: Time lapse of a participant dragging a blanket off his legs with corresponding signals from the IMU along the pitch and roll axes. The highlighted regions of the roll and pitch signals correspond to the matching colored video frames and correlate to the head motions that the participant performs to accomplish a specific robot motion. (A) Participant tilts their head to the left to make the robot turn towards their body. (B) After positioning the robot's base, the participant switches to arm mode and tilts their head to the right to extend the arm. (C) The participant switches to gripper mode and pitches their head down to open the gripper. (D) After grasping the blanket, the participant switches to arm mode and tilts their head up to raise the robot's arm up. (E) Lastly, the participant drives the robot forward by pitching their head forward.

robot actuators according to the following equation:

$$V(\theta, \theta_c, a) = \begin{cases} -v_{a,max} & \text{if } t_{h,b} > \theta \\ k_a(\theta - t_{l,b}) & \text{if } t_{l,b} \geq \theta \geq t_{h,b} \\ 0 & \text{if } t_{l,b} < \theta < t_{l,a} \\ k_a(\theta - t_{l,a}) & \text{if } t_{l,a} \leq \theta \leq t_{h,a} \\ v_{a,max} & \text{if } t_{h,a} < \theta \end{cases}$$

where $V(\theta, \theta_c, a)$ is the velocity command sent to actuator a, θ is the angle of the user's head, θ_c is the initial angle of the user's head during calibration, k_a is the proportional constant for actuator a, $v_{a,max}$ is the maximum velocity limit for actuator a , $t_{l,a} = 15° + \theta_c$, $t_{h,a} = 45° + \theta_c$, $t_{l,b} = -15° + \theta_c$, and $t_{h,b} = -45° + \theta_c$.

In drive mode, pitch of the head controls forward and backwards motion of the robot's base, whereas roll controls rotation clockwise and counterclockwise. In arm mode, shown in Fig. 4.2B, pitch controls the robot arm height and roll controls the extension. In wrist mode, pitch controls the wrist's pitch and roll controls the wrist's yaw. Lastly, in gripper mode, pitch commands the gripper open and close. The yaw (Z) axis of the IMU is only used for triggering speech recognition and not for any robot motion, allowing the user to rotate their head in this axis to look at objects in their environment. A time lapse of a participant completing a blanket task from the human study is shown in Fig. 4.3 with corresponding pitch and roll angles from the head-worn interface.

4.5 Study Design

The human study was conducted with 16 healthy participants (11 M, 5 F) with ages ranging from 18-27 (Mean = 22.8, STD = 2.5) and two participants with motor impairments. We obtained informed consent from all participants and approval from the Carnegie Mellon University Institutional Review Board. Participants were asked how much experience they have with controlling a robot on a 5-point scale ranging from "1 = No Experience" to "5 = Expert User". Participants had a mean experience of 2.2 (STD = 1.2, Median = 2).

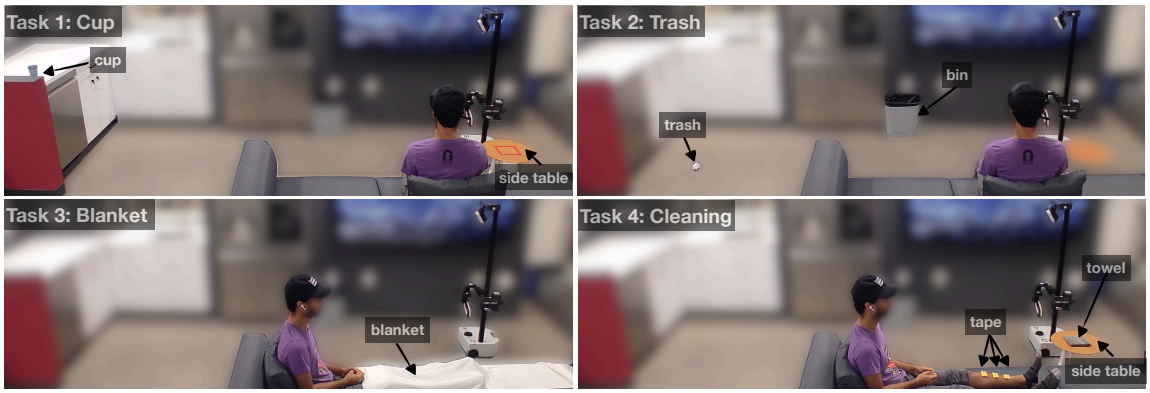


Figure 4.4: Setup of the 4 study tasks: Cup, Trash, Blanket, and Cleaning. The robot starts in the same position for each task, homed and with the wrist stowed. For Task 1 (Cup), participants must use the robot to grab a cup from a kitchen counter and place it within a square target on the side table. For Task 2 (Trash), participants are instructed to grab a piece of trash (crumpled A4 printer paper) on the floor, and place it in a trash bin. For Task 3 (Blanket), participants are told to remove the blanket from their legs using the robot. The task ends when the blanket is not touching any part of the participant. Finally, for Task 4 (Cleaning), participants must grab a towel from the side table and use it to clean 3 pieces of "dirt" (paper tape pieces) stuck to their left leg. The task ends when all pieces of tape are unstuck and moved from their original position.

The participants are instructed to do 4 tasks (cup retrieval, trash pickup, blanket removal, and leg cleaning / hygiene) by controlling the robot in a replicated home environment using the head-worn interface. The tasks are ordered in difficulty and shown visually in Fig. 4.4. Participants were allowed to ask researchers to reset the environment in cases when an object is dropped or they think it would be easier to restart; however, the task timer was not reset. Participants were given 14 minutes to complete each task and are instructed to keep their body still and only move their head and neck.

The head-worn interface serves as an alternative to a web interface for people who have a difficult time accessing traditional computing systems. Thus, for Task 1 (Cup), we additionally have participants use the modified Stretch web interface with modifiable speed control [33] as a baseline. The order of the web and head-worn interface for Task 1 was alternated for every participant. For the web interface, participants used a laptop placed in front of them with a head tracking software[2] to move the cursor and a single button mouse to click. Head tracking software is a common interface used by many individuals with head control but limited hand function to interact with computational devices [24, 180, 214]. All healthy participants used head tracking software with the web interface to standardize control between the head-worn and web interfaces.

Prior to the study, participants are led through practice sessions for both interfaces. Participants watched an instructional video for each interface[3] and were led through a series of commands available. For the web interface, participants watched the instructional video from [33]. Finally, participants completed a practice task, grabbing a wooden block on a kitchen counter, while the researchers provided tips. After the practice sessions, participants are not provided any help with using the interfaces.

After each completed task, participants answer two 7-point Likert items provided in Fig. 4.7. At the end of the study, participants answered the six question NASA Task Load Index (TLX), which measures perceived workload, and 7-point Likert items provided in Fig. 4.8. A researcher also asks participants the following open-ended questions for each interface: (1) What did you like about the interface? (2) What were the cons of the interface? (3) What were the reasons for any errors you encountered?

[2]Enable Viacam (https://eviacam.crea-si.com/index.php)
[3]HAT Instructional Video (https://youtu.be/v8wXM-cCss0)

(4) What were the biggest challenges while using the interface? Lastly, participants are asked how the two interfaces compare to each other.

4.5.1 Participants with Motor Impairments

The procedure described above was additionally followed for two participants with impairments, who were given the choice to conduct the study either in the same location as healthy participants or in an alternate location of their preference. Participants were allowed to use a device of their choosing for controlling the web interface and had no time limits to complete each task.

Participant I1 (M, 36) is paralyzed below the waist and additionally has limited movement in his hands and neck due to spinal cord injury. He is an expert robot user (marked 5 on survey) and has participated in robotics studies before. He chose to use a computer mouse for the web interface as he has enough hand control to do so, but uses his hand to operate the mouse in a non-standard manner. He conducted the study in an alternate location.

Participant I2 (M, 63) has limited hand function due to an essential tremor, a neurological disorder that causes involuntary shaking primarily while writing and using a keyboard. He has no prior experience with controlling a robot (marked 1 on survey). He chose to use a computer mouse for the web interface, but mentioned occasionally having hand tremors while doing so. He conducted the study from the same location as healthy participants.

4.6 Results and Discussion

4.6.1 Task Completion and Efficiency

The vast majority of participants in both groups successfully completed all head-worn interface tasks within the allotted time. Participant I1 and I2 are shown doing tasks from the study in Fig. 4.6. Task completion times are shown in Fig. 4.5. For Task 1 (Cup), the proposed head-worn interface allowed for faster task completion in comparison to the web interface with 12 out of 16 healthy participants and both participants with impairments finishing Task 1 faster. On average, Task 1 with the

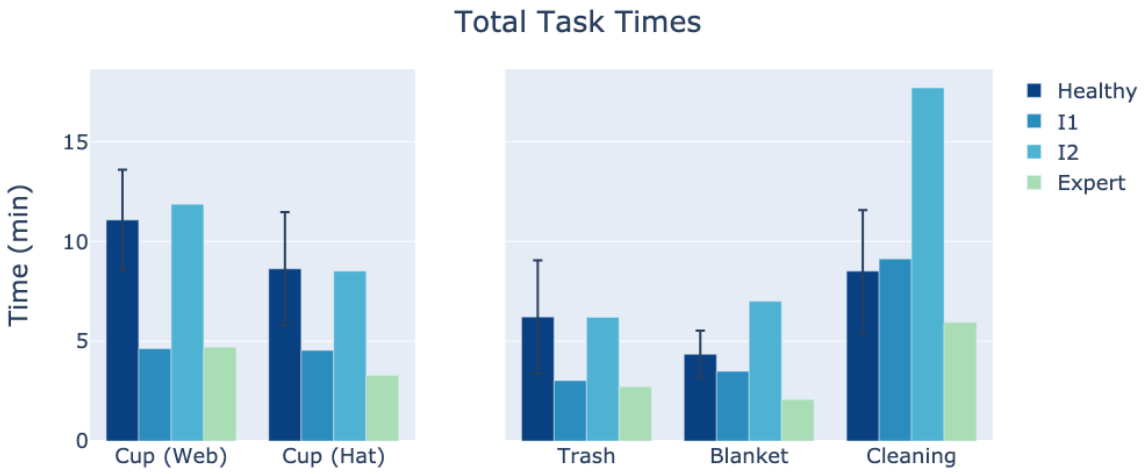


Figure 4.5: Average task times from 16 healthy participants, 2 participants with impairments, and 1 expert user. 4 healthy participants were unable to finish Task 1 (Cup) with the web interface due to reaching the time limit and 1 healthy participant additionally failed to complete Task 2 (Trash) and Task 4 (Cleaning) in the allotted time. If a task is not completed, the task time is recorded as 14 minutes.

head-worn interface took 147 seconds less than with the web interface for healthy participants as shown in Fig. 4.5. When asked the Likert item "I was able to complete this task efficiently using the control interface", shown in Fig. 4.7, 15 out of 16 healthy participants and both participants with impairments ranked the head-worn interface the same as the web interface or higher, with a median ranking of 6 (Agree) and 4 (Neutral) by each group respectively. By applying a Wilcoxon signed-rank test, we observed a statistically significant difference between task times ($p = 0.04$) and between Likert item responses on task efficiency ($p = 0.0016$) for the head-worn and web interfaces on Task 1 with healthy participants. The head-worn interface was more efficient as it allowed participants to have continuous control of robot motion and speed (i.e. degree of head tilt resulting in more/less speed), thus requiring less commands than the web interface, which only allowed clicks of constant distances and selection of speed through buttons.

As displayed in Fig. 4.7, participants agreed with the Likert item on task efficiency for the three remaining tasks with the head-worn interface. Note that participants with motor impairments were able to complete most of the tasks in similar times to healthy participants as shown in Fig. 4.5. Task 4 (cleaning) was the most challenging for participants (both healthy and with impairments) due to the added difficulty



Figure 4.6: Left: Participant I1 completing Task 4 (Cleaning) using the head-worn interface. Right: Participant I2 completing Task 3 (Blanket) using the head-worn interface.

of wiping off tape that was firmly attached to their leg. Despite the increased task time, all participants were successful in grabbing the towel from the side table and wiping their leg. Additionally seen in Fig. 4.5 are the times achieved by an expert user, a member of the research team, for performing the same tasks; these results serve as a baseline for what task times are achievable with sufficient experience with the interface. At the end of the study, all healthy participants agreed that "The [head-worn] interface enabled control of the robot in a reasonable amount of time" with a median score of 6 (Agree) and both participants with impairments reported 7 (Strongly Agree), as detailed in Fig. 4.8.

4.6.2 Errors and Recovery

After each task, when asked the Likert item "I was able to complete this task without any errors using the control interface", median results show that both healthy participants and participants with motor impairments agreed for all head-worn interface tasks, as shown in Fig. 4.7. Nonetheless, some errors were observed with participants operating the head-worn interface. Common errors included participants tilting their head too far resulting in the robot moving too fast and overshooting the intended position. Occasionally, participants tilted their head the opposite direction of their intended action; I2 summarized this well saying, "Biggest thing and it's not

unique to this interface is keeping straight which direction is which when you're not oriented the same way the vehicle is.". In contrast, 3 healthy participants and 1 participant with impairments mentioned liking the buttons of the web interface which would need to be clicked to cause robot motion, thus resulting in less inadvertent movements of the robot. For example, P11 mentioned "I had to click the button to do something. [There was a] low chance of any accident.". In most cases, these mistakes with the head-interface were quickly recognized by the participants and were easy to recover from. With regards to overshooting, 2 healthy and both participants with impairments mentioned similar downsides with the web interface as they were unable to predict how far the robot would move with each click of the web interface, especially after changing control speeds. For the head-worn interface, a calibration procedure, visual cues, and additional practice could help address these common sources of error identified through the human study. At the end of the study, we presented the following 7-point Likert item to participants regarding the head-worn interface: "The control interface allowed easy recovery from errors". As shown in Fig. 4.8, healthy participants agreed with a median ranking of 5.5 while participants with impairments agreed to the statement with a median response of 6.5 (out of 7).

4.6.3 Ease of Use

Another important factor to consider with the development of an assistive interface is perceived ease of use [139, 225]. Participants noted the continuous control of the robot using the head-worn interface which allowed for intuitively varying the velocity of the robot's actuators in proportion to the magnitude of head tilt. When participants were presented with the statement "The control interface was easy to use" (see Fig. 4.8), healthy participants agreed with a median ranking of 6 while participants with impairments agreed with a median ranking of 6.5 out of 7. Additionally, in the qualitative interviews, 9/16 healthy participants and both participants with impairments mentioned, without solicitation, that control using the head-worn was intuitive or more intuitive than the web interface. I2 likened the interface to a Segway, "If you wanted to go forward, you would pitch forward... so I would say strangely the hat was more intuitive."

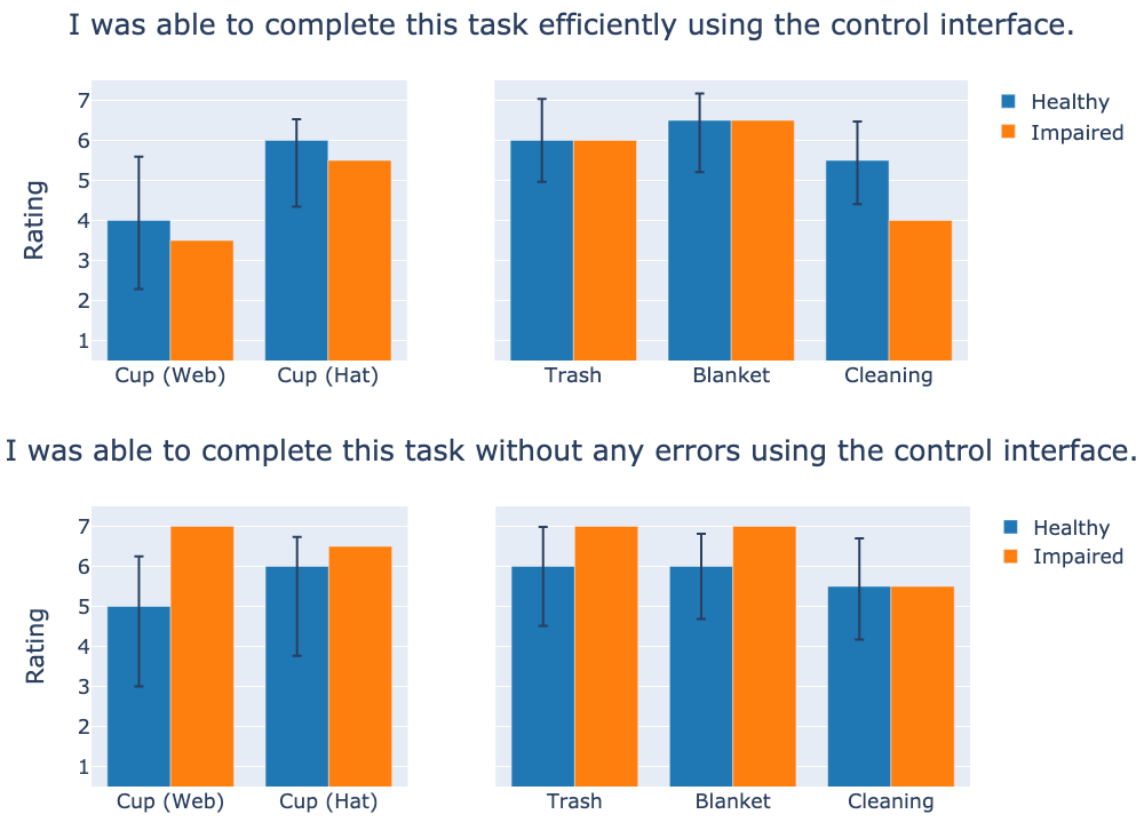
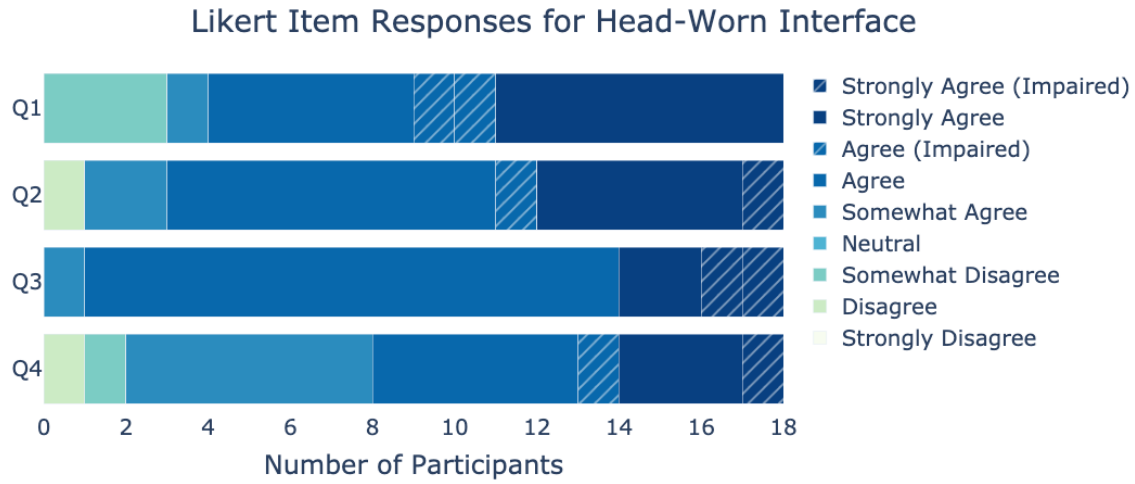


Figure 4.7: Median results from the two 7-point Likert items that participants are asked to evaluate after completion of each task.

4.6.4 Learning Curve

For the head-worn interface, both healthy participants and participants with impairments agreed that "The control interface was easy to learn", as seen in Fig. 4.8, with a median reported score of 6 (Agree) for both groups. 4 out of the 16 healthy participants felt that the web interface was faster to learn, but mentioned that use of the head-worn interface improved with more practice. One participant (P8) noted at the end of the study: "If I had never seen any of this before and you wanted me to do one thing and that was it, I would take the web interface but with time to practice, the hat was definitely way better".



Q1: The control interface was easy to learn.
Q2: The control interface was easy to use.
Q3: The control interface enabled control of the robot in a reasonable amount of time.
Q4: The control interface allowed easy recovery from errors.

Figure 4.8: Results from the 7-point Likert items presented to participants after completion of all tasks using the inertial head-worn interface. The majority of participants agree with all four statements about the proposed interface.

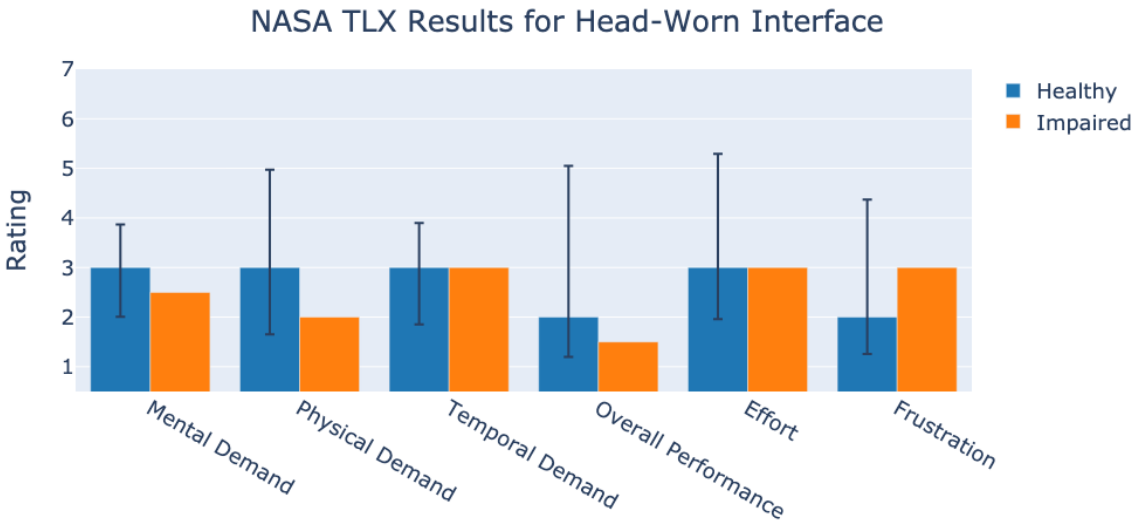


Figure 4.9: Median results from the NASA TLX survey asked to participants after completion of all tasks using the head-worn interface. A rating of 1 is best for all categories.

4.6.5 Workload

Workload and perceived effort are also important metrics to evaluate for new interfaces [74, 144]. The NASA TLX [85] is a 7-point scale that measures mental, physical, and temporal demand, overall performance, effort, and frustration. A rating of 1 is best for all categories. The median results are located in Fig. 4.9. The overall performance of the interface was strong for both healthy participants and participants with impairments with a median ranking of 2 and 1.5 respectively.

For mental demand, healthy participants and participants with impairments rated the head-worn interface as 3 and 2.5 respectively. However, 5 healthy participants mentioned facing mental demand as the head-worn interface required them to recall more commands in contrast to having them displayed on the computer screen in the web interface. Additionally, both participants with impairments mentioned demand associated with remembering to keep their head still while not trying to command robot motions. For physical demand, healthy participants and participants with impairments rated the head-worn interface as low with 3 and 2 respectively.

4.6.6 Challenges and Future Directions

Perception was a challenge for participants while using the head-worn interface. In Task 1, the lack of depth perception was an issue due to the positioning of the cup far away on a kitchen counter. In qualitative interviews, 11/16 healthy participants highlighted that the camera feed of the web interface helped with this problem and a few vocalized that a combination of the head-worn interface and the visual cues from the web interface would work the best. In the qualitative interview, 6 healthy participants and 1 participant with impairments mentioned preferring interaction directly with the robot through line of sight. One participant (P14) said, "When I'm trying to operate the robot with [the head-worn interface], I'm still a participant in the scene, experiencing reality, but when I'm looking at the screen that's taking all my attention." In addition, having some visual interface with a camera feed is necessary in situations where the robot is used out of line of sight, such as in another room. Participants also mentioned issues with maintaining vision of objects while tilting their head. This was specifically a problem while closing the gripper which required participants to tilt their head up, restricting line of sight to objects like the

blanket or towel that the robot was manipulating near the person.

A few participants also expressed a desire for more visual feedback with the head-worn interface, specifically a diagram to remind them of the direction to tilt their head and to show them where the thresholds were. Future work could explore providing participants with these visual cues. Several participants faced occasional difficulty using speech recognition, which presents a future opportunity to provide users with other methods for mode switching. Both participants with motor impairments also mentioned issues with depth perception and mode switching. I1 additionally mentioned issues with his impairment limiting head movement in the roll (Y) axis, preventing him from going as fast as he would have liked. He elaborated that a calibration procedure could help make the tilting range for the roll axis smaller, allowing him to reach faster speeds.

4.7 Conclusions

This work presents a unique inertial-based wearable assistive interface, embedded in a familiar head-worn garment, for individuals with motor impairments to teleoperate and perform physical tasks with a mobile manipulator. The interface presents individuals with impairments an alternative to conventional computer based teleoperation platforms. Our results from a study with both able-bodied and participants with impairments show that the head-worn interface is both intuitive, and efficient, allowing users to perform a variety of physical self-care and household tasks with few errors. Our study results contrast the head-worn interface with a conventional computer-based assistive web interface and we observed benefits in terms of efficiency and ease of use.

Our work, as previously noted, has certain limitations, including the absence of long-term testing with individuals with motor impairments in their home environments. To address these challenges, the next chapter presents findings from a week-long study conducted in the home of an individual with severe motor impairments. Additionally, we focus on enhancing performance by reducing workload and task completion times through the implementation of shared autonomy algorithms, which are detailed in the following section.

Chapter 5

Independence in the Home: A Wearable Interface for a Person with Quadriplegia to Teleoperate a Mobile Manipulator

5.1 Contributions

I formulated the core ideas for this project with Zackory Erickson and Carmel Majidi. I designed the device and led the design and implementation of the system. Janavi Gupta designed the human study and helped design, implement, and test the system. I assisted in the design of the human study and executed it with Janavi Gupta. Chen Chen, Jehan Yang, and I worked on the Driver Assistance algorithm with Chen Chen handling all the implementation and integration with the system. I wrote the IRB protocol and the paper draft with help on specific sections from Chen Chen, Janavi Gupta, and Jehan Yang. Vy Nguyen facilitated recruitment, provided feedback on the design of the human study and the paper, and provided the robot tools. Zackory Erickson and Carmel Majidi advised on all aspects of the project.

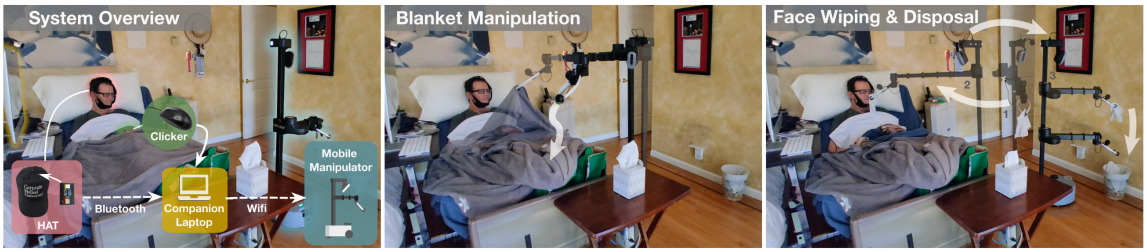


Figure 5.1: Henry, a non-speaking individual with quadriplegia, uses HAT, a head-worn assistive interface, to teleoperate a mobile manipulator to move a blanket below his knees, grasp a tissue to wipe his face, and throw the tissue in a trash can. The interface records and wirelessly sends Henry's head orientation angles to a companion laptop, which computes and transmits actuator velocities to the robot. A clicker is used by Henry to switch between modes, distinct operational states of the HAT system.

5.2 Introduction

In this work, we build on our existing, inertial measurement unit (IMU) based interface, HAT (Head-Worn Assistive Teleoperation), which can be embedded in any head-worn clothing article such as a baseball cap [167]. HAT serves as a direct teleoperation interface which maps head orientation angles to actuator velocities for a high degree of freedom mobile manipulator, Hello Robot's Stretch [110]. HAT can be used by individuals with head motion, which can often be retained after cervical spinal cord injury or quadriplegia [163]. Research has suggested that head-worn interfaces for mobile manipulators may offer individuals with impairments a strong alternative compared to traditional web-based interfaces [167].

We note several limitations in head-worn control interfaces that we aim to address in this research. First, prior evaluations with HAT were conducted through 2-hour long studies primarily with non-disabled individuals in confined lab environments [167]. Thus, to evaluate longer term usage of the interface in a real-world setting, we conducted and present results from a 7 day in-home case study with Henry Evans, an individual with quadriplegia. Case studies (N=1) are a proven research methodology in the fields of medicine, human computer interaction, and human robot interaction, that can provide in-depth analysis in a particular process or technology [37, 71, 146, 156, 241]. Second, despite participants' impression of a head-worn interface as efficient [167], task times were still much higher than if a caregiver or non-disabled

individual completed the task. To reduce these long task times, overshooting errors, and effort for grasp alignment especially with reduced depth perception, we introduce Driver Assistance for mobile manipulators, inspired by driver assistance features in motorized vehicles. Driver Assistance, in our case, is a shared control method for precision grasping of household objects. Lastly, the need for customization in a wearable interface is paramount as impairments vary across a spectrum encompassing a wide range of conditions and physical abilities. Thus, we introduce multiple wearable designs, capture an additional degree of head motion for robot control, and integrate a clicker for mode switching. We further generalize these results by connecting HAT to a GUI that allows for robust robot control out of line-of-sight and for control of other physical devices in the home.

The contributions of this work are as follows:

- We evaluate an inertial-based assistive interface, HAT, with a mobile manipulator in a week-long case study with an individual with quadriplegia in their home environment and assess a variety of metrics.
- We introduce Driver Assistance, a form of shared control for mobile manipulators, and show that, for our participant, it reduces task times and effort while still preserving the perception of control.
- We demonstrate the customizability and generalizability of the interface and show how it can adapt to individuals with severe impairments, to different physical devices in the environment, and to different contexts.

5.3 Related Work

5.3.1 Teleoperation Interfaces

Many forms of assistive interfaces, including hand/mouth operated joysticks, have been evaluated with mobile robots and wheelchairs [63, 178]. Web-based interfaces [33, 37, 79, 189], the prevailing norm for teleoperation of robots, enable direct control of actuators using screen-based buttons and camera feeds. These interfaces can be used with a standard mouse but are also compatible with alternate assistive devices such as head tracking [24, 180, 214] and eye tracking [95, 107]. Over the last decade,

numerous web-based teleoperation interfaces have been developed and evaluated with individuals with and without motor impairments [33, 37, 40, 79, 86].

5.3.2 IMU-based Interfaces

IMU-based interfaces have previously been explored for control of mobile robots [62, 126]. While highly efficient, these interfaces cannot be used by individuals without intact limb movement. Head-worn IMU-based interfaces have been used in conjunction with electric wheelchairs [31, 73, 140], but most have not been evaluated with mobile manipulators that have many additional degrees of freedom and which require the user to teleoperate the mobile base in third person. HAT is an IMU-based, direct teleoperation interface which maps head orientation angles to actuator velocities for a high degree of freedom mobile manipulator [167]. In the past study, participants used head tilts along the pitch and roll axes to control robot velocities. They used speech to switch between modes, which dictated the subset of actuators they could control with their head motions. Researchers tested the interface with 16 non-disabled and 2 participants with motor impairments in 2-hour study sessions per participant. Participants generally found the interface efficient and error-recovery friendly. Ease of use and learning curve were rated positively, and most participants reported low workload measures.

5.3.3 Shared Control

Shared control (SC) involves a robot adjusting its level of autonomy based on its comprehension of human intentions and the surrounding environment [17, 83, 201], resulting in a reduction in cognitive load for the user [75, 132, 145]. SC has been explored in diverse contexts, such as manipulators/humanoids [187, 233, 249], wheelchairs [59, 208], and surgical robots [234, 246]. Research has also examined SC across varying levels of autonomy [83]. Past works typically infer the user's intention via their input and provide assistance through control blending [97, 132, 201]. However, few SC methods have been tested in complex real-world environments, such as household settings. While developing SC methods, the user's feeling of control over the system is an important consideration as previous studies show that higher levels of autonomy can lead to a reduced feeling of control [27, 98]. In comparison to

past work, our proposed SC method, Driver Assistance, limits the shared control to specific robot joints to exemplify the human perception of control.

5.4 Methodology

We conducted an evaluation of the system with a single participant, Henry Evans (M, 62). Henry faces unique challenges as a non-speaking individual with quadriplegia due to brainstem stroke. He experiences some limitation in his neck's range of motion and he has slight flexion of his left thumb, allowing him to click a single button on a computer mouse. Henry has accumulated extensive robotics experience through his involvement in numerous studies with web-based interfaces [8, 37, 40, 79, 86, 106, 158] and assistive robots. Our study, approved by Carnegie Mellon University's Institutional Review Board, spanned a period of seven consecutive days from 8/14/2023 to 8/20/2023 in Henry's home environment. Recruitment was conducted over email and informed consent was obtained from Henry for use of name and image in publication. Before and throughout the study, we adhered to user-centered design principles, actively seeking input from Henry at various stages of system development. This process led to improvements to existing functionality and the creation of new features for HAT.

5.4.1 Device Design

As seen in Fig. 5.1, the system consists of HAT, a companion laptop, and Hello Robot's Stretch RE2 robot. HAT consists of a clothing article and a foam pad on which the electronics are mounted on. Velcro is glued to the pad and can be used to secure it to a head clothing article of the user's choosing. The IMU, attached such that it remains flat at the crown of the head, estimates and sends roll, pitch, and yaw absolute orientation angles at 20 Hz to the companion laptop over Bluetooth.

We evaluated three different head-worn garments (baseball cap, cloth headband, and silicone chin strap) for this study. The chin strap, suggested by Henry, is traditionally used to reduce snoring or minimize chin wrinkles. Henry proposed the chin strap for two reasons: 1. Its secure fit, preventing movement during spasms 2. Comfort due to its design without material at the back of the head. After

testing, Henry chose to use the chin strap for the study with slight modifications to enlarge the ear holes. Throughout the week, we discovered two additional benefits: it prevented his glasses from slipping and effectively reduced drooling by keeping his mouth comfortably closed.

5.4.2 Mode Switching

Mode switching involves transitioning between distinct operational states within the HAT system, each offering the user specific functionalities. For example, while in each of the robot control modes (drive, arm, and wrist) shown in Fig. 5.2, Henry can command a subset of robot actuators with head motions. For this study, we integrated Henry's clicker, which is a standard computer mouse shown in Fig. 5.1, as he is non-speaking. The clicker is plugged directly into the companion laptop, which announces the mode change with an audio notification. We additionally added two modes, idle and cursor control, to expand the available functionality. In idle mode, the HAT is deactivated allowing the participant to move their head freely without causing robot motion. In cursor control mode, head motions can be used to control the cursor on the companion laptop. To switch between the modes, Henry can use 4 distinct clicking patterns: single click, double click, triple click, and holding down the clicker. We conducted two rounds of iteration on these clicking patterns during the study at the start of Day 3 and Day 6, incorporating Henry's feedback. More information on the clicking patterns can be found in the published work [168]. The integration of an alternative mode-switching method and the ability to easily modify clicking patterns demonstrates the flexibility of the HAT system for customization.

5.4.3 Robot Mapping and Thresholds

At the start of the study, we recorded Henry's max head rotations in the roll (Y), pitch (X), and yaw (Z) axes, seen in Fig. 5.2. while wearing HAT. From a resting position, looking straight ahead, his max rotations were ± 17 degrees in the roll axis, ± 59 degrees in the pitch axis, and ± 53 degrees in the yaw axis. While his pitch and yaw motion is unrestricted in relation to a non-disabled user, his roll motion is more constrained. The adaptability of the HAT interface allowed Henry to use only the pitch and yaw axes of his head motion for robot control.

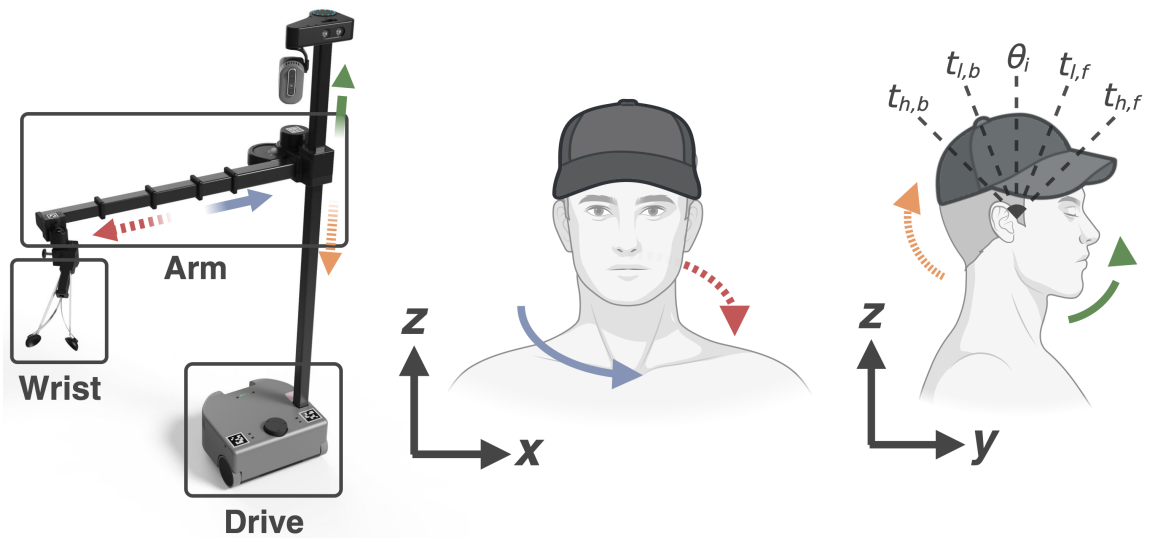


Figure 5.2: Three robot control modes: Drive, Arm, Wrist and mapping of head orientation to robot motion in arm mode.

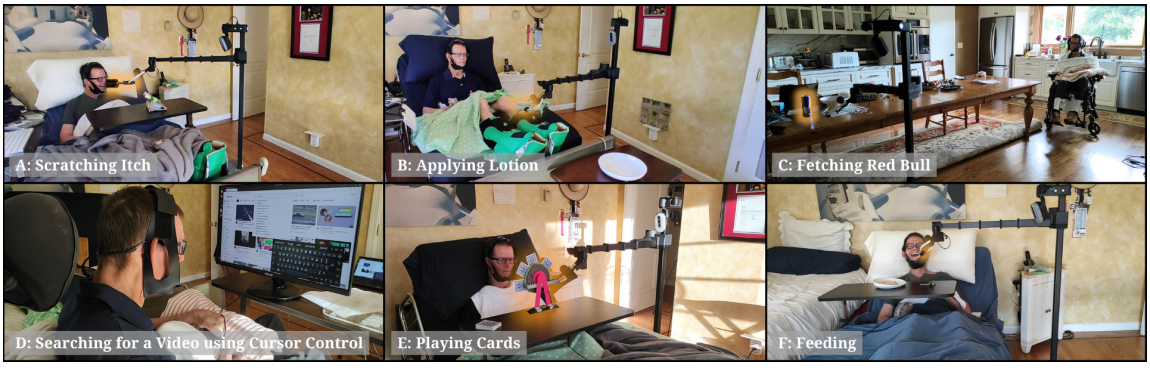


Figure 5.3: A subset of the tasks conducted during the study. Pertinent objects and tools are highlighted in orange.

The robot receives velocity commands over WiFi from the companion laptop at 20 Hz. In drive mode, the pitch angle is mapped to the robot base's forward and backward movement, while the yaw is mapped to clockwise and counterclockwise rotation. When the robot is in arm mode, as depicted in Fig. 5.2, pitch controls the height of the robot arm, and yaw manages its extension. Lastly, when in wrist mode, pitch controls the opening and closing the gripper, while yaw controls the wrist's rotational movement. Upon entering either robot control or cursor control mode, Henry must first initialize the orientation of HAT by clicking once. During

initialization, the orientation of the interface is saved and used for setting four thresholds: the two minimum motion thresholds, $t_{l,f}$ and $t_{l,b}$, and the two maximum motion thresholds, $t_{h,f}$ and $t_{h,b}$, as shown in Fig. 5.2. In conjunction with the user's head orientation, these thresholds are used for calculating the velocity sent to each robot actuator. After testing with Henry at the start of the study, we set the minimum thresholds to $\pm 10°$ from the initialized position, θ_i, and the maximum thresholds to $\pm 35°$ from the initialized position, θ_i. These thresholds are shown along the pitch axis in Fig. 5.2 and work as follows. If the user's head is tilted less than $t_{l,f}$ and greater than $t_{l,b}$, in both the X (pitch) and Z (yaw) axes, the robot stops all motion. While the angle measurement along an axis is within a pair of minimum and maximum motion thresholds (e.g. $t_{l,f}$ and $t_{h,f}$), the measurement is proportionally scaled to a velocity command for the robot's actuators. Further details and a mathematical formulation for velocity scaling are provided in the published work [168].

5.4.4 Cursor Control

To investigate the adaptability of the interface for use with different physical devices, we explored using HAT to control a computer cursor. This idea was proposed by Henry, who currently uses a computer vision-based head tracking interface that uses a webcam to track a reflective dot attached to his glasses. Prior to the start of the study, we implemented velocity control of the cursor, which uses the same equation as robot control. In this setup, tilting your head along the pitch axis moves the cursor up or down, and tilting it along the yaw axis shifts the cursor left or right. The degree of head tilt determines the cursor's speed. Based on Henry's suggestions, provided on Day 1, we developed and tested a position control version on Day 2 and Day 3, which acts more similarly to his vision-based head tracking interface. Using position control, the user's head orientation is directly mapped to a cursor position on the screen using linear interpolation, described further in the published work [168].

5.4.5 Graphical User Interface

Previous instantiations of head-worn interfaces have relied on line-of-sight [31, 73, 140, 167]. In contrast, when we pair HAT with a graphical user interface (GUI), users can control a mobile manipulator at further distances, performing tasks in other

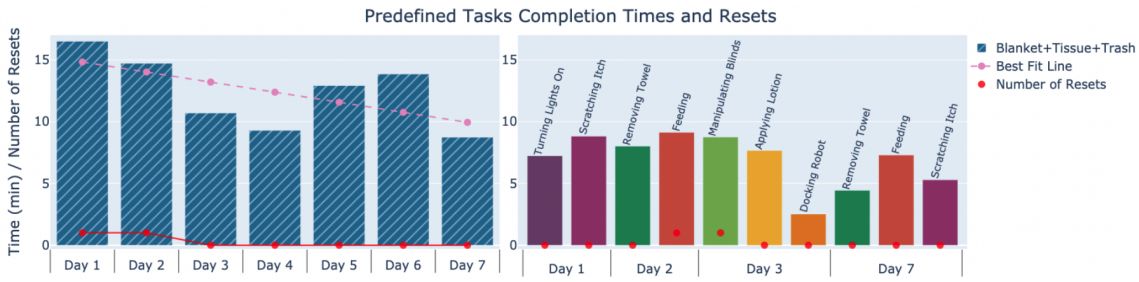


Figure 5.4: Left: Completion time and resets for the repeated blanket+Tissue+Trash task. A best fit line is plotted, showing a downward trend in completion times over the course of the study. Right: Completion time and resets for all other predefined tasks. Improvement is seen in task completion times and resets required for all repeated tasks over the course of the study.

rooms, such as retrieving a drink from the kitchen to the bedroom. The GUI provides information about the current mode, how to access other modes, and notably, it features two views from fisheye cameras, one mounted at the top of the robot for navigation and one mounted at the gripper for precision alignment and grasping.

5.4.6 Driver Assistance

To minimize cognitive load for grasping tasks, we propose a shared control method for HAT called Driver Assistance. When activated by the user, the system first processes input language queries that define target objects and matches these to detected objects within the environment. It then automatically aligns the robotic gripper's position with the intended object, while ensuring that the user maintains a sense of control.

Modeling

We model the Stretch robot, shown in Fig. 5.2, as a nonholonomic mobile manipulator. The velocity kinematics of the robot are $\mathcal{V} = \boldsymbol{J}(\boldsymbol{\theta})[v,\, \omega,\, \dot{\boldsymbol{\theta}}]^{\mathsf{T}}$, where $\mathcal{V} \in \mathbb{R}^6$ is the twist of the end-effector, v is the mobile base's forward velocity and ω rate of rotation, and $\boldsymbol{\theta} \in \mathbb{R}^3$ consists of the joint angles and joint positions of the arm. Lastly, $\boldsymbol{J}(\boldsymbol{\theta}) \in \mathbb{R}^{6\times 5}$ is the Jacobian matrix mapping between joint velocities and end-effector twist in robot base frame.

Open-Vocabulary Perception

To handle perception in complex household environments, we use the robot's Intel RealSense D435i RGB-D in conjunction with OWL-ViT [147], an open-vocabulary object detection model that accepts natural language (text-based) queries describing the object of interest for grasping. After obtaining a bounding box from OWL-ViT, the system segments the corresponding object from the point cloud, and then locates the pose of the object in the world frame for grasping. In this study, we input the language queries, for our participant based on the Driver Assistance tasks detailed in Section 5.5. We envision that users will select their own queries via a GUI in future iterations of our system.

Intent Recognition

The perception system provides a set of potential target objects denoted as goal set $\boldsymbol{g}$. From the goal set, we must infer the most probable object $g^* \in \boldsymbol{g}$ that the user wants to grasp based on the robot's state $\boldsymbol{\Theta}$. For each target object $g \in \boldsymbol{g}$, we design the probability of the object as the intended goal with $P(g \mid \boldsymbol{\Theta}) = 1/(1 + \rho d_{\text{ee}}(g))$, where $\rho = 5$ is an adjustable parameter and $d_{\text{ee}}(g)$ is the Euclidean distance between the end-effector and the object. Thus, the most probable goal is defined as

$$g^* = \underset{g \in \boldsymbol{g}}{\operatorname{argmax}}\, P(g \mid \boldsymbol{\Theta}) = \underset{g \in \boldsymbol{g}}{\operatorname{argmin}}\, d_{\text{ee}}(g), \tag{5.2}$$

which is the closest object to the end-effector. The confidence of the current inference among all goal candidates is computed as:

$$C(\boldsymbol{g}) = P(g^* \mid \boldsymbol{\Theta}) - \max_{g \in \boldsymbol{g} \setminus g^*} P(g \mid \boldsymbol{\Theta}), \tag{5.3}$$

which is the difference between the probability of the most likely target object and that of second most likely target object [97].

Driver Assistance Controller

Consider a kinematic controller where the control input, $\boldsymbol{u} \in \mathbb{R}^5$, is executed on joint velocities. To share control between the user and robot, we use a control

blending method $\boldsymbol{u} = \boldsymbol{u}_h + \alpha \boldsymbol{u}_a$ similar to past literature [54, 75], where $\boldsymbol{u}_h \in \mathbb{R}^5$ is the operator's input, $\boldsymbol{u}_a \in \mathbb{R}^5$ is the input from the robot's controller, and α is a weighting factor determining the extent of assistance the robot should provide. In our design, we set $\alpha = C(\boldsymbol{g})$, resulting in the robot providing little to no assistance when its confidence of the goal is low. The robot's controller is formulated as a Cartesian space proportional controller $\boldsymbol{u}_a = \boldsymbol{K}_p \boldsymbol{J}^\dagger(\boldsymbol{\theta})(\boldsymbol{x}_d - \boldsymbol{x})$, where $\boldsymbol{K}_p \in \mathbb{R}^{5\times5}$ are the control gains, $\boldsymbol{J}^\dagger(\boldsymbol{\theta})$ is the pseudo inverse of the Jacobian matrix, $\boldsymbol{x}_d$ is the desired end-effector configuration which is the pose of g^*, and $\boldsymbol{x}$ is the current end-effector configuration. To ensure a user retains the feeling of control of the robot, we constrain the robot's assistance on the joint that the user is controlling by multiplying a factor $\rho_c = 0.2$ on the corresponding row of $\boldsymbol{u}_r$. We additionally prevent the robot from controlling the arm extension, base translation, and wrist rotation by setting $\boldsymbol{K}_p = \text{diag}(0, 1, 1, 0, 0)$. By constraining these specific joints, the robot's motion only facilitates alignment with the object, allowing the user to remain in control of how the robot approaches and grasps the object with the arm.

5.5 Study Design

A full, comprehensive study schedule can be found in the published work [168]. The IVs are the task and the use of driver assistance and the DVs are task completion time and perception of the system. Henry completed a number of predefined tasks with line-of-sight of the robot: blanket+tissue+trash, scratching an itch, feeding, and applying lotion. Detailed task descriptions and completion criteria can be found in the published work [168]. He additionally completed several tasks without line-of-sight using the GUI. These tasks include turning the light on, manipulating blinds, removing soiled towel, and docking robot. The predefined tasks were primarily chosen based on tasks that Henry had previously completed in past assistive robotics studies [37, 40, 79, 86, 158]. Each day of the week-long study, he additionally performed a consistent challenging task, blanket+tissue+trash, shown in Fig. 5.1, to track performance with increasing experience over time. Except for this task, all predetermined tasks were executed once in the initial three days of the study. On the final day, we also repeated three of the predefined tasks (scratching itch, removing soiled towel, and feeding) to further assess improvement. Some tasks involved custom

tools, like a soft silicone spoon with an adapted handle for improved robot grip, as shown in Fig. 5.3f.

On the 5th day of the study, we had Henry do 3 within line-of-sight tasks to systematically evaluate our Driver Assistance method. The tasks were fetching Red Bull, cleaning up cups, and removing soiled towel. The tasks were completed with and without Driver Assistance, with the starting order alternated for each consecutive task. Two of the tasks were purposefully designed to be challenging, to explore the potential benefits of Driver Assistance. Specifically, for the fetching Red Bull task shown in Fig. 5.3c, we placed the drink can about 10 feet from Henry, making it challenging for depth perception. The cleaning up cups task also presented a challenge as it demanded manipulating two cups in sequence.

Throughout the study, we encouraged Henry to suggest any tasks he may want to attempt with the robot. On the 2nd day of the study, Henry used the robot for playing cards with his wife using a card wheel, shown in Fig. 5.3e, that can be rotated by the robot. On the 6th day of the study, we conducted an open day of testing, allowing Henry to dictate the tasks he wanted to try for an entire day. Henry did 4 tasks (opening drawer+fetching tape, grasping+fetching bottle, grasping towel+wiping table, and opening fridge), some with driver assistance and some out of line-of-sight.

After each task and at the end of each day, Henry responded to multiple 7-point Likert and NASA TLX workload items about robot control. He was additionally asked open-ended questions.

5.6 Results and Discussion

Full details on robot control tasks, task completion times, number of resets, and responses for Likert items and the NASA TLX scale can be found in the published work [168]. 6 tasks are shown visually in Fig. 5.3. Task videos are included on the project website.

5.6.1 Metrics

Task Completion and Efficiency

Henry attempted and completed a total of 17 predefined tasks, whose completion times are shown in Fig. 5.4. Times were modified to account for any technical issues. Through the course of the study, Henry became increasingly more efficient, showing a clear improvement in usage of the interface over time. This was observed in the recurring task, blanket+tissue+trash task, where there is a noticeable downward trend in completion times as seen on the left side of Fig. 5.4. Specifically, from day 1 to day 7, there was a 47% (7 m 46 s) decrease in time. With all 3 repeated predefined tasks, scratching itch, removing soiled towel, and feeding, we see clear, positive improvements from the start to the end of the study with a 40% (3 m 32 s), 45% (3 m 35 s), and 20% (1 m 50 s) reduction respectively in task times.

Over the week long study, Henry's perception of the interface's ability to enable task completion in a reasonable amount of time improved with him answering 4 (Neutral) and 2 (Disagree) on Days 1 and 2 to 6 (Agree) on Day 7. When asked about the efficiency of the interface in comparison to a direct teleoperation web interface without autonomy, Henry said "HAT is slightly more efficient [than a web interface] once you have mastered it. Certain specific tasks, however, are definitely more efficient with HAT. HAT in all cases is markedly better, and significantly more efficient and effective at 1) face to face things like playing simple games, or whenever a computer in your face is a problem like [during] blanket manipulation, 2) fine motor control of [the] gripper, and 3) [tasks that] use Assisted Grasping (Driver Assistance)–which is awesome and should be added to every Stretch interface." At numerous points of the study, Henry mentioned that the "analog" (continuous) control of the robot (i.e. degree of head tilt mapped to velocity of actuators) was useful, leading to higher efficiency and more precise control in comparison to direct teleoperation web interfaces.

Errors and Recovery

Our findings show that with minimal practice Henry was able to limit errors during tasks and consistently recover from any errors that occurred, two essential charac-

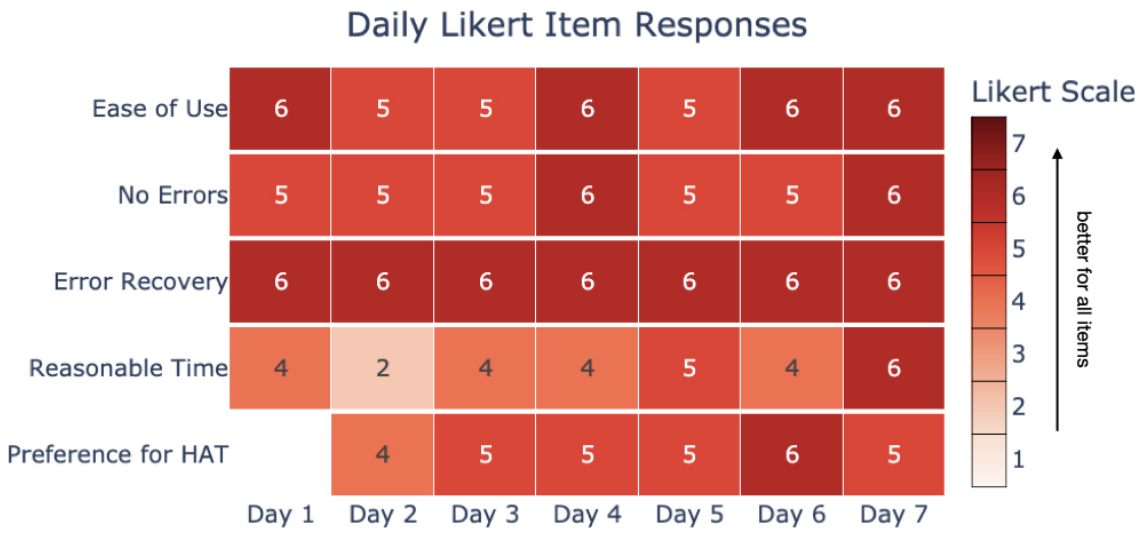


Figure 5.5: 7-point Likert item responses for Ease of Use, No Errors, Error Recovery, Reasonable Time, and Preference for each day of the study. For all items, a higher rating is better.

teristics for any proposed teleoperation interface for mobile manipulators in a home environment. As seen in Table 5.1, across all predefined tasks, Henry's median response was Agree (6) to the Likert item "I was able to complete the previous task without any errors using the control interface". Additionally, as seen in the daily responses to the same Likert item in Fig. 5.5, Henry replied Somewhat Agree (5) or Agree (6) on all days. Henry completed 13 out of 17 predefined tasks without any resets (critical errors that required resetting of the environment or robot by a researcher). The number of resets per predefined task is shown in Fig. 5.4. For the blanket+tissue+trash task, Henry required a reset on the first two days of attempting the task. After Day 3, no resets were needed as Henry succeeded in instinctively putting the robot back into idle mode and re-initializing the HAT when he faced any inadvertent robot motion. As seen in the median responses across all predefined tasks (Table 5.1) and daily responses (Fig. 5.5), he responded Agree (6) to the Likert item "If there were errors, the control interface allowed easy recovery".

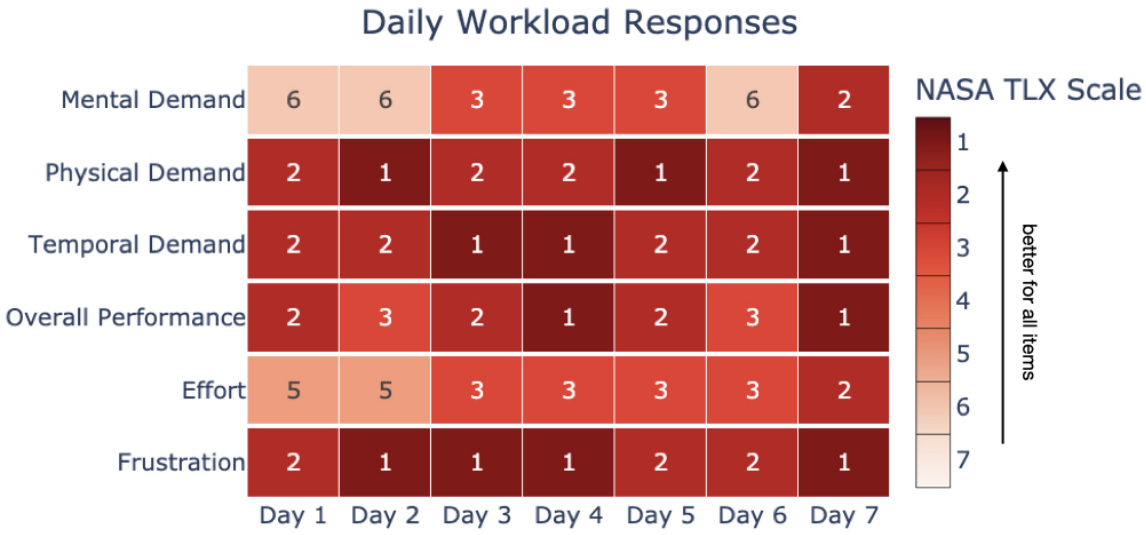


Figure 5.6: 7-point NASA TLX responses for each day of the study. For all items, a lower rating is better.

Table 5.1: Predefined Tasks Likert Responses

Likert Item	median	interquartile range	min	max
Ease of Use	6	0	3	7
No Errors	6	1	3	6
Error Recovery	6	0	5	7
Reasonable Time	4	1	2	5
Preference for HAT	5	0	4	6

Ease of Use

As seen in Table 5.1, across all predefined tasks, Henry's median response was Agree (6) to the Likert item "The control interface was easy to use." Specifically, Henry gave a 7 (Strongly Agree) ranking to 3 tasks: Playing Cards, Blanket+Tissue+Trash, and Manipulating Blinds. This was similarly the case for his daily responses; as seen in Fig. 5.5, Henry replied either Somewhat Agree (5) or Agree (6). As discussed in the Task Completion and Efficiency section, Henry found the interface well suited for any precise tasks due to having continuous control and for any tasks around his body due to not having a monitor directly in front of him. At the end of the study, Henry

specified there are certain tasks that he would use HAT and the robot to do instead of relying on a caregiver. He said, "Definitely scratching itches. I would be happy to have it stand next to me all day, ready to do that or hold a towel to my mouth. Also, feeding me soft foods, operating the blinds and doing odd jobs ('freestyling') around the room would be helpful." Starting on the second day of the study, we additionally asked the Likert item, "I prefer to use HAT over a web interface overall" assuming direct teleoperation and no robot autonomy for both interfaces. As seen in Fig. 5.5 and Table 5.1, Henry either answered Neutral (4) or higher for all tasks, with a median response of Somewhat Agree (5). When asked if the HAT interface is intuitive to use, Henry said, "It takes a bit of practice, but all controls become reflexive after a few days."

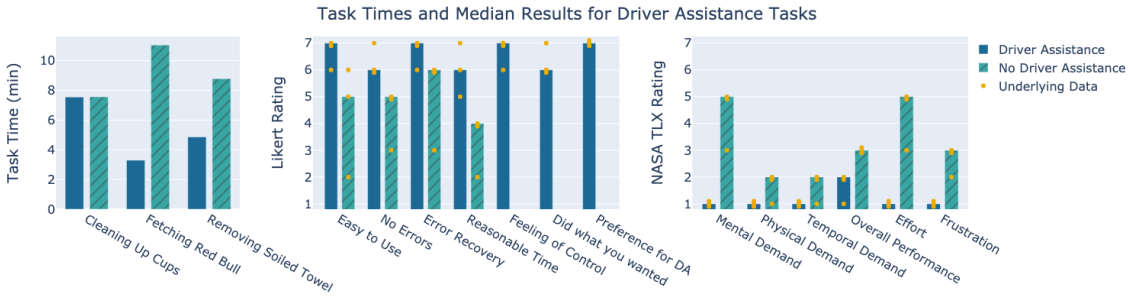


Figure 5.7: Left: Comparison of task times for the 3 tasks completed with and without Driver Assistance. Middle: Median Likert items across all 3 tasks with and without Driver Assistance. For all Likert items, a higher rating is better. Right: Median NASA TLX responses across all 3 tasks with and without Driver Assistance. For all NASA TLX items, a lower rating is better.

Learning Curve

By comparing task times and errors for the daily blanket+tissue+trash task, we can more thoroughly explore the interface learning curve. After two days, Henry was able to complete the task without any resets required, showing familiarity with the interface. From the second to third day of the study, his perception of errors also improved with a jump in response from 3 (Somewhat Disagree) to 6 (Agree) for the Likert item on no errors. Despite some outliers, Henry continued to improve his task

time over the course of the study as shown by the best fit line in Fig. 5.4. Henry commented on the learning curve saying, "It actually had a fairly steep learning curve, although not wide. An easy way to reduce the learning needed would be to produce a 'Quick Start' card providing, on one side, all the information you need to get started, including what each head movement does. After a few days, the card will be unnecessary. It does get easier every day for a while."

Workload

Daily workload measures using a 7 point NASA TLX Scale are shown in Fig. 5.6. On the first two days, Henry reported higher values of mental demand (6) and effort (5). He specified this was due to the clicking patterns (which we changed after day 2) and due to lack of familiarity with the interface. On day 6, the open day, there was another spike in mental demand due to some of the challenging tasks (opening drawer+tape fetching and opening fridge) that Henry did with the interface and robot. Throughout the study, Henry had strong ratings for HAT in physical and temporal demand, overall performance, and frustration. Henry gave his strongest workload ratings for HAT on the last day with a 1 (Low) for physical demand, temporal demand, and frustration, a 1 (Successful) for overall performance, and a 2 for mental demand and effort. While evaluation with more individuals with impairments is needed, these strong workload measures highlight the potential for HAT for individuals with motor impairments who retain some degree of head and neck mobility.

5.6.2 Effect of Driver Assistance

We find that Driver Assistance led to clear improvements in task times, Likert responses, and workload measures while still preserving user control of the robot in the home. As seen in Fig. 5.7, for cleaning up cups, fetching Red Bull, and removing soiled towel, Driver Assistance led to a reduction in task times by 1 s, 7 m 45 s (70%), and 3 m 55 s (46%) respectively. Despite similar task times for the cleaning up cups task, while not using Driver Assistance, Henry required a reset of the environment for knocking over one of the cups while trying to align the robot. For the fetching Red Bull task, the automatic alignment was especially essential as the can was far away from Henry, making it extremely challenging to perceive depth to align the gripper

with the can.

After testing Driver Assistance for the first time, Henry expressed his fondness for the system saying, "that's better than anything I have tried for grasping." As seen in Fig. 5.7, Henry reported improvements in ease of use, errors, error recovery, and completion time for all tasks. Similarly, in Fig. 5.7, Henry reported a substantial reduction of 4 points in both mental demand and effort and improvements in physical/temporal demand, overall performance, and frustration. In contrast to prior work in shared control where users have reported reduced feeling of control with increasing robot autonomy [27, 98], Henry did not feel this for HAT. For the Driver Assistance tasks, as seen in Fig. 5.7, Henry agreed (6) or strongly agreed (7) with the Likert items related to feeling in control, having the robot do what he wanted, and preference for the assistance over full teleoperation.

Further testing will be necessary to assess the advantages of driver assistance in improving robotic manipulation within a natural home environment for a broader population of individuals facing significant motor impairments. Nevertheless, these preliminary findings show that incorporating open-vocabulary object detection perception models alongside shared control for alignment during grasping tasks may play a vital role in alleviating the workload involved in robotic teleoperation within household settings.

5.6.3 Out of Line-of-Sight Tasks

Henry completed 5 predefined tasks (turning the lights on, manipulating blinds, docking robot, and removing soiled towel twice) out of line-of-sight using HAT coupled with a GUI displaying the robot's camera feeds. On the open day, he completed two more tasks, opening drawer+fetching tape and opening fridge, using HAT and the GUI. Across all 7 tasks, when asked the Likert Item, "The control interface was harder to use without line-of-sight than with line-of-sight," Henry responded with a median response of 1 (Strongly Disagree) with an interquartile range (IQR) of 1, showing that the experience of using HAT with the GUI was similar to using it without. On Day 3, after completing all the predefined tasks at least once, Henry said controlling the robot using the GUI is "much easier [than without] since [the] yaw [axis] control makes it difficult to look around." Although, at times, Henry found the GUI camera

views useful, he mentioned on numerous occasions that he preferred a screen-free experience when performing tasks involving his face and body.

5.6.4 Control of Multiple Devices

On multiple days of the study, Henry evaluated HAT as an interface to control the cursor on his computer. Two tasks that Henry completed were typing sentences in a notepad and searching for a video on YouTube as seen in Fig. 5.3d. On Day 3, for a quickly implemented position-based cursor control, the system worked well with Henry saying it was "pretty close" in comparison to his current head tracking interface though "about 60% as fast as [the] head tracker because [the] cursor jumped around". One notable advantage, as highlighted by Henry, was that HAT required less frequent recalibration compared to his existing head tracking system. Our changes after Day 3 concentrated on mitigating cursor jumps caused by noise in the IMU-estimated angles. Testing of cursor control with the HAT showed that the interface is viable for control of other physical devices. Much like a web interface, one can envision HAT seamlessly enabling concurrent tasks such as controlling a robot, composing emails on a computer, and changing channels on a TV, all without the need for caregiver assistance.

5.6.5 Lessons Learned

Viability of HAT

In conjunction with our previous study on this interface which involved 18 participants of primarily low expertise with controlling robots, the results of this case study show that HAT is a strong option as a direct teleoperation interface for mobile manipulators as it performs strongly in terms of task efficiency, errors, ease of use, learning curve, and workload. These results are further supported by our participant's preference for HAT in comparison to similar direct teleoperation web interfaces he has used in the past. We expect these results to generalize to a broad range of individuals with motor impairments, though head and neck motor function along at least 2 axes must be retained for the interface to be effective. Since extensive testing was only conducted with a non-speaking individual with quadriplegia due to a brainstem stroke, future

testing with additional individuals with motor impairments is crucial to fully validate these findings.

Customizability

This work highlights the importance of personalizing assistive interfaces to individual needs. In the case of Henry, utilizing prior versions of HAT would have been infeasible without switching from roll to yaw axis control and from speech to clicker for mode switching. Furthermore, without the ability to adjust mode switching clicking patterns to suit his specific needs and condition, Henry would have also had significantly higher mental demand while using the interface. Lastly, the ability to modify the physical design of HAT holds significance, allowing users to incorporate their comfort and fashion preferences into the interface.

Combining Interfaces

Wearable interfaces, such as HAT, offer distinct advantages when compared to standard interfaces for robotic teleoperation. First, wearable interfaces embedded in clothes are readily accessible when needed and inconspicuous when not in use. Such interfaces can offer a more direct and intuitive means of controlling mobile manipulators. Additionally, the absence of a screen or input device (head tracking, sip and puff, etc.) directly in front of the user's face allows for increased flexibility when performing tasks around the body and enhances situational awareness. On the other hand, web-based interfaces come with their own set of benefits, including a greater number of inputs through on-screen buttons and the capability to display camera feeds. They also simplify the activation of autonomous routines, allowing for completion of repetitive tasks such as navigating between rooms or turning on/off the lights with a single click. At the conclusion of the study, Henry specified that the largest limitation of HAT was the lack of autonomous routines other than Driver Assistance. His feedback was that we could "use a web interface to 'program' the automated function and either HAT or voice controls to execute it".

Combining wearable technologies like HAT with a screen-based interface for triggering semi-autonomous or autonomous robot actions could present a robust interface for many individuals with motor impairments. It's conceivable that Henry

might employ HAT to fully teleoperate the robot to assist him at times of rest when he doesn't have his computer screen in front of him. Many of his needs at this time such as wiping his face, itch scratching, or moving his blanket may be benefited by limited autonomy, as this would allow him to exercise preference. In contrast, at times when his caregiver has set up his computer in front of him, he could use HAT with the screen-based GUI. With the presented cursor control functionality, he could use the GUI to trigger a variety of autonomous actions that pair with direct teleoperation using HAT. To reduce a view-obstructing screen, future research could explore the use of augmented reality or the display of information via a projected screen at a distance from the user's body.

Adoption in the Home

Intuitive teleoperation of assistive mobile manipulators in the home is approaching a technologically feasible state with a wearable interface like HAT. For wider adoption, a few more development steps are needed for safety and robustness.

Henry occasionally encountered IMU drift, likely due to magnetic interference, sometimes leading to unintended robot movements as a result of shifting thresholds. We found that this drift was notably pronounced when Henry was seated in his wheelchair and his legs were involuntarily shaking, causing his head to also shake. Under the existing system setup, Henry had no means of realizing drift was occurring except when it resulted in unintentional robot motions. Many IMUs, including the one utilized, provide calibration status indicators which could be used to require the user to reinitialize HAT when the confidence in the fusion estimate is low.

During the human study, Henry also faced inadvertent robot motions due to spasming and looking around the room without putting the robot into idle mode. Another potential safety feature is an 'anomaly detection' classifier [154, 174] to distinguish the aforementioned head motions from intended head tilting movements for robot control. Given that head tilting motions with HAT tend to be notably slower than other head movements, this classifier could predict intent using head accelerations captured by the IMU and could automatically switch the robot to the idle mode upon detection of a non-head tilting motion.

Lastly, a safety module which forecasts potential collisions based on depth images

and robot velocities, could enhance suitability for home use. This functionality would enable the robot to automatically decrease actuator velocities while in close proximity to the body, increasing safety and reducing mental workload.

Head-worn interfaces offer a promising solution for individuals with severe motor impairments who retain head movement, but their utility is restricted for those without this capability. In contrast, speech is a powerful alternative for individuals who can speak, allowing them to give both broad commands and more specific preferences to robots. With advancements in large language models (LLMs), speech interfaces could become even more effective for controlling robots. The next chapter delves into the use of LLMs as speech interfaces for physically assistive robots.

Chapter 6

VoicePilot: Harnessing LLMs as Speech Interfaces for Physically Assistive Robots

6.1 Contributions

I formulated the core ideas for this project with Zackory Erickson and Henny Admoni. Jessie Yuan designed and implemented the LLM based speech interface with guidance from me. I wrote the IRB protocol. I designed and executed the user study with Jessie Yuan and input from Zackory Erickson and Henny Admoni. I wrote the paper with input from all authors, especially Jessie Yuan and Janavi Gupta. Janavi Gupta conducted the thematic analysis with the guidance of Zulekha Karachiwalla and Henny Admoni. Jessie Yuan and I formalized the framework, while Janavi Gupta, Zulekha Karachiwalla, and I formalized the design guidelines. Advising on all aspects of the project was provided by Zackory Erickson and Carmel Majidi.

6.2 Introduction

Interfaces for mobile manipulators often have physical requirements, dictating which individuals can effectively use them. For example, individuals without the ability to

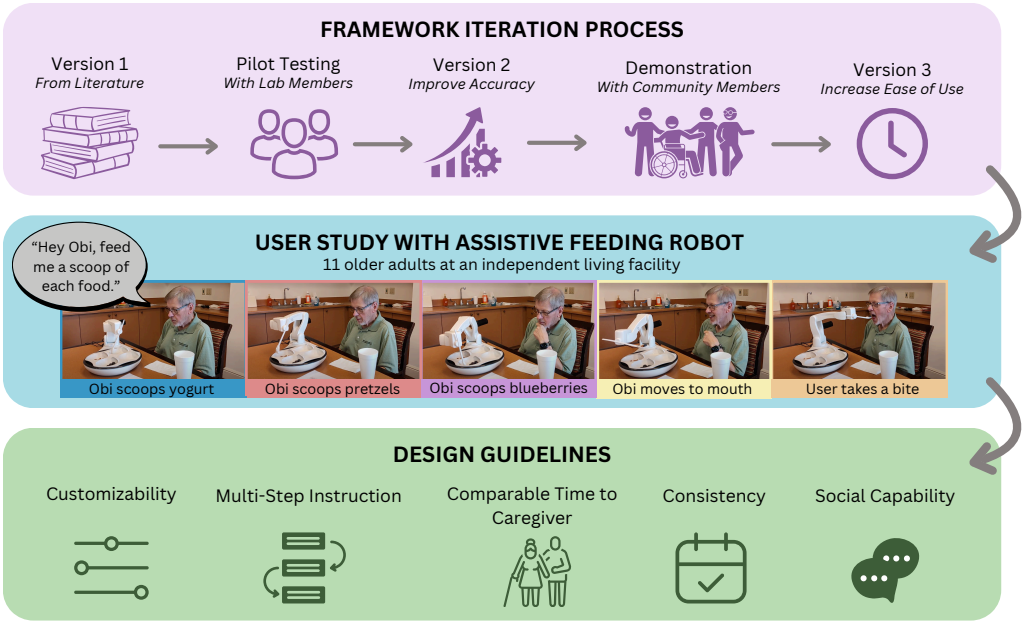


Figure 6.1: Top: Our work introduces a framework that was iteratively developed for integrating LLMs as speech interfaces for assistive robots. Middle: We develop a speech interface for a feeding robot using the framework and evaluate it with 11 older adults at an independent living facility. Bottom: Using insights from the study, we present a reflection and 5 design guidelines.

move their head may use eye tracking with a web-based interface [33, 37] while those with head motion may prefer inertial interfaces that capture their head movements and directly map them to robot motion [167, 168]. Among these interfaces, speech excels as an option for those with the ability to speak, as it allows individuals to naturally provide robots with both high-level commands and nuanced preferences or customizations. Speech interfaces for assistive robots have been explored extensively for control of both mobile robots and robot arms [92, 122, 185] and can be more intuitive than existing interface options like joysticks [182]. With the advent of Large Language Models (LLMs) and ongoing research combining LLMs with robots, robust speech interfaces for robots are becoming increasingly viable in the near future. Existing works have presented frameworks and guidelines for integrating LLMs with robots for high-level task planning and code generation [11, 30, 128, 137, 207, 223, 229, 245, 248]. These works concentrate on prompt engineering and code generation to enhance the accuracy of large language models (LLMs). No prior works involve human subjects interacting directly with an LLM-integrated robot for physically

assistive tasks. Consequently, they lack insight and guidelines on the human-centric considerations essential for using LLMs as assistive interfaces.

To address this gap, we develop a framework and design guidelines for integrating LLMs as speech interfaces for physically assistive robots. Our presented framework underwent extensive iterative development and consists of components related to prompt engineering and system rollout, the process of deploying the LLM-based speech interface and robot with users. First, we leveraged insight from existing literature to develop an initial version. We applied this Version 1 framework to integrate an existing LLM, GPT-3.5 Turbo from OpenAI [5], with the commercial feeding robot Obi [1]. After testing this interface with lab members, we refined the framework, iterated on our LLM-based speech interface for the Obi, and demonstrated it to community members at a disability awareness event (Version 2). We used insights from the demonstration to develop a final framework and interface, evaluated through a user study involving 11 older adults at an independent living facility (Version 3). Lastly, we use both quantitative and qualitative data from the study with older adults to present design guidelines.

Our final framework, shown in Fig. 6.3, and design guidelines, shown in Fig. 6.8, provide insight on important human-centric considerations to researchers, engineers, and product designers for integrating off-the-shelf LLMs with assistive robots and devices. The contributions of this work are as follows:

- We present an iteratively constructed framework, shown in Fig. 6.3, for integrating LLMs as assistive speech interfaces for a robot.
- We integrate an existing LLM with a commercial feeding robot using our framework and evaluate the interface through a human study with 11 older adults from an independent living facility.
- We present a reflection and design guidelines, shown in Fig. 6.8, based on the findings of the human study.

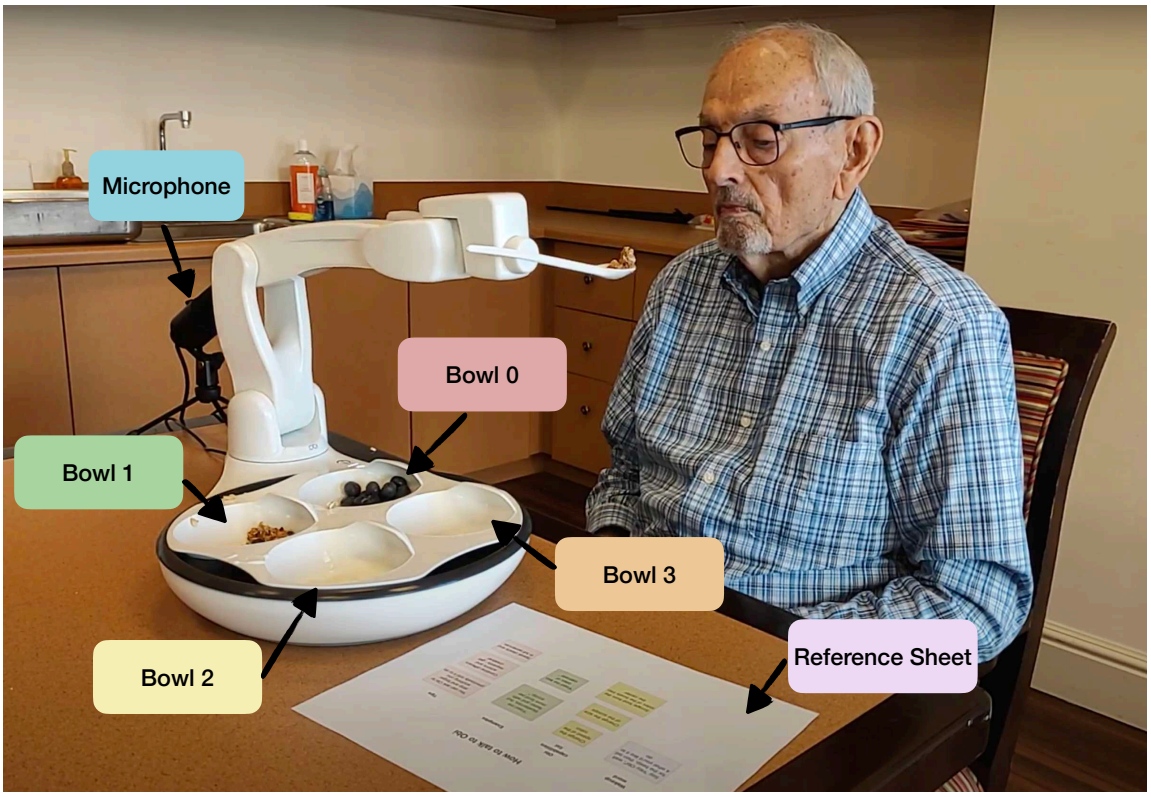


Figure 6.2: Obi robot and setup for study with 11 older adults at an independent living facility. The Obi robot's arm moves towards a participant's mouth with a spoonful of granola. A microphone is positioned to the right of the participant and a cheat sheet with example commands is placed to the left of the participant.

6.3 Related Work

6.3.1 LLMs for Robotics

There have been numerous advancements in using LLMs for robotics [11, 30, 68, 108, 128, 137, 207, 223, 224, 229, 245, 248]. For the scope of this paper, we will focus on works that use off-the-shelf LLMs for high-level task planning through code generation, as these works are most pertinent to using LLMs as assistive interfaces. Liang et al. utilize codewriting LLMs with perception and control APIs to create reactive and waypoint-based trajectories for various robotic platforms, while Vemprala et al. develop design guidelines and a pipeline for integrating ChatGPT with different robots and simulation environments [128, 223]. Similarly, Singh et al. and Arenas

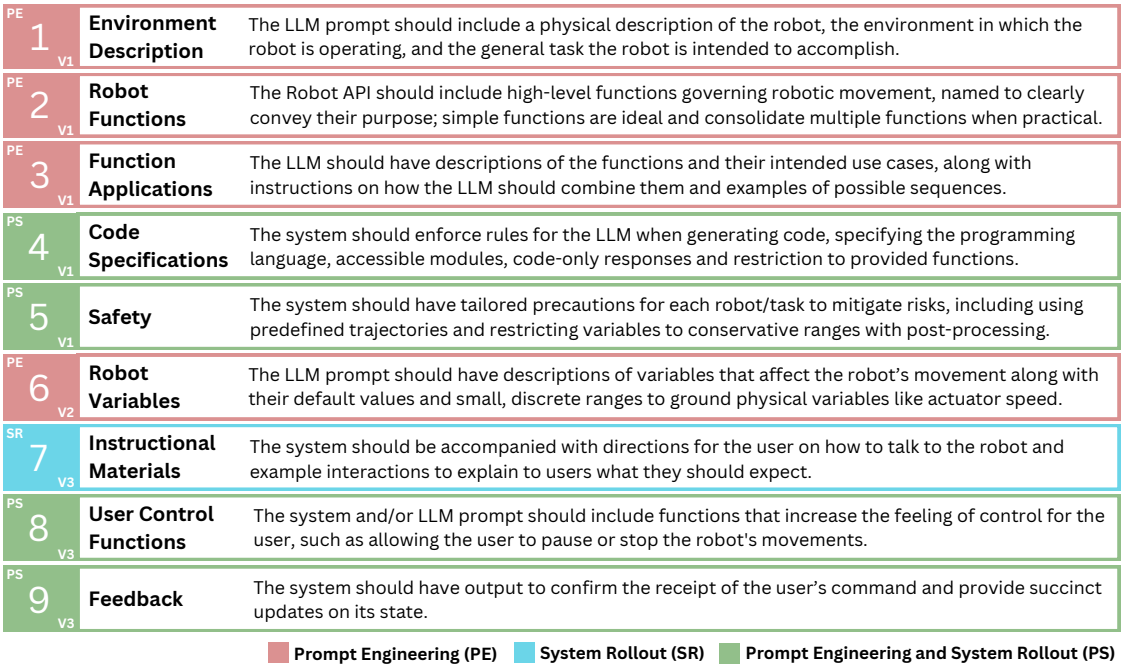


Figure 6.3: Our final framework consisting of 9 components is shown. The color and annotation at the top left of each numbered component indicates whether the component is related to prompt engineering (PE), system rollout (SR), or both prompt engineering and system rollout (PS). System rollout refers to the process of deploying the LLM-based speech interface and robot with users. The annotation at the bottom right of each numbered component shows which framework iteration it was added in.

et al. provide prompting guidelines and methodologies for using LLMs in robotic tasks [16, 207]. Other work has used ChatGPT as an interface for surgical robots and quadrupeds [134, 173]. These prior works primarily focus on high level planning and code generation which is just one aspect of designing a speech interface. Most utilize prompt engineering and evaluate their LLMs using accuracy, specifically if the LLM succeeds in generating the "correct" code that the researcher expects based on the provided prompt and command. While some papers include human studies to evaluate their methods, none evaluate their system with participants who interacted directly with the LLM-integrated robot. In comparison, our work includes multiple rounds of direct interaction between the robot and participants including with older adults who are close to the target population of individuals who need feeding assistance.

This process allows us to identify human-centric elements essential to positive user perception of the system.

Past works have also explored using LLMs with social and service robots [93, 111, 112, 232]. Wu et al. [232] focus on learning personal preferences through language input for TidyBot, a cleaning robot, while Hu et al. [93] benchmark LLM performance for code generation for various service tasks. Both papers focus on how LLMs can be utilized to perform a specific task and similar to other papers, do not have participants interacting directly with the robot. The absence of these human studies limits the insights into improving the general interface design and human-robot interaction. Lastly, Kim et al. explore using LLMs with a social robot and focus on conversational tasks, such as having the robot guide the participant in making a drink [112]. The authors compare their robot against text and voice based agents and present design insights from a thematic analysis of the data obtained during the human study. Most of their findings are only applicable to social, conversational robots and do not generalize to physically assistive robots and tasks such as robot-assisted feeding. In comparison to all past work on using LLMs for robots, we iterate on our presented framework through multiple rounds of testing with users and we introduce design guidelines derived from our final human study.

6.3.2 Robot-Assisted Feeding

Robot-assisted feeding systems employ robotic manipulators and perception systems to feed users either autonomously or semi-autonomously using input from users through assistive interfaces [23, 25, 77, 100, 175]. Voice control interfaces have been explored in past works on robot-assisted feeding, but use keyword detection in comparison to LLMs, which could enable more natural speech interactions with the robot [28, 151]. Bhattacharjee et al. find that participants prefer using speech over a web interface while dining alone [28] while Nanavati et al. also find that participants see value in speech interfaces in quieter social settings [151]. In both studies, the researchers find that participants may not want to use speech in social dining settings as it may interfere in conversations. Lastly, Nanavati et al. find that participants want customizable feeding robots that can be tailored to specific needs and take into account their preferences [151]. These works motivate the need for the development

of customizable speech interfaces for assistive feeding robots.

Alongside research progress in assistive feeding, commercial feeding robots, such as the Obi feeding robot, are being used by hundreds of individuals across the United States. Obi, shown in Fig. 6.2, is a FDA compliant Class 1 medical device from DESĪN LLC [1]. The robot consists of a 6 degree-of-freedom robot arm with a spoon end-effector and a dish tray consisting of 4 bowls. The commercial system uses hard-coded trajectories for the utensil to scoop food from a bowl and transfer that food to a predefined location near the user's mouth. For this work, we utilize a research version of the Obi robot with a Python API that allows us to send trajectories of joint angles to the robot. Obi is currently only commercially sold with a two button interface, with one button to select the bowl and the other button to scoop from the selected bowl. The button interface can be limiting especially for users who may want to customize their assistance; in comparison, an LLM-based speech interface can allow a user to provide verbal instructions to Obi, much like they would to a human caregiver, allowing them to customize how the robot feeds them.

6.4 Version 1: Framework from Literature

6.4.1 Framework Development

Prior work has shown that off-the-shelf LLMs can be used for high level planning and code generation for robots when provided with a well-designed prompt [16, 207, 223, 224]. We start our design process by drawing on this existing work on prompt engineering for LLMs for robotic control to develop an initial framework.

Our Version 1 framework identifies five preliminary components, shown and described in Fig. 6.3. We describe how these preliminary components were motivated from past literature:

- Environment Description: Vemprala et al. suggest to include information in the prompt about the environment in which the task is taking place and the goals and objectives of the task [223]; likewise, Arenas et al. recommend including an overview of the robot and the task at hand [16]. We combine insights from these past works to define Environment Description as a physical description of the robot, the environment in which the robot is operating, and the general

task the robot is intended to accomplish.

- Robot Functions: Past papers on using LLMs for high-level task planning and code generation suggest defining and listing a set of robot functions, named the Robot API, in the prompt so that the LLM can access these function names and use them to control the robot [16, 207, 223]. Similar robot functions are required for LLM-based assistive speech interfaces. For example, for an assistive cleaning robot, you may include a function to perceive objects and a function to grasp an object which can be used in conjunction by an LLM to generate code to detect and pick up toys from the floor of a cluttered room.
- Function Applications: Beyond listing the names of the robot functions, Vemprala et al. and Arenas et al. recommend including documentation of what each function does in the prompt [16, 223]. Both works also suggest providing solution examples to demonstrate to GPT how to use the functions to accurately respond to users commands; Arenas et al. especially emphasize the importance of examples because the custom control functions are not in an LLM's training set, so the model is unfamiliar with them [16]. Therefore, we dedicate a component of our framework to specifying what each function does and providing examples of how they are intended to be used.
- Code Specifications: In order to use an off-the-shelf LLM for code generation for robots, past work has shown that the LLM must be prompted with general details of the programming environment, including what programming language should be used and what packages are accessible [16, 207, 223].
- Safety: While integrating LLMs with assistive robots, safety needs to be evaluated on a case by case basis depending on the robot and the task at hand. Vemprala et al. suggest having a human in the loop to inspect the code that GPT generates, which isn't practical for an assistive interface that should respond immediately to the user [223]. Other options include control-barrier functions [43, 190] and testing extensively in simulation before deployment [80, 224]. Due to the especially crucial role of safety in the context of physically assistive robots, we suggest using lightweight or compliant robots and predefined trajectories, thus barring GPT from modifying low-level joint angles or robot positions.

6.4.2 Implementation

To validate our framework, we implemented an LLM-based speech interface for the Obi feeding robot. We chose to use the Obi as it is a commercially available assistive robot designed with safety considerations, namely a lightweight arm and a detachable spoon in case of collisions. Additionally, as discussed in Section 6.3.2, prior studies on robot-assisted feeding indicate a preference for voice control interfaces in specific scenarios like individual dining, underscoring the necessity for such interface development.

Using our initial framework, we crafted a tailored prompt for GPT-3.5 Turbo, subsequently referred to as GPT, designed for the Obi feeding robot. Our prompt, included in the published work [169], integrated the five identified components from our Version 1 framework.

We addressed the components as follows: to implement the Environment Description, within our prompt, we included details on the task at hand, feeding the user, and the working environment, such as the foods in each of the bowls, as the Obi robot does not have a built-in perception system. We incorporated the Robot Functions and Function Applications components by providing the names of three high-level robotic control functions in the prompt, along with a short description of what each function did and how it should be used. These three functions moved the robot arm along hard-coded trajectories and were described as follows: `obi.scoop_from_bowlno(bowlno)` moved the robotic arm to the specified bowl and scooped food from it; `obi.move_to_mouth()` moved the robot to the user's mouth position; and `obi.scrape_down_bowlno(bowlno)` moved the robot to the specified bowl and subsequently scraped food from the sides to the center of that bowl so that the robot could pick up additional food. We designed the Robot API to include these three functions because they correspond to actions performed by the commercially available Obi robot. Specifically, the robot's default two-button interface allows users to select the bowl with one button and scoop/feed themselves with the other. Additionally, the scraping motion occurs automatically after every few scoops.

At the end of the prompt, to satisfy the Code Specifications component, we included sentences instructing GPT to provide Python code for each request. Lastly, to address the Safety component, each function only moved the robot along a

predefined trajectory, as opposed to allowing GPT to directly dictate the joint angles of Obi's arm. This prevents GPT from generating unsafe or unanticipated trajectories. Choosing to use the Obi robot, which is specifically designed to be safe for individuals with motor impairments, instead of other robotic feeding systems, is another implementation decision for enhancing safety.

At this stage, users would hold down a physical button while speaking to record a command to the robot, which would be transcribed to text using OpenAI's Whisper API [6]. The parsed speech-to-text commands were appended to the bottom of the prompt and provided to GPT, which generated Python code that could subsequently be deployed on the Obi robot.

6.4.3 Pilot Testing with Lab Members

We conducted pilot testing of our Version 1 interface for the Obi robot within our research team and members of our lab group. Four lab members watched videos of the robot performing four feeding tasks and suggested different command phrasings to accomplish the task, resulting in 48 phrases. These commands were tested for accuracy, with mistakes analyzed and prompts adjusted iteratively. Later, three researchers directly interacted with the robot to further assess and improve the system, focusing on user comfort during assisted feeding. More details on this process is included in the published work [169].

6.5 Version 2: Improved Accuracy and Customizability

6.5.1 Framework Iteration

Based on our pilot testing, we noted instances where GPT made errors, enabling us to identify gaps or areas within the framework that required iteration. Specifically, the Code Specifications were not detailed enough, causing GPT to import nonexistent modules and define unnecessary new functions. To address this, the framework was modified to emphasize the importance of explicitly outlining how GPT should write code. Additionally, the Function Applications component was expanded to

better describe the relationships between functions and provide examples, while the Robot Functions component was modified to suggest merging common sequential functions to improve performance in tasks that required multiple robot actions. More information on these changes can be found in the published work [169].

Finally, we added a new field to the existing framework:

- Robot Variables: Users wanted to have more ability to customize the way in which the robot moved. In the case of the Obi robot, the first iteration had only a single speed and scoop depth, yet, based on user feedback, we found that preferences for these can vary significantly from user to user, between different foods, and even from situation to situation. Accordingly, we added a component to the framework providing descriptions of variables affecting the way in which the robot moves, their ranges, and their default values.

Due to the user safety implications of the new Robot Variables component which gives the LLM speech interface control of motion variables such as speed, we proactively modified the Safety component of the framework to suggest restriction of robot variables such as speed to conservative ranges and to check the validity of robot variables in LLM-outputted code to ensure they are within expected bounds.

6.5.2 Implementation

At this stage, we changed the prompt for the Obi system to reflect the modifications to the framework and thus improve the accuracy of GPT's responses.

In summary, we updated the prompt to emphasize the production of responses exclusively in Python code, without importing extra modules or creating new functions, as specified in the revised Code Specifications component. To adhere to the updated Function Applications component, we added detailed instructions on how the robot functions should be combined, emphasizing Obi's movement to the user's mouth when asked to be fed, and included a 4-second default pause between bites for better user experience, with customization options for the delay. Additionally, the Robot Functions component was modified to ensure that after scraping down a bowl, Obi would immediately scoop from the same bowl to improve accuracy, addressing previous issues where the sequence was misunderstood. For the Robot Variables component, we added new variables for Obi's movements (speed, acceleration, and scoop size) to



Figure 6.4: 9 of the 11 participants are shown at various stages of the Version 3 study at an independent living facility with older adults.

enhance customizability, with post-processing by the system to ensure safety. More information on these changes can be found in the published work [169].

6.5.3 Demonstration with Community Members

After the modifications to the implementation in Version 2, we brought the system to a disability awareness event in Pittsburgh, PA, USA, where we demonstrated the speech interface to about 15 members of the community, some of whom had motor impairments or other types of disabilities. During the event, community members who wanted to interact with the robot provided commands to the feeding robot to have it feed a lab member either M&Ms or Cheerios cereal placed in two bowls of the Obi. We encouraged visitors to be creative with their requests to the robot. Most individuals who interacted with the robot provided only 1-2 commands and the interactions were short, lasting only a few minutes. Nevertheless, this process was

helpful for evaluating the system, enabling us to identify where users were confused or struggled to use the interface, as well as hear feedback on what people liked and disliked.

6.6 Version 3: Increased Ease of Use

6.6.1 Framework Iteration

Using insights from the Version 2 demonstration, we identified essential modifications and new components for our framework. First, we made adjustments to the Robot Variables component due to GPT's inconsistent handling of continuous variables, which caused unpredictable speed changes that confused users. By grounding variables like actuator speed on a 0-5 scale, speed adjustments became more consistent and aligned with user expectations. This process allowed us to identify that while increasing the granularity of robot function and variables can enhance the user perception of robot customization, they can negatively impact the language model's accuracy. More details are included in the published work [169].

Additionally, we added three new components to the framework, all of which focus on human-robot interaction:

- Instructional Materials: Users tended to struggle with giving commands to the system at first, as they were unsure what they were allowed to say or what the robot was capable of and were wary of experimenting with new commands. Thus, it became evident that the framework required a field on how to effectively prepare users on how to use a speech interface by providing guidelines for user communication with the robot and sample interactions to set their expectations. This component was added to emphasize the importance of giving users more guidance and support with using the system beforehand to reduce user confusion.
- User Control Functions: A component was added to incorporate functions that increase the feeling of control for the user, an important consideration for assistive interfaces [28, 99, 168]. Example functions include allowing the user to pause or stop the robot's movements using voice commands.
- Feedback: Users struggled to discern whether the system heard them or not,

especially given the few second delay between completion of the command and robot movement due to networking and GPT processing the request. As a result, they would sometimes attempt to repeat themselves or become confused or frustrated. Thus, we added a Feedback component to the framework to convey the need for feedback from the system to the users informing them about the current state of processing and execution and providing a clear error message if it is unable to process their command correctly.

In comparison to all related work, these components move beyond governing only prompt engineering, instead providing recommendations for designing the system as a whole in a user-centric way.

6.6.2 Implementation

We updated the implementation for the Obi robot using our latest framework, aiming to enhance user acceptance of the system. First, to adhere to the updated Robot Variables component on grounding of physical variables, we changed the continuous speed and acceleration ranges provided to GPT from [0, 80] degrees per second and [0, 250] degrees per second squared respectively to discrete [0, 5] ranges. Next, for the Feedback component, we changed how users interact with the system by replacing the need to hold down a button with a wakeup phrase, "Hey Obi," implemented via the Porcupine Python API from PicoVoice. We also added audio cues to inform users when the wakeup phrase is registered, when their request is being processed, and when the robot is about to execute or has completed an action. For the Instructional Materials component, we created an instructional video and a reference sheet, included as Supporting File 1 and 2, to help users understand and use the system effectively. These materials are expected to be temporary aids as users become familiar with the robot. More details can be found in the published work [169].

Finally, we implemented three new User Control Functions that GPT is able to call, giving users the ability to interrupt and halt the movement of the robot with a verbal command for any reason: `obi.start()`, which begins or resumes execution of any robot code; `obi.stop()`, which permanently stops execution of any currently running robot code; and `obi.pause_indefinitely()`, which suspends execution of any currently running robot code.

6.6.3 Testing with Older Adults

To evaluate Version 3 of the implementation, we conducted a formal study at an independent living facility, Baptist Providence Point in the greater Pittsburgh, Pennsylvania metropolitan area, with 11 non-disabled older adults (5M, 6F) ranging in age from 72-91 (mean = 81.1, SD = 5.9). A subset of the participants are shown in Fig. 6.4. The ethnicity of all participants was white. Older adults were chosen in the study as they share more similarities with the target demographic of individuals with physical impairments as compared to the prior populations the interface was tested with due to their age. Our study, approved by Carnegie Mellon University's Institutional Review Board, involved the researchers traveling to the independent living facility and conducting one-hour-long studies with each participant. Recruitment was conducted in person through an information session held at the independent living facility as well as over email, and informed consent was obtained from the participants for use of image in publication. Participants were asked how much experience they have with controlling a robot on a 7-point scale ranging from "1 = no experience" to "7 = expert user" and how they feel about using robots to help with everyday tasks ranging from "1 = very negatively" to "7 = very positively". Participants had a median experience of 1 (IQR = 1, min = 1, max = 4) and a median attitude of 6 (IQR = 1, min = 4, max = 7).

As seen in Fig. 6.2, we placed the Obi robot and a microphone directly in front of the participant. If needed, we adjusted the food delivery location of the robot to a comfortable location for the participant directly in front of their mouth. As discussed previously, in line with the "Instructional Materials" field of our framework, we started the study by showing the participant an instructional video and providing them with a reference sheet—both included as Supporting Files.

After the instructional session, the participant was asked to select three foods out of the following options: blueberries, yogurt, granola, pudding, pretzels, and Cheerios. As seen in Fig. 6.2, these foods were added to bowls 0, 1, and 2 while bowl 3 is left empty. The first part of the study consisted of 1 practice task, eat a single scoop from any bowl, and 5 predefined tasks: (1) eat a single scoop from bowl 0 quickly, (2) eat a single large scoop from bowl 2, (3) scrape bowl 0 and eat a single scoop from bowl 0, (4) eat three scoops from bowl 1, (5) eat a single scoop from bowl 2 followed by

single scoop from bowl 0. For each task, participants were instructed to only provide a single command. After each task, participants were asked a yes/no question, "Did the robot adequately complete the intended task?" If the participant answered no, they were given an additional attempt to complete the task, up to 3 total attempts. For each task, interaction memory, the previous commands given by the user to the robot and the previous outputs by GPT, was provided to GPT for new attempts.

The next part of the study was a 7 minute open feeding session. If needed, we refilled the bowls with more food. The instructions provided to the participants were "Your goal is to finish the meal in a pleasant and efficient way. You are free to command the robot to eat the food however you like. Feel free to be creative in how you accomplish this goal." For this part of the study, interaction memory was provided to GPT for the entire session.

At the end of the study, we asked participants for responses to the NASA TLX workload items on a 0-100 scale, the System Usability Scale (SUS), multiple 7-point Likert items, and 4 open ended questions. We also audio and video recorded the entire study to capture participant reactions, feedback, and responses to the open-ended questions.

Both qualitative and quantitative data from our study show that our framework and methodology were successful in creating a positively-received speech interface for the Obi feeding robot. As seen in Fig. 6.5, the majority of participants were able to complete the predefined tasks within 3 attempts, showing adequate accuracy for the LLM interface. From the responses to Likert items shown in Fig. 6.6, we find that participants found the interface easy to learn, answering with a median score of 6 (Agree) to L1 (Likert Item 1), "It was easy to learn how to speak to the robot", with one participant (P10) saying, "I would imagine that most people would learn to use the system quickly." They also found the robot easy to control through speech and felt in control when using the robot with median scores of 7 (Strongly Agree) for both items L2 (Likert Item 2) and L5 (Likert Item 5). P5 summarized this at the end of the study saying, "What I liked was it gave me a sense of control. For somebody who's in an incapacitated situation, that would be very important... And it was easy to talk to." Lastly, participants also gave a strong median rating of 7 (Strongly Agree) for items L4 (Likert Item 4) and L6 (Likert Item 6), related to enjoyment while using the robot and if they would use the robot and speech interface if they were unable to

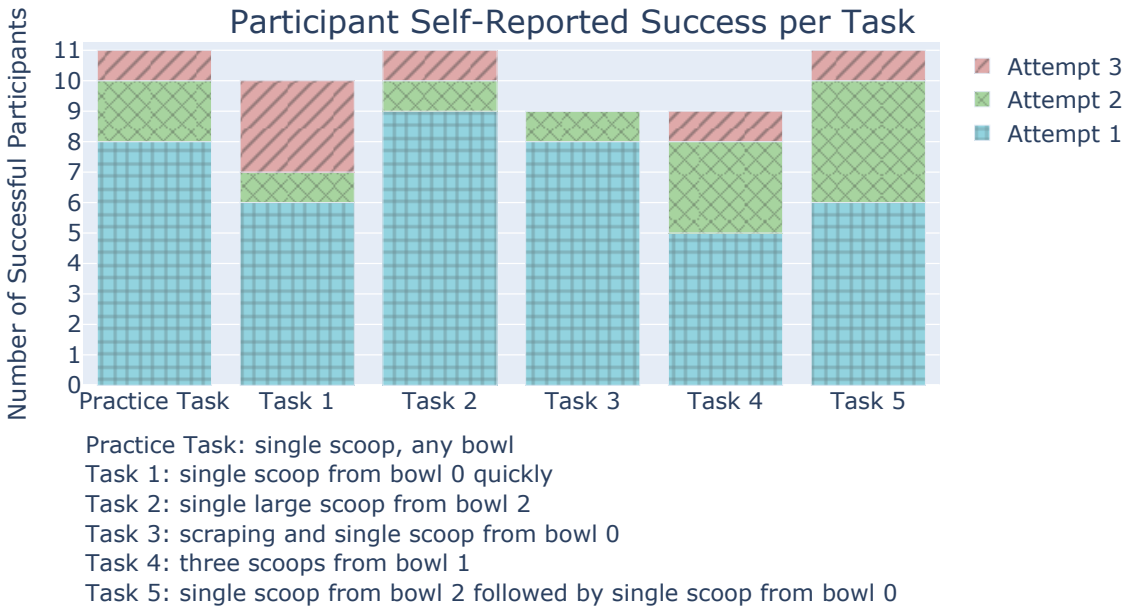


Figure 6.5: Participant self-reported success per task for each attempt. After both the practice task and the 5 predefined tasks, participants reply Yes/No to the question "Did the robot adequately complete the intended task?"

eat independently.

Using the NASA TLX scale, we find that participants generally reported low workload measures across all categories as seen in Fig. 6.7. Lastly, the average System Usability Scale score was 73.0 (SD = 18.6, min = 50.0, max = 95.0), which shows decent performance, but leaves room for improvement. The strong Likert item responses, workload measures, and SUS scores from older adults show promise for using LLMs as assistive interfaces for robot-assisted feeding and other physical tasks. Future testing with more individuals, especially those unable to feed themselves, is needed for full validation of the system.

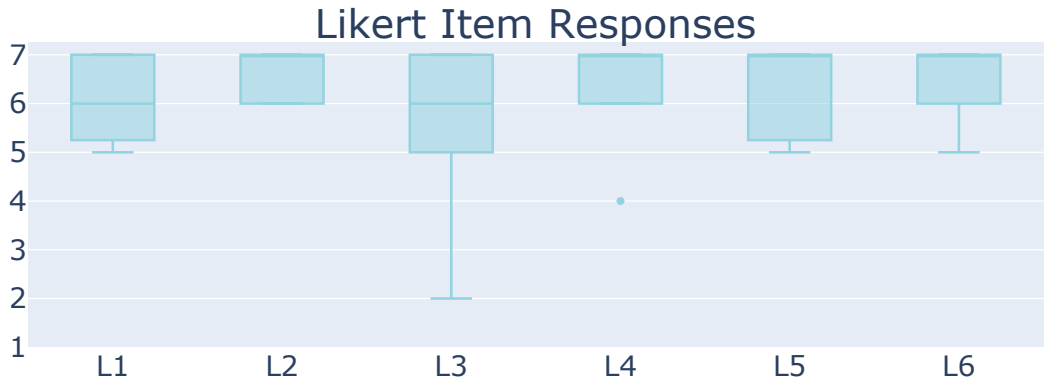


Figure 6.6: Responses from six 7-point Likert Items answered by the 11 participants at the end of the study. A high score is best for all items, with 1 = Strongly Disagree and 7 = Strongly Agree.

6.7 Reflection

6.7.1 Final Framework

We present our final framework for integrating LLMs as assistive speech interfaces, outlined and detailed in Fig. 6.3. The framework now comprises a total of nine components: Environment Description, Robot Functions, Function Applications, Code Specifications, Safety, Robot Variables, Instructional Materials, User Control Functions, and Feedback. In the figure, we indicate if each component pertains to prompt engineering (PE), system rollout (SR), or both prompt engineering and system rollout (PS).

6.7.2 Analysis

To gain deeper insights into the needs of users of LLM-based speech interfaces for robots, we conducted a thematic analysis on audio recordings collected during our

study with older adults. Thematic analysis is a common technique used for analyzing qualitative feedback from participants in human-robot interaction studies [48, 112, 192, 211]. This analysis informed the development of design guidelines for integrating LLMs as speech interfaces for assistive robots. While our framework provides assistance with integrating an LLM as a speech interface, our design guidelines should be strongly considered during development and testing of such interfaces as they can further enhance the user experience.

Initially, we transcribed the audio recordings from our user studies. Subsequently, we uploaded these transcripts into ATLAS.ti [4] and employed deductive coding to extract users' perceptions of the system. Our aim was to comprehend both the effective and challenging aspects of our system; therefore, we coded user feedback into three distinct categories: (1) negative responses to the interface, (2) positive responses to the interface, and (3) desired features.

After coding the transcripts using the defined categories, we proceeded to analyze the quotes within each category using affinity diagramming. This process allowed us to identify and group similar feedback provided by users. We conducted two iterations of coding and affinity diagramming to ensure that we captured all relevant data comprehensively. During this iterative process, we focused on identifying challenges and needs expressed by users and grouped similar feedback together. By examining the collective feedback, we synthesized our findings into actionable guidelines provided below.

6.7.3 Design Guidelines

Based on our analysis of user feedback from the study at an independent living facility, we provide design guidelines, shown in Fig. 6.8, for the incorporation of LLMs in assistive interfaces. The recommendations include (1) Customization (2) Multi-Step Instruction, (3) Consistency, (4) Comparable Time to Caregiver and (5) Social Capability. We discuss each guideline in more detail below.

Customization

Customization refers to the ability of the user to incorporate their preferences while using an interface to complete an assistive task. All 11 participants provided feedback

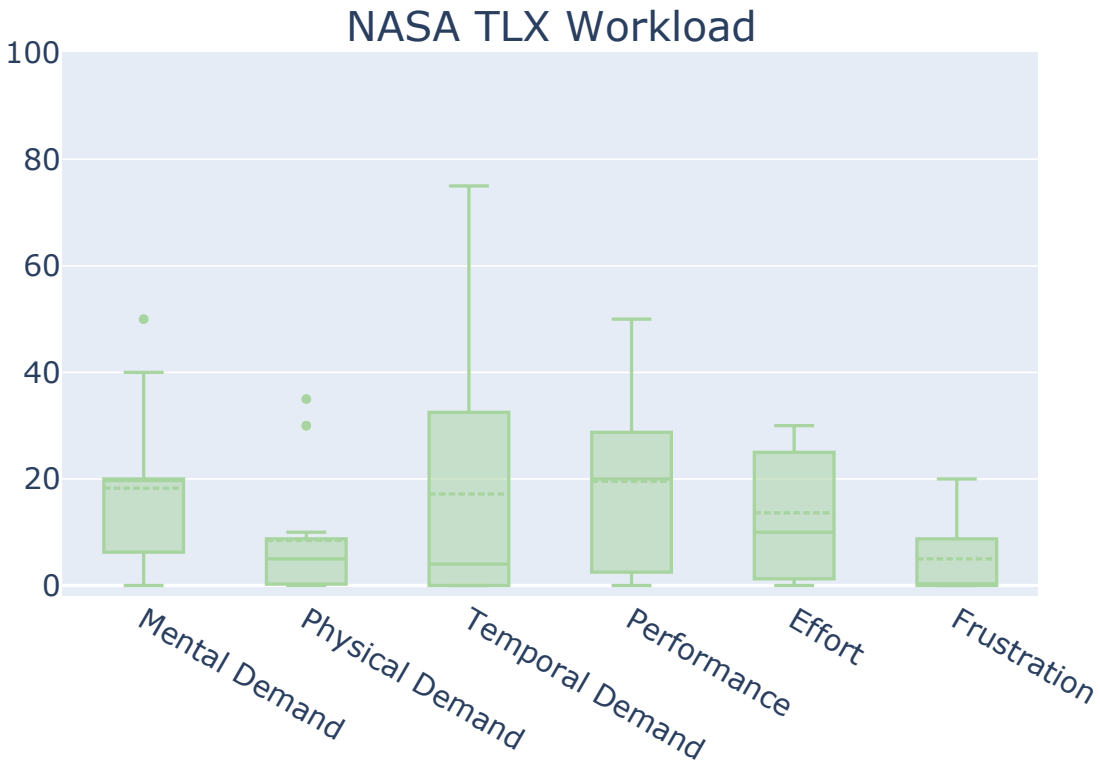


Figure 6.7: Responses to a 0-100 NASA TLX Scale assessed at the end of the study with all 11 participants. A low rating is best for all categories.

CUSTOMIZATION	MULTI-STEP INSTRUCTION	COMPARABLE TIME TO CAREGIVER	CONSISTENCY	SOCIAL CAPABILITY
Should offer customization options, specifically the ability to execute non-predefined commands and modifiers to predefined commands.	Should be capable of executing multiple functions sequentially from a single user command to increase the efficiency and comfort of users	Should execute commands with speeds comparable to a caregiver, ensuring reasonable and personalizable execution times.	Should execute commands consistently, with the same LLM output for differently worded and repeated command phrasings.	Should have the ability to socially engage with the user during execution of the physically assistive task to more closely mimic the role of a caregiver.

Figure 6.8: Design Guidelines. Using both quantitative and qualitative data from a human study with 11 older adults, we establish 5 guidelines for integrating LLMs as speech interfaces for physically assistive robots.

and comments regarding the customization of the robot through the interface, with select quotes from participants depicted in Fig. 6.9. Through our thematic analysis, we find that the ability for users to provide Obi commands customized to their

own preferences (such as how big of a scoop they wanted) was pivotal for successful interactions. This design guideline directly aligns with insights from past works on the importance of user customization of robotic interfaces [168, 188, 189]. From our analysis, there are two types of ways that LLM-based speech interfaces must be able to provide users with customization: (1) Modifiers and (2) Ability to process non-predefined commands.

Modifiers: Modifiers refer to an attempt by the user to customize the execution of commands, using words like "faster," "slower," "larger," and "smaller." For example, for the Obi robot, this might involve a desire for the robot to move at a faster speed while moving to the mouth. Throughout our study, both for the predefined tasks and the open feeding session, participants frequently issued commands to the robot accompanied by modifiers, but some expressed concerns with the system's ability to consistently understand their modifiers.

The quantitative data, shown in Fig. 6.5, specifically Task 1 and 2 which involved increasing robot speed and scoop size, highlights some of these challenges that participants faced with modifiers. For Task 1, only 6 participants were successful in one attempt, which shows room for improvement. P10 requested Obi to increase the speed during feeding, noting that "the change in speed seemed to be not great." Similarly, P2 highlighted issues when asking for different bite sizes saying, "what I didn't like was... [Obi was] not very accurate, especially in terms of sizes of scoops and what it was picking up each time." Specifically, 6 participants expressed challenges with modifying the scoop size. We would like to note that scoop size challenges were primarily due to the lack of a perception system to dynamically change the depth of scoop based on the amount of food in the bowl. Our system instead used a naive approach and included a Robot Variable named "scoop depth" in the prompt that could be modified by the LLM to adjust the depth of the scoop but didn't take into account the amount of food left in the bowl, leading to inconsistency in scoop sizes.

Nevertheless, the importance of modifier customization was underscored by participant quotes that unveiled frustration and confusion when the system didn't respond properly to their modifiers. For example, P5 expressed frustration with changing the scoop size saying, "Trying to get the amounts is a challenge." We want to emphasize that modifiers are personalized, as they may carry varying interpretations for different users. Thus, the system should ideally adjust its execution to align with

the user's expectations. Possible approaches to achieve this include accurate sensing and perception of the environment and learning user preferences, with past work demonstrating LLMs are well suited for this task [232].

Ability to process non-predefined commands: Predefined commands refer to the Robot Functions, Function Applications, and User Control Function components of our framework; for the Obi robot, these commands include scooping from a bowl, moving to the mouth, etc. On the other hand, non-predefined commands denote additional functionalities that users may want, such as mixing for the Obi robot, which are not explicitly defined and described to the LLM. Non-predefined commands are essential to an assistive interface because they allow for increased customization of robot motions in addition to modifiers, allowing the robot to more closely adhere to the user's preferences. However, having an LLM understand and execute on non-predefined commands is challenging, especially while hard-coding robot trajectories in alignment with the Safety component of our framework. This challenge was apparent in our study, with 6 participants expressing difficulties in executing non-predefined commands, depending on the specific request and phrasing of the command. Some common non-predefined commands from the open-feeding session with Obi included mixing of foods and selection of a specific quantity of a food, such as 2 blueberries in one scoop.

While the vast majority of speech interfaces for assistive robots are unable to handle non-predefined commands, an advantage of LLMs is they have the ability to sequentially arrange predefined commands, which serve as 'primitives,' to perform non-predefined commands. In the case of the Obi, during our human study, 5 participants requested the robot to mix foods, a non-predefined command. For 2 participants, GPT was successful, as it consecutively ordered scooping functions from each bowl leading to a bite with multiple foods as seen in the sequence of images in Fig. 6.1. However, the inability to execute non-predefined commands for some participants led to confusion. For example, P1 remarked "Maybe with the mixing of the food together... [it] didn't seem to know what mix meant."

Summary: Both predefined/non-predefined commands and modifiers aren't exclusive to the Obi robot; interfaces for other assistive robots, such as mobile manipulators, often contain predefined commands such as autonomous grasping of an object and modifiers such as the ability to change robot speed. Our thematic

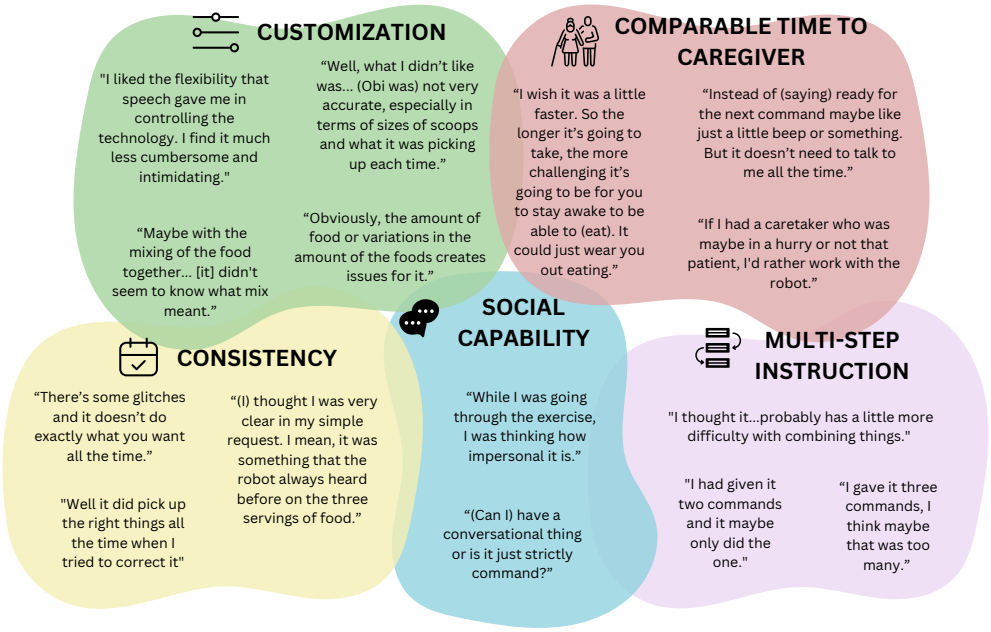


Figure 6.9: A subset of quotes from our study involving 11 older adults is presented and organized according to the categories outlined in our design guidelines.

analysis showed that participants wanted the ability to customize their assistance using both modifiers and non-predefined commands. When these requests were not fulfilled, this led to frustration and confusion for users, underscoring the importance of incorporating these customization considerations while designing LLM-based speech interfaces. In contrast, when the LLM-based interface worked well for customization, users expressed strong approval; for example, P10 said "I liked the flexibility that speech gave me in controlling the technology. I find it much less cumbersome and intimidating."

Multi-Step Instruction

Multi-Step Instruction refers to the capability of an LLM-based speech interface to execute sequential robot actions encompassed within a single user command to the robot. This guideline is closely tied to the Function Applications component of our framework in which we discuss how to prompt an LLM to understand how to combine robot functions sequentially. From our thematic analysis, we find that multi-step instructions are important to users as they can help increase system acceptance by

reducing frustration and user workload. 5 participants commented on multi-step instructions, with a subset of quotes shown in Fig. 6.9. Some examples of multi-step instructions for the Obi robot are predefined Task 4, eating three scoops from bowl 1, and predefined Task 5, eating a single scoop from bowl 2 followed by a single scoop from bowl 0. As seen in Fig. 6.5, the system worked adequately for the respective tasks, with 9/11 and 11/11 participants completing the task within 3 attempts. However, there is still additional room for improvement as the system didn't work for a subset of users during their initial attempts. P3 summarized their experience, saying "I thought it... probably has a little more difficulty with combining things" while P7 stated, "I gave it three commands, I think maybe that was too many." Consistent functionality of multi-step instructions is crucial to consider when using LLM-based speech interfaces for assistive robots as they can increase the efficiency and comfort of users, while reducing fatigue and frustration. Multi-step instructions also enable users to conduct activities such as watching TV or conversing, while having confidence that the robot will continue the task without additional input for an extended period of time.

Comparable Time to Caregiver

Utilizing assistive interfaces to control robots often results in individuals with impairments taking substantially more time to complete tasks in comparison to non-disabled individuals and caregivers [37, 168]. LLM-based speech interfaces for robots should strive to execute commands with speeds comparable to a caregiver, thereby enhancing user comfort and satisfaction. In total, 4 participants provided feedback on different aspects of the robot's execution time, with a subset of quotes shown in Fig. 6.9. The robot's execution time encompasses networking and code generation time by the LLM, the duration for the robot to move and carry out the command, and the time for the robot to provide feedback to the user on its state. P10 commented on the feedback time, saying "Instead of (saying) 'ready for the next command' maybe like just a little beep or something. But it doesn't need to talk to me all the time." This insight ties directly to the Feedback component of our framework, and users of our framework should be mindful of providing feedback to users as succinctly as possible. In comparison, P1 highlighted the time taken for the entire feeding process, saying

"I wish it was a little faster. So the longer it's going to take, the more challenging it's going to be for you to stay awake to be able to [eat]. It could just wear you out eating." P1's comment highlights the impact of slow execution times on user fatigue, emphasizing the need for interfaces to align with the personalized time-frames typically provided by caregivers. In contrast to the 4 participants who felt the feeding process was too slow, P11 commented on their preference for the system over a rushed caretaker, saying, "If I had a caretaker who was maybe in a hurry or not that patient, I'd rather work with the robot."

Additionally, as seen in Fig. 6.6, participants responses to L3 (Likert Item 3), related to controlling the robot in a reasonable amount of time, had high variance, indicating that preferences regarding time-frames are strongly individualized. This observation is also evident in participant responses to temporal workload on the NASA TLX scale, shown in Fig. 6.7.

Our findings indicate that executing commands within a similar time-frame as a caregiver would help improve user experience with LLM-based speech interfaces. This can be accomplished by giving precedence to preventing fatigue caused by temporal demand, identifying the specific moments during robot execution when users are prone to experiencing it, and streamlining or allowing customization to the level of communication needed between the user and the interface.

Consistency

Consistency refers to the ability of an LLM-based speech interface to reliably generate and execute commands. Specifically, these interfaces should produce the same output when given identically phrased commands repeatedly, and when given commands with similar intention, but that are differently worded. 6 participants commented on the system's limitations in reliably executing commands as expected. Some of these quotes are shown in Fig. 6.9. For example, P4 said, "(I) thought I was very clear in my simple request. I mean, it was something that the robot always heard before on the three servings of food," suggesting that Obi performed differently when executing the same command on different occasions. Despite participants encountering occasional consistency issues, we found that they were often able to rectify the behavior of the LLM. P2 remarked, "Well it did pick up the right things all the time when I tried to

correct it."

A possible solution for improving LLM consistency is to modify parameters such as the temperature parameter for GPT-3.5 Turbo, which defines the randomness of LLM responses. In Section 6.7.4, we also discuss how LLM fine-tuning can be a possible remedy. Nevertheless, from the perspective of the user, it is essential that these interfaces consistently execute commands, as this builds trust and improves the system's usability. We recommend that LLM-based speech interfaces should be evaluated extensively for consistency before deployment.

Social Capability

Social capability pertains to allowing a user to interact socially with a robot while receiving physical assistance, similar to interactions with a human caregiver. 3 out of 11 participants provided insights, with some of the comments showcased in Fig. 6.9, regarding the integration of social capabilities into the Obi system. P11 said "While I was going through the exercise, I was thinking how impersonal it is." They further elaborated that they were unsure about whether they would prefer to be fed by Obi or a caretaker because of its lack of conversational abilities. While none of the participants attempted to converse with Obi, P8 inquired about it, saying "(Can I) have a conversational thing or is it strictly command?" The feedback from users demonstrates the significance of integrating social capabilities, even though such capabilities are often not the primary focus of physically assistive robots. The presented social capability guideline is in agreement with past work on using LLMs as speech interfaces for social robots that shows that robot's social aspects can enhance user engagement [112]. We encourage researchers to explore the social capabilities of LLM-based speech interfaces for physically assistive robots, as this could allow the robot to provide a sense of companionship and engagement similar to that of a caregiver, thereby potentially increasing user acceptance.

6.7.4 Considerations and Future Work

In addition to the testing already conducted, a follow-up study involving individuals who are unable to feed themselves, including long-term deployments, would be beneficial to our framework and design guidelines. These additional studies will

enable us to determine if the LLM interface in conjunction with the robot meets specific needs and desires of users. Furthermore, it would be useful to conduct our framework iteration process with another assistive robot such as a mobile manipulator that can do multiple tasks in the home. This testing could present unique insights that users of our framework may need to consider for their specific robot and tasks.

Additionally, as seen in Fig. 6.5, the accuracy of our speech interface for certain feeding tasks was subpar, especially on first attempts by participants. Consistency of the interface is an important consideration for LLM speech interfaces as presented in our design guidelines, Fig. 6.8, and explained in Section 6.7.3. Since the testing of our interface, which used OpenAI's GPT-3.5 Turbo model, there have been several advancements in LLMs and their performance. We urge other researchers to test out various closed and open-source LLMs and benchmark their performance as speech interfaces before selecting one. Fine-tuning of the LLM using commands gathered from study participants could also help address subpar accuracy [112, 224].

In Chapters 4, 5, and 6, we explore active and shared control interfaces for caregiving robots, which require users to issue explicit commands. In Chapter 7, we introduce a passive sensing wearable interface for feeding robots, allowing the robot to infer appropriate moments to act without direct input. This approach has the potential to further reduce user workload while preserving a sense of control.

Chapter 7

A Wearable Approach to Bite Timing Estimation in Robot-Assisted Feeding

7.1 Contributions

For the following chapter, the contributions are as follows. I formulated the core ideas for this project with Zackory Erickson, Tapomayukh Bhattacharjee, Jessie Yuan, and Rajat Kumar Jenamani. I led the project and contributed to all aspects of it. Anthony Wertz designed the custom Inertial Measurement Units (IMUs) and programmed the firmware for the IMUs and the IMU hub). Victoria de León assembled the IMU boards. Jessie Yuan wrote all code for the Phase 1 and Phase 2 studies. Tanisha Mehta and I wrote the ML pipeline with help from Jessie Yuan. Jessie Yuan and I compiled all final results. Jessie Yuan, Tanisha Mehta, Janavi Gupta, and I ran the Phase 1 and Phase 2 studies at CMU. Rajat Kumar Jenamani and Ben Dodson set up Phase 3 studies with the FEAST robot at Cornell University and ran these studies. I wrote the paper and all authors provided feedback. Zackory Erickson, Tapomayukh Bhattacharjee, and Carmel Majidi advised on all aspects of the project.

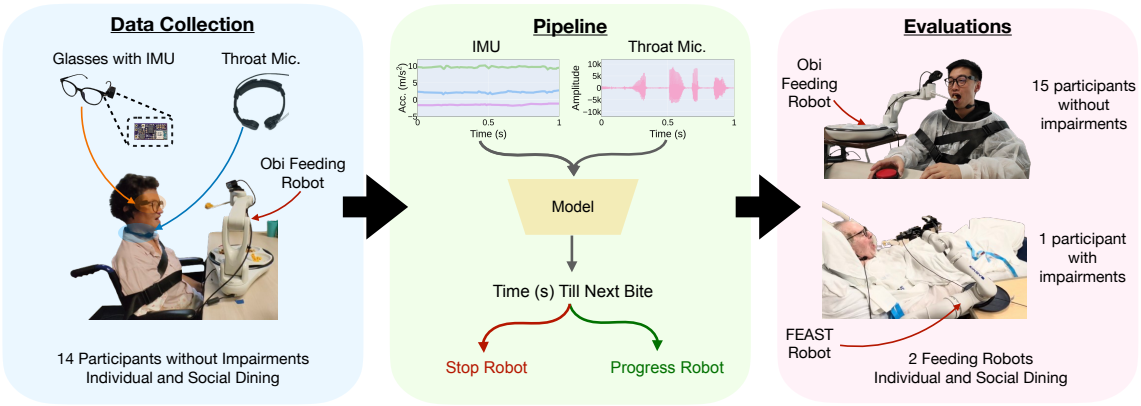


Figure 7.1: We present **WAFFLE**, **W**earable **A**pproach **F**or **F**eeding with **LE**arned Bite Timing, a bite timing algorithm for assistive feeding robots that generalizes across users with and without motor impairments, feeding robots, robot positioning, feeding trajectories, foods, and both individual and social dining settings. We collect data from 14 participants without impairments using the Obi robot, then train a machine learning model to estimate the amount of time until the next bite. For deployment, a user-selected assertiveness threshold allows for the conversion of predictions into robot stop and proceed commands. Lastly, we evaluate WAFFLE with 15 participants without impairments and 1 participant with motor impairments using two different robots.

7.2 Introduction

According to the World Health Organization, nearly 190 million people worldwide live with motor impairments [215], often needing assistance with Activities of Daily Living (ADLs) such as eating, bathing, and dressing [57, 88]. Specifically, around 3.5 million people in the United States alone require assistance with feeding, with this estimate increasing as the population ages [88]. For individuals with impairments, feeding assistance can yield feelings of anxiety and self-consciousness, and a sense of burden [96, 151, 203]. Providing feeding assistance is also time-consuming and adds to the high workload for caregivers, with studies indicating that each meal takes about 35-40 minutes of their time [25, 70, 205].

Physical caregiving robotics have shown their utility in a variety of activities of daily living (ADLs) including feeding assistance, increasing independence, quality of life, and well-being for individuals with impairments [136, 151, 152, 167, 168, 238]. Although commercial feeding robots like the FDA-compliant Obi Robot [1] are available, these systems have not seen widespread adoption, in part due to technical limitations in food acquisition, bite transfer, bite timing, and user interfacing. Notably, Obi relies on manually programmed trajectories and does not adapt its motions to variations in user behavior or food positioning, limiting its flexibility in real-world use. Recent research has focused on addressing these challenges [100, 102, 153, 162, 169, 213]. In this work, we focus on bite timing, the appropriate moment for a robot to deliver food to a user. Bite timing is an important component of robot-assisted feeding with prior research highlighting the value users place on having multiple options for triggering bite timing and on the adaptability of the timing itself [25, 102, 151]. Most existing approaches rely on users manually initiating bite timing, often through actions like opening their mouth. However, such methods can be distracting and intrusive, particularly in social dining contexts [151, 162]. Ondras et al. presented the SoNNET algorithm, which utilizes depth cameras and a microphone array to predict bite timing during social dining situations [162]. While their study showed that participants preferred the predicted bite timing over two baseline methods, the approach relies on highly constrained conditions, specifically requiring two accompanying diners in a social setting, precisely positioned cameras and a microphone array at the center of the dining table, and minimal background noise. These constraints limit adoption in

the real world including in setting such as restaurants where there may be privacy concerns due to the use of the cameras and microphone. Moreover, the method does not generalize to common scenarios such as individual dining and was not tested with a diverse population, including individuals with motor impairments who may exhibit different behaviors. These limitations underscore the need for a more generalizable and privacy-preserving approach to bite timing estimation.

Caregivers often rely on behavioral cues, such as head movements, chewing, and speaking, to determine when to offer the next bite during assisted feeding. Visual approaches using external mounted cameras have also shown promise for chewing and swallowing detection but require the user to be within the view of the camera continuously [91, 220]. To overcome these limitations, we present a system named **WAFFLE**: **W**earable **A**pproach **F**or **F**eeding with **LE**arned Bite Timing, consisting of wearable sensors for estimation of bite timing for both individual and social dining scenarios, designed to work with for any robotic feeding platform positioning, any robot feeding trajectory, any foods, and any user. The wearables, an inertial measurement unit (IMU) mounted on a pair of glasses and a throat contact microphone worn around the neck, provide our system with cues often used by caregivers during the feeding process including a participant's head movements and when they are chewing and talking. The wearables are privacy-preserving as the throat microphone does not record ambient sounds. In order to develop a supervised learning algorithm for bite timing estimation, we conduct a data collection study with 14 individuals without motor impairments with the Obi feeding robot [1] feeding from the front. We use this wearable data to train a bite timing regression algorithm to estimate the time until the next bite. To include user preference into how assertive the robot is during deployment, we use a simple user-selected assertiveness threshold to convert predictions into progress / stop commands for the robot. We conduct two evaluation studies to evaluate WAFFLE. The first is with 15 individuals without motor impairments using the Obi robot and feeding from the side. In this study, we compare our WAFFLE (Wearables) method with two baselines, Mouth Open and Fixed Interval, and show that our method performs statistically on par with or better than the baseline methods in both individual and social dining based on all metrics including feeling of control, robot understanding, and seamlessness. We also find that the Wearables method is strongly preferred by participants with 13 out of 15

participants ranking it 1st among all methods. Our method additionally performs strongly in a qualitative study with a single participant with motor impairments and the FEAST system [102], consisting of a Kinova 7DOF robot arm, feeding from the front, showing generalization to participants with impairments and other feeding robots.

The contributions of this work are as follows:

- We introduce WAFFLE, a novel, reactive bite timing approach trained on data from two wearable sensors collected from 14 participants.
- We conduct a human study with 15 participants without motor impairments to evaluate WAFFLE against two baseline methods, demonstrating appropriate bite timing, high perceived feeling of control, and low workload in both individual and social dining contexts.
- We deploy WAFFLE on a second type of feeding robot and provide qualitative evidence of its effectiveness through a human study involving one participant with motor impairments.

7.3 Related Work

7.3.1 Robot-Assisted Feeding

Past research in robot-assisted feeding has primarily focused on food acquisition [26, 64, 72, 76, 78, 101, 109, 131, 212, 213], bite transfer [23, 100, 202], and system/interface design [20, 25, 28, 77, 89, 102, 151, 153, 169, 176, 242]. For bite timing, many of these works have relied on manual triggering by participants through a button or assistive interface, fixed-interval timing, or open-mouth detection [28, 102, 151, 176, 179, 200]. In contrast, Ondras el al. introduced the SoNNET algorithm, which employs depth cameras and a microphone array to predict bite timing during social dining scenarios [162]. The authors gathered an initial dataset where participants without impairments ate normally with 2 co-diners with 3 RGBD cameras and a microphone array in the middle of the table. Their algorithm performs binary classification utilizing features extracted from RGBD cameras and microphone arrays to determine whether the user intends to take a bite and commits to progressing along the entire

feeding trajectory when it predicts that the user is prepared. These algorithms are not generalizable in the real world as they are constrained to very specific conditions: having two accompanying diners and placement of sensors in the middle of the table. The researchers also didn't test their system with participants with impairments, with other types of feeding robots, and with various positioning of the robot and sensor array. While their findings also indicated that participants preferred the predicted bite timing over the baseline strategies (Mouth Open and Fixed Interval timing), no statistical significance was observed between mouth open and their learned timing for the majority of metrics, suggesting room for improvement. In comparison, our method works in both individual/social dining scenarios, uses common wearable sensors, and was evaluated with a participant with impairments and two different robotic systems in various orientations. Our method also achieves statistical significance across many metrics in comparison to the baselines.

7.3.2 Wearable Sensing

Wearable sensing involving accelerometers, inertial measurement units (IMUs), and contact microphones has been useful for identification of various human movements [87, 166, 239] and for control of mobile robots for activities of daily living during robotic caregiving tasks [167, 168, 171]. Past work has shown that wearables can successfully capture chewing and swallowing using acoustic, inertial, electromyography (EMG), piezoelectric, or a fusion of the aforementioned sensing methods [22, 184, 199, 226]. Positioning of the sensors on the human body is important with wearable sensors often mounted on the inner/outer ear, head, neck, and chest [226]. Insights from these past works on the types of sensing methods and the placements of these sensors show promise for using a fusion of wearable sensing modalities, specifically IMUs and contact microphones, for estimation of bite timing for robot-assisted feeding.

7.4 Methods

Bite timing is an extremely challenging problem: it often depends on a subconscious sense of when it feels appropriate to take a bite, especially during social dining scenarios. The problem becomes more complicated since there are often many valid



Figure 7.2: Study Setup for the Three Phases. **A.** Phase 1 Study Setup. In this study, we collect bite timing data from 14 participants with no motor impairments with the Obi robot feeding from the front of the participant. **B.** Phase 2 Study Setup. In this study, we compare our wearables approach, WAFFLE, to two baselines with 15 participants without motor impairments with the Obi robot feeding from the side of the participant. **C.** Phase 3 Study Setup. In this study, we evaluate our wearables approach with 1 participant with motor impairments with the FEAST robot [102] feeding from the front of the participant.

timings and robot motions that result in an appropriate bite. Bite timing is also very dynamic; for example, a person may lift food to their mouth, pause, and only proceed to eat at a very specific moment. We hypothesize that a robot can achieve appropriate bite timing by being highly reactive to natural user cues from wearable sensor data, including chewing, speaking, and head movements. Our goal is to develop a system that is generalizable across all variables in a robot-assisted feeding setting, including different users, feeding robots, robot positioning, feeding trajectories, foods, and individual and social dining scenarios.

To train a model capable of using wearable data to estimate bite timing, we propose a supervised, data-driven machine learning algorithm. In the following sections, we describe our methodology for data collection used to train the algorithm, the architecture of our machine learning model, and our deployment approach, followed by a description of two evaluation studies involving both individuals with and without impairments. The setup for all three studies are shown in Fig. 7.2 and compared in table format in Table 7.1.

Table 7.1: Overview of the Three Human Studies

	Phase 1	**Phase 2**	**Phase 3**
Type of Study	Data Collection	Evaluation	Evaluation
Number of Participants	14	15	1
Demographic	No Motor Impairments	No Motor Impairments	Motor Impairments
Robot Type	Obi [1]	Obi [1]	FEAST [102]
Robot Positioning	Front	Side	Side
Feeding Trajectory	Front	Side	Front
Foods	Assorted	Assorted	Bananas
Settings	Individual & Social	Individual & Social	Individual
Bite Timing Methods	Participant Controlled & WoZ	Wearables & 2 Baselines	Wearables

7.4.1 Phase 1: Data Collection Study

To collect a dataset of wearable data for training a supervised learning algorithm, we conducted a human study, referred to as Phase 1, with 14 participants (7 M, 7 F) ranging in age from 19 to 60 (Mean = 26.07, SD = 10.16). Three additional participants were excluded, two because they took very few bites because they were not hungry during the session, and one due to robot and sensor issues. We received approval from Carnegie Mellon University's Institutional Review Board. Participants were asked how much experience they have with controlling a robot and their level of hunger.

As illustrated in Fig. 7.2, participants wear two wireless Inertial Measurement Units (IMUs) mounted on glasses, two additional wireless IMUs embedded in earbuds, and a commercially available throat microphone. The IMU boards [14], shown in Fig. 7.3, are custom designed and fabricated and are housed in 3D-printed enclosures affixed to the glasses and in the 3D-printed earbuds.

Participants sit in a wheelchair with their movement restricted by two straps, simulating the condition of someone with severe motor impairments as suggested by an Occupational Therapist consultant. The commercially available Obi assistive feeding robot from DESĪN LLC [1] is situated in front of the participant. We additionally mount an Intel Realsense D435i RGBD camera on the robot arm.

A companion laptop running ROS 1 Noetic is used to simultaneously collect RGB images and depth images from the Realsense at 15 Hz, 3-axis accelerations and quaternions from each of the four IMUs at 200 Hz, and audio from the throat microphone at 44.1 kHz. There are two push buttons additionally attached to the laptop, which are used to control the Obi robot. Lastly, a small display monitor is

placed in front of the user to indicate which bowl the robot will be scooping from next.

We allow participants to pick a soft, medium, and hard food and an additional wildcard food from any of the 3 categories to fill the 4 bowls of the Obi robot. Before starting data collection, we kinesthetically move the robot to record the feeding position directly in front of the user's mouth. Next, we introduce the participant to the system by having them eat 2 bites of food and to try rotating between the 4 bowls using the red button.

We structure our data collection as follows. The robot automatically scoops food from the selected bowl and moves a staging point about 15 inches from the participant's mouth. Six participants do a participant-controlled version of the study, while the other eight participate in a Wizard of Oz (WoZ) version. For the participant-controlled version, between the staging and feeding points, the participants use a yellow button, which moves the robot towards the mouth when it is pushed and held down. When the button is released, the robot doesn't move. For the Wizard of Oz version, participants are told that the robot is fully autonomous, and a researcher uses the yellow button to control the robot between the staging and feeding points. The researcher specifically looks for cues from the user that caregivers often use, including chewing, talking, and glancing at the robot. They follow a set of guidelines we developed to decide when to progress the robot and stop it. These guidelines are listed in order of priority as sometimes cues can conflict: for example, the person may still be chewing while also glancing at the robot to indicate they want another bite.

We conduct both the Wizard of Oz and participant-controlled versions of the study because each provides valuable bite timing data for training our algorithms. The Wizard of Oz condition offers more structured training data as the experimenter directly controls the robot's movements, ensuring consistent robot movements across participants. This structure makes it easier for the model to learn generalized patterns by minimizing variability. Additionally, the Wizard of Oz condition enables us to collect data on interactions with an 'autonomous' robot, including participants' reactions to occasional bite timing errors. However, structured data alone may not capture the full range of nuanced user preferences in robot motion and bite timing, which is why participant-controlled data is additionally valuable. For both conditions, the researcher presses a button on the companion laptop after the participant takes

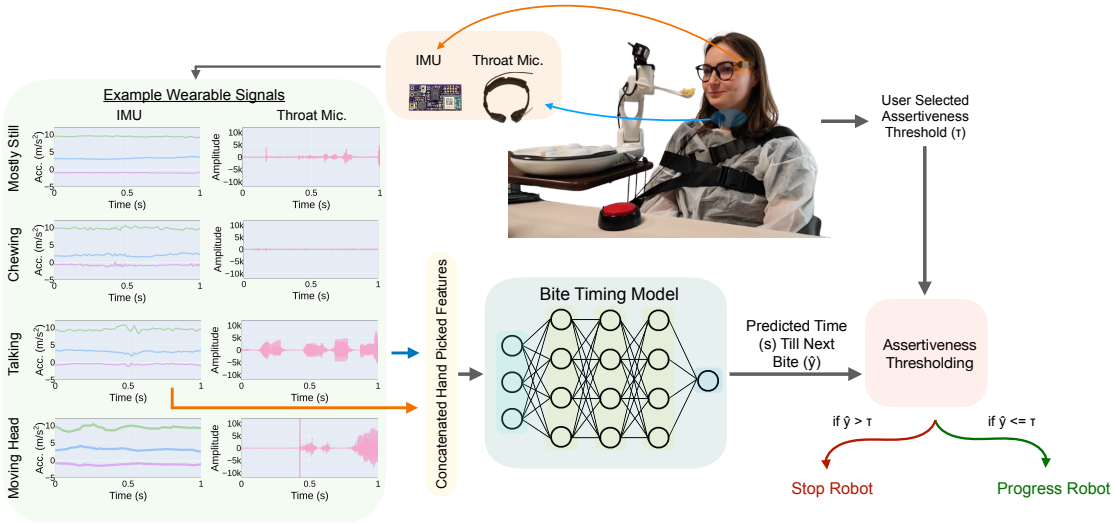


Figure 7.3: WAFFLE Deployment Pipeline. Features from 1 second windows of IMU and Contact Microphone data are inputted into our bite timing machine learning model which outputs $\hat{y}$, the predicted time (s) until the next bite. We use a user selected assertiveness threshold, τ, to stop and progress the robot along the trajectory.

the bite, commanding the robot to proceed with the next scooping action.

For both participant-controlled and WoZ studies, for every participant, we conduct both individual and social dining scenarios, with the order counterbalanced for all participants. For individual dining, the user has 6 minutes to eat using the robot without any external interactions. For social dining, the user has 12 minutes to eat using the robot while conversing with a researcher who sits in front of them. During the debriefing of the Wizard of Oz version of the study, none of the participants suspected that the robot was being controlled by the researcher.

7.4.2 Bite Timing Model

We propose a machine learning regression algorithm that estimates the number of seconds until the next bite using a throat microphone and a single IMU mounted on the left side of a glasses frame. To ensure that our approach generalizes across different robot platforms, participant-robot orientations, and feeding trajectories, we deliberately exclude RGB camera data collected during the study. In real-world settings, feeding robots do not consistently face the user while waiting for bite

initiation; instead, they are often positioned above the plate or in a retracted location to avoid obstructing the user's social interactions. Even when the robot is oriented toward the user, the participant's face can move out of frame due to natural head motion during feeding. Moreover, because the camera is mounted on the robot arm, features such as facial keypoints are heavily influenced by the robot's own movement rather than the participant's behavior. These factors make camera-based features unreliable for capturing user intent, and thus less useful for accurate bite timing estimation. For the scope of this paper, we also omit data from the other three IMUs (right glasses, left earbud, right earbud), as their signals closely resemble those from the single IMU located on the left side of the glasses.

As shown in the raw data signals in Fig. 7.3, the IMU's acceleration signals and the throat microphone's signals offer complementary insights into head movements, chewing, and talking. The IMU effectively captures head movements and chewing activity, while the throat microphone, positioned on the throat, excels at detecting vibrations from swallowing and vocal cord activity during speech.

We use a regression algorithm that estimates the time until the next bite in seconds instead of directly predicting a binary progress/stop command for two reasons. Firstly, it allows for a generalizable model that can be used with various deployment strategies based on the robot and the participant. For example, predictions can be binarized into proceed or stop commands for the robot or predictions can be used directly as the number of seconds the robot should pause until progressing continuously through the feeding trajectory. Additionally, we observed that participant controlled progress and stop commands were often noisy and inconsistent, even within the same individual, making it difficult for the model to reliably learn associations between wearable signals and robot motion. As an example, during one feeding trajectory, a participant might partially advance the robot mid-chew and then pause to finish chewing, while in another, they might wait until completely done chewing before commanding the robot to proceed fully to the mouth along the trajectory; both of these trajectories are valid for correct bite timing.

Our regression algorithm is visually shown in Fig. 7.3. First, 1-second windows of 3-axis IMU accelerations and throat microphone data are resampled using 1 dimensional interpolation to 200 Hz and 100 Hz respectively. We specifically select short 1 second windows as this minimizes latency and allows our feeding robot to be very reactive

to human behavior and cues. Next, we split each 1 second sample into two 500 ms low-level windows. For each low-level window, for each of the 3 acceleration axes of the IMU and the 1 axis of the throat microphone, we extract 6 hand-picked features from the acceleration: maximum, minimum, mean, standard deviation, range, and root mean square (RMS). Using low-level windows with hand-picked features has shown promise for activity recognition including chewing detection in past works [22, 170, 194]. We concatenate the features from both sensing modalities and normalize them to zero mean and unit variance. In total, there are 48 features (2 low level windows $\times$ 4 axes $\times$ 6 features per axis). The concatenated, normalized features, $f \in \mathbb{R}^{48}$ are inputted to our bite timing ML model.

Our bite timing algorithm is a multi-layer perceptron with three fully connected layers of sizes (128, 64, and 1). ReLU activation and dropout with a probability of 0.1 are applied after the first and second layers. The final layer directly outputs the predicted time, $\hat{y}$, until the next bite in seconds. We train the model using ground truth labels of time until the next bite, collected from our human study. We observe that behavior more than 10 seconds before a bite, such as chewing, talking, or head movements, is similar to behavior around the 10 second mark. To limit the influence of these distant and less informative events during model training, we cap all ground truth labels at 10 seconds. This reduces noise from larger labels that will not improve bite timing prediction.

The model is trained using mean absolute error loss and the Adam optimizer for 100 epochs with a learning rate of 0.0001 and batch size of 128. To assess how well our algorithm performs, we evaluate it using Leave-One-Subject-Out Cross-Validation (LOSO-CV) on all participant data. Model performance is measured using mean absolute error (s). We train the final model for our evaluation studies on data from all 14 participants.

7.4.3 Assertiveness Thresholding

To deploy our trained model and enable reactive robot behavior, we use a thresholding approach to convert the predicted time until the next bite into a binary command (0 to stop, 1 to progress the robot). We refer to the threshold value, τ, as the user-selected assertiveness threshold, as it reflects how confidently the robot should

act on its prediction. Lower thresholds result in more cautious behavior while higher thresholds produce more assertive feeding actions. This thresholding incorporates user preferences and interpretability into the system and is motivated by observations from participant videos in our Phase 1 study that revealed that participants preferred different levels of assertiveness. For example, in the participant-controlled studies, we observed that some participants would only progress the robot when they were fully done chewing, while others would do so even when they were still lightly or heavily chewing. These observed behaviors highlight the need for user-controlled assertiveness, which cannot currently be predicted by the current system, and may additionally change over the course of the meal or during different dining scenarios.

Our thresholding method is shown visually in Fig. 7.3. If the predicted time ($\hat{y}$) until the next bite is such that $\hat{y} > \tau$, a stop moving command (0) is sent to the robot. Otherwise, if $\hat{y} <= \tau$, then a proceed moving command (1) is sent to the robot. Our bite timing model generates predictions at 2 Hz. The predictions are thresholded using the user selected assertiveness threshold and the binary progress/stop command is sent to the robot at 2 Hz.

To quantify how well our bite timing model would perform in deployment with thresholding, we use the ground truth 0 and 1 robot motion labels from the Phase 1 study, which were provided by the participant for the participant-controlled studies and by the researcher for the Wizard of Oz studies. We generate predictions (in seconds) from our model and convert these to binary (0 and 1) predictions using the threshold. We then calculate binary metrics using LOSO-CV, evaluating only the participants who completed the participant-controlled Phase 1 study. The binary classification metrics we use are accuracy and the normalized Matthews Correlation Coefficient (nMCC), where 0 indicates perfect negative correlation, 0.5 represents no correlation, and 1 indicates perfect positive correlation. We chose nMCC instead of F1 score because it better accounts for class imbalance and more effectively captures the quality of both true positive and true negative predictions. Both are equally important in our context, as they represent the robot's decisions to reactively proceed or stop along the feeding trajectory. These binary metrics do not reflect the appropriateness of the bite timing itself, but instead serve as an indicator of how closely the robot's start and stop actions (as determined by thresholding) align with the actual trajectories from the Phase 1 study.

We calculate the aforementioned binary metrics using a fixed threshold of $\tau = 6$ seconds which produced the highest nMCC and accuracy scores across all discrete threshold values from 4 to 8 seconds. We also found a fixed threshold of $\tau = 6$ seconds worked best in practice through testing among the research team and pilot studies. To show the benefit of user-selected variable thresholds, we also calculate binary metrics using the optimal threshold for each participant, out of $\tau \in \{4, 5, 6, 7, 8\}$ seconds, that leads to the highest nMCC score. In a deployment, we assume that participants would be able to choose this optimal threshold for themselves after getting fully acclimated with the system. During deployment, we transform the assertiveness threshold (originally in units of seconds) into a mapped assertiveness threshold. This new scale ranges from 1 (least assertive) to 5 (most assertive).

7.4.4 Phase 2: Evaluation with Individuals without Impairments

In order to directly compare our wearables method with bite timing baselines, we conducted a comparison study, named Phase 2, with individuals without impairments. This study was completed with 15 participants (6 M, 9 F) with ages ranging from 19-28 (Mean = 22.27, SD = 3.1). We received approval from Carnegie Mellon University's Institutional Review Board. Similar to the Phase 1 study, participants were also asked how much experience they have with controlling a robot and their level of hunger.

The two baselines we used were Fixed Interval and Mouth Open. We do not compare our method directly with Ondras et al.[162], due to the custom hardware and setup needed and as it can only be used for social dining. In line with past work on bite timing [162], for the Fixed Interval method, the robot feeds the participant every 45 seconds. The Mouth Open method is the most commonly used trigger for bite timing in robot-assisted feeding studies [28, 151, 176, 179, 200]. For the Mouth Open baseline, in line with [162], after reaching the staging point, the robot announces "Please open your mouth when you are ready to eat." The participant must then turn towards the Intel Realsense D435i RGBD camera mounted on the robot's arm and open their mouth to indicate they are ready for the next bite. Once the robot detects the user's mouth is open, it proceeds automatically through the entire feeding

trajectory.

Similar to the Phase 1 study, as seen in Fig. 7.2, participants sit in a wheelchair with their movement restricted by two straps, simulating the condition of someone with severe motor impairments. Participants are outfitted with only a single wireless Inertial Measurement Units (IMU) mounted on the left side of the glasses and the throat microphone. Additionally, we place the commercially available Obi assistive feeding robot from DESIN LLC [1] to the side of the participant, to highlight the ability of our algorithms to generalize to any robot positioning and feeding trajectory. This study setup is shown in Fig. 7.2B.

We allow participants to pick 2 medium and 2 hard foods to fill the 4 bowls of the Obi robot. We remove the option of soft foods because there is more risk of food spillage on the participant due to robot feeding from the side. Before starting the data collection, we kinesthetically move the robot to record the feeding position directly in front of the user's mouth.

We begin with a practice session to familiarize participants with the system and the three feeding methods. For the two baseline methods, we explain their mechanisms and allow participants to try each once. For the wearables method, participants complete three practice trials to experience how specific cues including speaking, chewing, and large head movements, pause the robot's motion, while minimal head movement allows it to continue along its trajectory. To minimize participant confusion during the study and due to the limited number of trials per method, we fix the mapped assertiveness threshold at a value of 3, which based on quantitative results presented in Section 7.5.1 and piloting, offers a balanced level of performance for most participants.

We structure our study as follows. There are two parts to this study: individual dining and social dining. We counterbalance the order of these parts. For each part, we conduct three consecutive trials for each method. The methods and their relative order are blinded to the participant and counterbalanced across participants. The relative order of the methods stays consistent for both individual and social dining scenarios for each participant to reduce confusion and minimize cognitive load. The social dining scenario is conducted with a researcher. After testing each method, the participant answers the following questions:

- Would you consider the timing of the bites appropriate? (Yes/No)

- How was the bite timing of this method on average? (Very Early, Early, Slightly Early, Right Time, Slightly Late, Late, Very Late)
- L1: I was not distracted by this bite timing method.
- L2: I felt in control of the robot with this method.
- L3: The robot understands when I want a bite.
- L4: The robot moved at the appropriate moments.
- L5: This bite timing method is seamless.
- L6: This bite timing method required no mental workload.
- L7: This bite timing method required no physical workload.
- L8: For social dining only: The robot allowed me to have natural conversation.

All Likert items (L1–L8) are rated on a 7-point scale.

After each part of the two parts of the study, the participant blindly ranks the three methods based on their preference. At the end of the study, the methods are unblinded and the participant ranks the three methods overall. Then, they answer the following qualitative questions:

- Why did you prefer your highest ranked method? Please elaborate and compare the methods for both the individual and social settings.
- Which methods did you feel in control while using? Why or why not?
- Which method had the most appropriate bite timing? Elaborate on why it had the most appropriate timing and compare it to the other methods.
- What cues do you think a caregiver or you use to know when to feed a care recipient or yourself a bite?
- Which method was closest to how a caregiver or you would perform this task? Please elaborate on why.
- Were the glasses comfortable to wear while eating? If not, please explain.
- Was the throat microphone comfortable to wear while eating? If not, please explain.

7.4.5 Phase 3: Qualitative Evaluation with Individual with Motor Impairments

We conducted a final study, named Phase 3, with 1 individual with motor impairments. We received approval from Cornell University's Institutional Review Board. The participant was 78 years old, male, white, and has quadriplegia due to spinal cord injury (SCI). He answered 1 (No Experience) to the 5-point question, "How much experience do you have with interacting with robots?". The participant has no prior experience with any feeding robots. We use the FEAST system, which consists of a Kinova 7DOF arm, with the same safety features as described in Jenamani et al. [102]. The participant is only fed pieces of bananas as they are unable to eat harder foods.

For this study, we only evaluate our Wearables method. The FEAST robot automatically acquires the pieces of banana and moves to a staging point. From there, the robot receives stop and proceed commands based on assertiveness thresholding on the predictions from our bite timing model. We set a fixed outside-mouth bite transfer distance, and the FEAST robot autonomously perceives the participant's mouth location and moves to that position. We make a minor modification to the Wearables method by having the robot commit to completing the remainder of the trajectory if it is within 5 centimeters of the participant's mouth. This change was motivated by observations in our Phase 2 study, where participants would open their mouths in anticipation of the robot finishing the trajectory, inadvertently causing it to stop.

Similar to the Phase 2 study, we walk the participant through using the robot and the Wearables method. Due to limited time with the participant, we only test a few bites with the participant in an individual setting with a fixed mapped assertiveness threshold of 3 to minimize participant confusion. The day after the study, the participant answered similar qualitative questions to the Phase 2 Study.

Table 7.2: Bite Timing Model Performance using LOSO-CV with Various Inputs

	IMU & Throat Mic.	IMU only	Throat Mic. only	Naive Predictor (mean of labels)
Mean Absolute Error	**2.83 ± 0.33 s**	2.98 ± 0.25 s	3.16 ± 0.27 s	3.25 ± 0.15 s

Table 7.3: Accuracy and nMCC after Thresholding using Optimal Thresholds for each Participant and a Fixed Threshold

	IMU & Throat Mic.	IMU only	Throat Mic. only	Always Feed
Accuracy (Optimal Thresholds)	**0.700 ± 0.090**	0.680 ± 0.071	0.532 ± 0.137	0.3957 ± 0.1215
Accuracy (Fixed Threshold)	0.668 ± 0.066	0.633 ± 0.069	0.523 ± 0.121	0.3957 ± 0.1215
nMCC (Optimal Thresholds)	**0.674 ± 0.088**	0.655 ± 0.053	0.576 ± 0.065	0.500 ± 0.000
nMCC (Fixed Threshold)	0.650 ± 0.082	0.624 ± 0.065	0.554 ± 0.102	0.500 ± 0.000

7.5 Results and Discussion

7.5.1 Model Performance using LOSO-CV

As detailed in Section 7.4.2, our bite timing regression model outputs predictions of time until the next bite in seconds. Our model performance using mean absolute error (MAE) and LOSO-CV, is shown in Table 7.2. Our results show that the combined IMU and throat microphone model achieves the lowest MAE value of 2.83 s. The IMU only model performs slightly worse with an error of 2.98 s, while the throat microphone only model achieves 3.16 s MAE. In comparison, a naive predictor using the mean of the training labels for each fold achieves 3.25 s MAE. These metrics show that our model is able to learn a general relationship between 1 second wearable data windows consisting of motion, chewing, and talking signals and the time until the next bite. They also show the value of using the combination of sensor modalities over a single sensing modality.

As described in Section 7.4.3, in order to deploy our method and enable reactive robot behavior based on wearable signals, we need to threshold predictions from our bite-timing model. We include binary metrics, nMCC and accuracy, using both a fixed assertiveness threshold, $\tau = 6$ s, equivalent to a mapped assertiveness threshold of 3 and using the optimal threshold per user as specified in Section 7.4.3. As previously mentioned, bite timing is a challenging problem and is largely subconscious, with multiple timings and robot motions that participants would accept. These binary metrics should not be interpreted as an indication of whether or not the bite timing was appropriate and instead be used to provide a signal on how well the robot's start

and stop motions after thresholding align with the executed trajectories in the Phase 1 study.

As seen in Table 7.3, our IMU & Throat Mic. method achieves the best alignment performance of 0.70 accuracy and 0.674 nMCC using the optimal threshold per participant. This performance slightly outperforms the IMU only method which achieves 0.68 accuracy and 0.655 nMCC. We see significantly better alignment performance than an Always Feed method which achieves 0.3957 accuracy and 0.5 nMCC. We also find that our combined IMU & Throat Mic. algorithm outperforms both the IMU only and Throat Mic. only algorithms using these binary metrics, further backing our proposed approach of using the combination of the two modalities.

Our metrics additionally show the value of incorporating user-selected thresholding. We find that using the optimal threshold for each participant results in an increase in alignment performance in comparison to the fixed threshold. Specifically, we see an increase from 0.650 nMCC to 0.675 nMCC and from 0.668 accuracy to 0.700 accuracy for the IMU & Throat Mic. algorithm. In addition to these quantitative results, we find the user-selected thresholding to improve performance of the system during pilot testing. Further work is needed to validate how users tune the thresholds during deployment.

7.5.2 Bite Timing Appropriateness

We present results for bite timing appropriateness from our Phase 2 Study in Table 7.4 for the yes/no question "Would you consider the timing of the bites appropriate?". In Fig. 7.4, we present participants' responses to the question "How was the bite timing of this method on average? (Very Early, Early, Slightly Early, Right Time, Slightly Late, Late, Very Late)".

Table 7.4: Number and percentage of responses indicating bite timing was appropriate

Context	**Fixed Interval**	**Mouth Open**	**Wearables**
Individual	8 (53%)	15 (100%)	15 (100%)
Social	13 (87%)	11 (73%)	13 (87%)
Overall	21 (70%)	26 (87%)	**28 (93%)**

Individual Dining

As seen in Table 7.4, for individual dining, we find that the Wearables method performs significantly better than both baseline methods. All 15 participants rated the timing of the Wearables method as appropriate, comparable to the Mouth Open method, while only 8 out of 15 participants (53%) rated the Fixed Interval method as appropriate.

As shown in Fig. 7.4, the Wearables method received 13 Right Time responses and only 2 Slightly Early responses. In contrast, the next best method, Mouth Open, received only 9 Right Time responses. The Fixed Interval method received no Right Time responses and was consistently rated as Slightly Late, Late, or Very Late. These results highlight the clear timing advantages of the Wearables method in individual settings. By utilizing wearable signals, the system is able to dynamically initiate feeding when chewing subsides and to pause when chewing is detected, potentially enabling bite timing that aligns more closely with the user's readiness for the next bite. In contrast, the Fixed Interval method performs worst, feeding every 45 seconds regardless of user behavior. Even if the interval were shortened, it would still lack responsiveness to varying chewing durations across different foods. Interestingly, the Mouth Open method, despite being manually triggered, does not outperform Wearables. This is likely due to the delay between a participant opening their mouth and the robot completing its trajectory to deliver the bite.

Qualitative responses from multiple participants also strongly support the quantitative results, with P2 saying "In terms of the individual portion, I liked this [Wearables] timing the best because it didn't force me to wait for another bite like the other ones, it kind of naturally matches my pace, the other ones felt a little slow or fast at times." Our testing in Phase 3 with 1 participant with motor impairments further supports the data from Phase 2. When asked the qualitative question, "Did this method have appropriate bite timing for this setting?", the participant answered that the bite timing was "Appropriate". Additionally when the participant was asked "Did you like this bite timing method for this setting?", they answered "[I] Like this method; It was convenient."

Social Dining

As seen in Table 7.4, from our Phase 2 data for social dining, we find that the Wearables and Fixed Interval methods perform similarly in terms of appropriateness with 13 out of 15 (87%) of participants finding the methods appropriate. Mouth Open performed the worst for social dining with only 11 out of 15 (73%) participants finding the method appropriate.

As shown in Fig. 7.4, responses exhibit greater variability across all methods compared to individual dining, reflecting the increased inconsistency in social dining settings due to conversation. Both the Wearables and Fixed Interval methods received 8 *Right Time* responses. However, the remaining Wearables responses were more tightly clustered around *Right Time*, with only 2 rated as early and 1 as late, whereas Fixed Interval received 1 early and 3 late ratings.

Qualitative responses also reflect mixed preferences between the Fixed Interval and Wearables methods. Some participants found the Fixed Interval method disruptive; for example, P15 noted that "fixed interval seemed to make the robot come at random points during the conversation, which also disrupted conversation a bit." Others, like P3, appreciated its predictability, stating it was "easier to anticipate." Participant reactions to the Wearables method were similarly varied. Some appreciated its adaptive nature, such as P2, who remarked, "it would wait for me if I started chewing or talking after I signaled that I was ready to eat. I didn't have to control the conversation for a pause around the time that the bite would be ready." In contrast, P8 found the timing intrusive, saying, "Because I speak more slowly and take time to think between ideas, I felt the robot was smothering me with food when I was hoping to talk." These varied responses underscore the value of the user-controlled assertiveness threshold, which we set to a middle level (3) during the study to minimize participant confusion.

Most participants found the Mouth Open method to have the least appropriate timing, with only 5 *Right Time* responses. Similar to the individual dining setting, this is likely due to the delay between when a participant opens their mouth and when the robot completes its trajectory to deliver the bite, a lag that becomes even more disruptive in social dining settings.

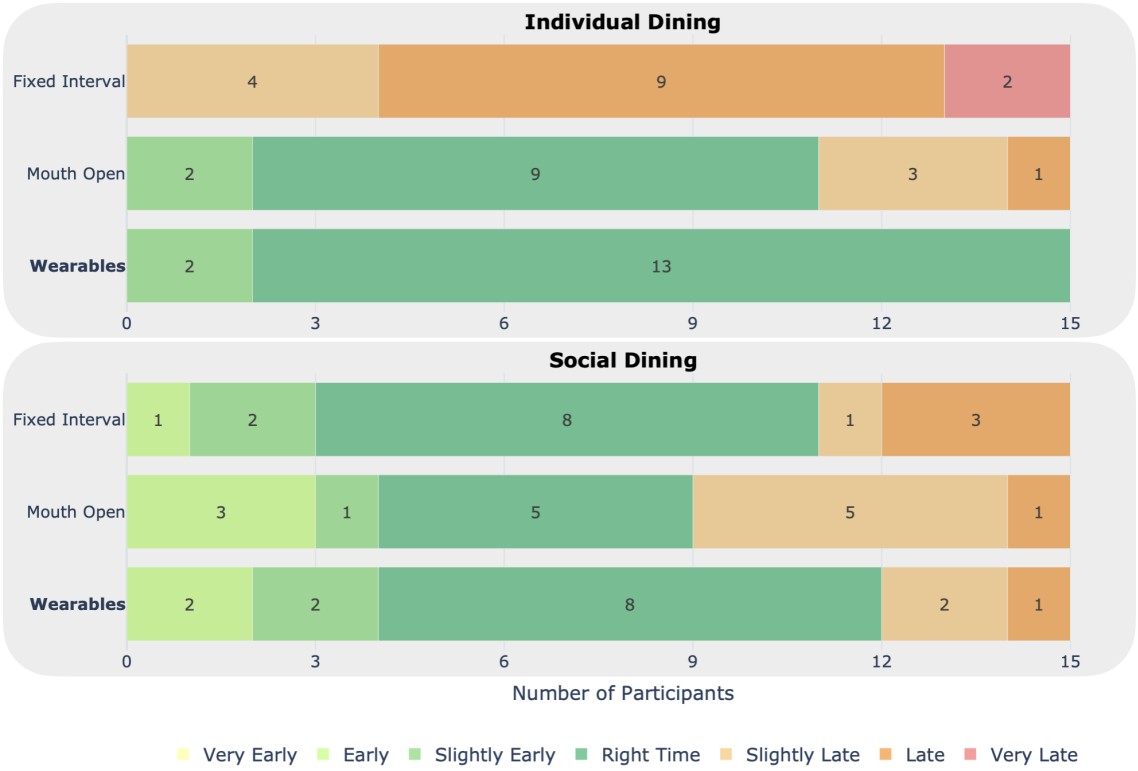


Figure 7.4: Phase 2 participant responses to the question, "How was the bite timing of this method on average? (Very Early, Early, Slightly Early, Right Time, Slightly Late, Late, Very Late)" for both individual and social dining scenarios.

7.5.3 Likert Items

For the Likert items described in Section 7.4.4, we present the mean and standard deviation for each item for individual and social dining in Fig. 7.5 and Fig. 7.6 respectively. We additionally present results from statistical testing in the plot. We specifically use the non-parametric Friedman test to assess overall differences among the three methods for each item. Significant differences ($p < .05$) are found for all items except *L4: Appropriate Movement* in the social dining context. For items showing significant overall effects, we conduct post hoc Wilcoxon signed-rank tests for all pairwise method comparisons, applying the more conservative Bonferroni correction factor of 3.

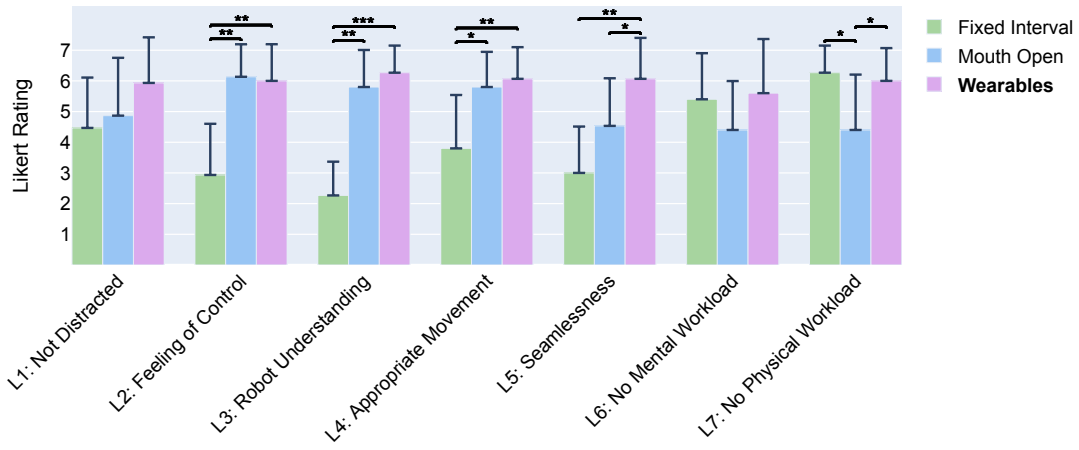


Figure 7.5: Phase 2 Individual Dining Mean Likert Ratings with error bars representing the standard deviation. Higher scores indicate more favorable outcomes for all items. Asterisks denote statistically significant pairwise differences based on Wilcoxon signed-rank tests with Bonferroni correction ($p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$). Prior to pairwise comparisons, a non-parametric Friedman test was used to confirm overall differences among the three methods for each Likert item.

Individual Dining

As seen in Fig. 7.5, the Wearables method received the highest mean Likert ratings across multiple items in the individual dining condition, with statistically significant differences observed for *L2: Feeling of Control*, *L3: Robot Understanding*, *L4: Appropriate Movement*, *L5: Seamlessness*, and *L7: No Physical Workload* compared to at least one of the baseline methods.

The Wearables method achieved similar levels of *Feeling of Control* as the Mouth Open method, but with statistically significantly lower *Physical Workload*. This finding is particularly insightful, as prior studies have noted a common trade-off in assistive systems where increasing autonomy often leads to users feeling less in control [27, 98]. In contrast, our findings suggest that our Wearables approach offers a promising balance, enabling the system to act proactively and reduce workload through autonomy while still respecting user intent. As P6 described, the system "made me feel in control in the sense that it understood my natural body movements

and I did not have to exert any additional effort." P15 similarly emphasized the reduced effort and efficiency of the system, stating that "it required the least amount of work on my part and had the best timing. I didn't have to open my mouth or wait unnecessarily long for the food to come." Our testing in Phase 3 with 1 participant with motor impairments further supports the data from Phase 2. When asked the qualitative question, "Did you feel in control while using the robot with this bite timing method for this setting?", the participant answered "Feel in control; Robot follows commands".

Interestingly, the Wearables method also matched Mouth Open on ratings of *L3: Robot Understanding* and *L4: Appropriate Movement*, indicating that participants perceived the robot as equally responsive and well-timed in its actions. This is notable because Mouth Open is an explicit cue, whereas the Wearables method relies on subtle, passive signals like chewing and motion. The strong Wearables performance on these items highlights that it is capable of interpreting nuanced user behavior without requiring overt signaling and extra workload from users.

Lastly, Wearables was rated statistically significantly seamless than both baseline methods based on responses to *L5: Seamlessness*, suggesting that it offered a smoother and more intuitive experience. This supports the idea that leveraging natural user behaviors can reduce friction in human-robot interaction, enhancing user comfort without sacrificing control.

Social Dining

In the social dining setting, as seen in Fig. 7.6, the Wearables method received the highest ratings for *L1: Not Distracted* and *L8: Natural Conversation*, significantly outperforming the Mouth Open method on both. These results suggest that the method allows the robot to feed without interrupting conversation, in contrast to requiring explicit cues like mouth opening which can feel disruptive in social settings. P3 noted, "I think the wearable method was especially beneficial for the social setting, since the actions for pausing the feeding aligned well [with] social cues where I'd rather not be disrupted." Similarly, P15 remarked, "in the social setting I could continue having a conversation and the robot would stay still and wait. Thus, it wouldn't make the conversation awkward." In contrast, the Fixed Interval method

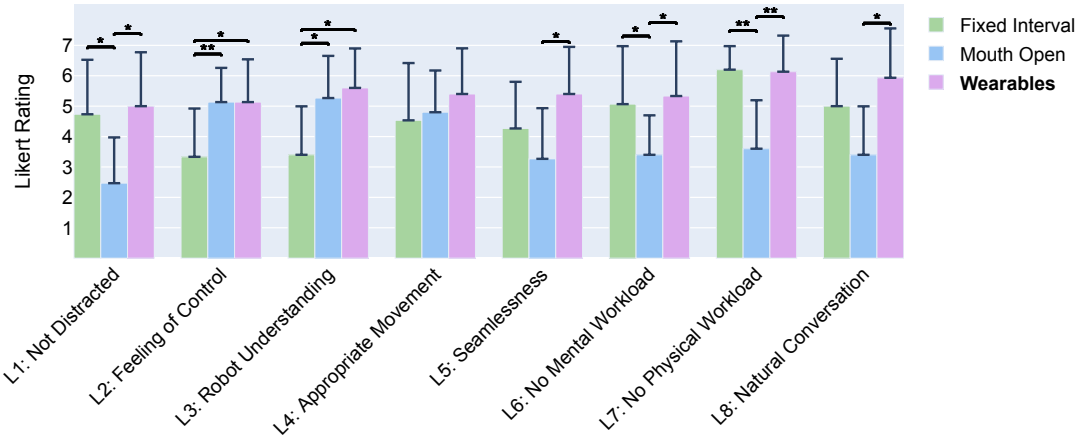


Figure 7.6: Phase 2 Social Dining Mean Likert Ratings with error bars representing the standard deviation. Higher scores indicate more favorable outcomes for all items. Asterisks denote statistically significant pairwise differences based on Wilcoxon signed-rank tests with Bonferroni correction ($p < 0.05 = *$, $p < 0.01 = **$). Prior to pairwise comparisons, a non-parametric Friedman test was used to confirm overall differences among the three methods for each Likert item.

did not achieve a significant difference for *L8: Natural Conversation* compared to Mouth Open.

Wearables also matched Mouth Open on *L2: Feeling of Control* and *L3: Robot Understanding*, while significantly outperforming Fixed Interval on both. This indicates that even in more dynamic, social contexts, the wearable sensing approach remains responsive and more aligned to user intent than the baselines.

Finally, Wearables was rated as more seamless based on responses to *L5: Seamlessness* and was additionally associated with significantly lower *L6: Mental Workload* and *L7: Physical Workload* compared to Mouth Open. Notably, Wearables matched Fixed Interval, a fully autonomous method, on workload while offering more adaptive and responsive behavior. P1 captured this sentiment succinctly: "It [Wearables] was the most seamless and recreated the way I usually eat."

7.5.4 Preference Ranking

As detailed in Section 7.4.4, participants were asked to blindly rank their preference for the three methods after completing both the individual and social dining scenarios. As shown in Fig. 7.7, participants preferred the Wearables method over the baseline methods in both dining contexts and overall.

In the individual dining setting, the Wearables method was strongly blindly favored, receiving 13 out of 15 first-place votes (87%). The Mouth Open method followed, with 2 out of 15 first-place votes (13%) and 12 out of 15 second-place votes (80%). The Fixed Interval method was least preferred, receiving the majority of third-place votes.

In the social dining setting, the Wearables method remained the top choice, with 10 out of 15 first-place votes (67%). Unlike the individual setting, the Fixed Interval method was the next most preferred, receiving 4 out of 15 first-place votes (27%) and 10 out of 15 second-place votes (67%). The Mouth Open method was the least preferred in this context as it interrupted the flow of conversation the most.

Lastly, in the overall ranking at the end of the study, the participants again strongly preferred the Wearables method, which received 13 of 15 first-place votes (87%). We find that Fixed Interval is the second most preferred method with Mouth Open coming in last.

7.5.5 Similarity to Caregiver

At the end of the Phase 2 and 3 studies, we additionally gauged participants on what cues a caregiver or they themselves would use to decide when to deliver a bite of food, and which bite timing method best reflected that behavior. When describing the cues they would rely on, 8 of the Phase 2 participants mentioned chewing or stopping chewing and 7 participants mentioned explicit gestures such as opening the mouth, head turning, or nodding. Other cues included talking or stopping talking (6 participants), swallowing (3 participants), looking at the food or caregiver (2 participants), and being still (1 participant). When asked which method most closely resembled how a caregiver would decide when to feed, a majority (9 out of 15) selected the Wearables method. In comparison, 3 chose Mouth Open, 1 chose Fixed Interval, and 2 participants mentioned both Wearables and Mouth Open. P5 summarized why

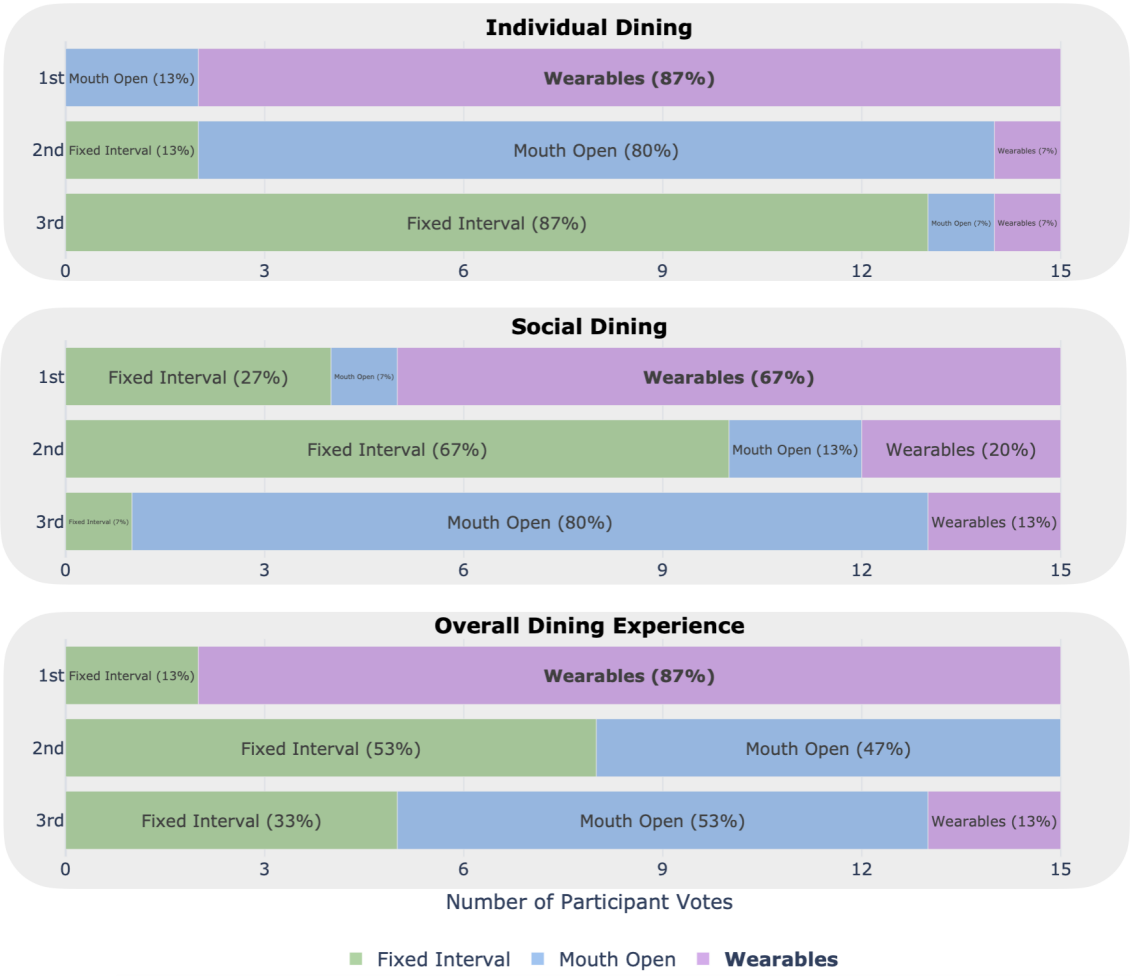


Figure 7.7: Phase 2 Preference rankings for the three methods. The Wearables method received the most 1st-place rankings across individual, social, and overall contexts. Rankings for individual and social settings were collected blindly, while the overall ranking was unblinded.

the Wearables method was the closest to a caregiver saying, "it wouldn't give the bite when I clearly didn't want it, and it would adjust to the environment to decide when to give the next bite, like a person would."

These results suggest that the Wearables method most consistently aligned with caregiver-like behavior across participants. This was further supported by the participant with motor impairments in the Phase 3 study. In terms of cues that a caregiver uses, the participant with impairments specified "Voice, eye contact". When asked

"Was this method close to how a caregiver would perform this task for this setting?", the participant answered "Close, no difference between caregiver and robot".

7.5.6 Adoption of the Wearable Interface and Future Work

Our findings support the hypothesis that a robot can achieve appropriate bite timing by being highly reactive to natural user cues, including chewing, speaking, and head movements, from wearable sensors. Furthermore, the strong performance across many participants, types of feeding robots, robot configurations, feeding trajectories, foods, and individual and social dining scenarios suggests that our system is generalizable across key variables in robot-assisted feeding. These results in addition to strong participant acceptance, performance on various Likert metrics, and ranking of our WAFFLE system over current baselines for bite timing, indicate strong likelihood of adoption of our interface by users in future commercial feeding robotic systems.

Given this promising outlook, it is worth noting that wearable sensing with IMUs and throat microphones is increasingly practical for real-world deployment. IMUs are already integrated into commercially available smartglasses and throat microphones currently have utility as general-purpose microphones. In terms of robot-assisted feeding specifically, these wearables could also offer multiple use cases beyond bite timing, such as tracking food intake or sensing spasms. We also note that real-world deployment of our system faces few barriers for adoption. Specifically, the wearables are privacy preserving and the throat microphone does not pick up ambient speech or sounds, allowing for reliable operation in noisy environments. The wearables are unobtrusive, further supporting adoption in daily life. These insights are in line with past work that has shown that wearable interfaces offer key benefits over conventional interfaces for caregiving robots, enabling intuitive, accessible, and unobtrusive control [167, 168, 171]. We find comparable advantages in our study, with the proposed Wearables method outperforming the baseline approaches that are more disruptive because they rely on screens, assistive input devices, or require the robot to approach from the front or the user to face the robot each time they want a bite. In contrast, WAFFLE reduces disruption and supports natural interaction in both individual and social dining situations.

To further assess potential adoption, we qualitatively evaluated the comfort of the

two wearable devices, as wearability plays a crucial role in sustained use. We find that the vast majority of participants considered both the IMU-equipped glasses and the throat microphone comfortable to wear. Six out of fifteen participants initially found the throat microphone uncomfortable or distracting to wear, but all clarified, without prompting, that they stopped noticing it once the study began. One participant, however, mentioned that they would not want to wear it for an extended period. For users who may find the throat microphone undesirable, we believe IMU-only models could provide a reasonable alternative based on our bite timing results in Section 7.5.1, though further testing is needed to confirm this. The single participant with motor impairments in the Phase 3 study found both devices comfortable.

7.6 Conclusion

In this work, we presented WAFFLE, a reactive, wearable-based system for estimating bite timing in robot-assisted feeding. By leveraging natural wearable signals, WAFFLE enables intuitive and socially appropriate interactions, responding in real time to passive user cues like chewing, speaking, and head movements. Our results from studies with both participants with and without motor impairments demonstrate that WAFFLE matches or exceeds baseline methods across key metrics such as user feeling of control, robot understanding, and workload, and is the preferred method for the vast majority of participants. These findings support the promise of reactive wearable sensing for bite timing as an unobtrusive and generalizable solution for real-world robotic feeding across users, robots, and settings. Future work should include additional testing with individuals with impairments, particularly through long-term deployment in the home and in more diverse dining environments. Another promising direction for improvement of the system is to incorporate additional cues that participants identified as relevant for bite timing, such as eye gaze.

Chapter 8

Conclusion

8.1 Summary

This dissertation addresses the pressing need for technological solutions that move beyond reactive care and support the long-term, personalized management of chronic conditions. By integrating multimodal sensing into compact, user-centered systems, the work presented here demonstrates how hardware, algorithms, and interface design can collectively bridge gaps in both clinical and home-based care.

In Part I, we introduced novel approaches for monitoring dermatological conditions, particularly chronic itch and changes in skin morphology. Our contributions include a wearable ring for real-time scratch detection and intensity estimation, as well as the first handheld, validated system for in-vivo 3D wrinkle height reconstruction across various body locations. These tools enable new forms of quantitative assessment, offering continuous, interpretable metrics that can enhance diagnosis, treatment evaluation, and patient self-management.

In Part II, we expanded the concept of multimodal sensing into the realm of assistive robotics, developing and evaluating novel interfaces for individuals with motor impairments. Through active, shared, and passive control paradigms, we demonstrated that wearable and speech-based interfaces can empower users to perform complex self-care tasks with robotic assistance. Our work culminated in week-long home deployments and testing, validating performance and usability of these systems in the real world.

Together, these contributions advance the scientific understanding of how multimodal sensing and learning-based algorithms can support autonomy, well-being, and quality of life for individuals with chronic health needs. They lay the groundwork for future systems that are adaptive, scalable, and grounded in the lived experiences of their users.

8.2 Future Work

There are many promising avenues for future work, building on the foundations laid out in this thesis. Future work in dermatological sensing should focus on improving the sensing hardware and validating it in long-term, real-world deployments with patients managing chronic conditions like eczema or psoriasis. For the GelSight-based skin surface reconstruction system, increasing the elasticity of the gel layer would enable better conformity to skin surfaces while minimizing distortion of the skin. The system should be explored for monitoring treatment response over time and could potentially be extended to applications like moisture estimation or hair density assessment. For the scratch tracking wearable, future work should prioritize developing resource efficient algorithms and adaptive sampling strategies to minimize compute, memory, and power consumption, enabling extended wear. Future device development should also enable uploading this data to patient records in the cloud, where it could be aggregated, analyzed, and visualized in a mobile application. This information could be synced with other contextual data such as sleep patterns, environmental conditions, and patient-reported information like diet to help individuals identify correlations between potential itch triggers and itch symptoms. Through longer term testing, it would be especially valuable to assess whether patients feel more empowered in managing their condition and whether clinicians adjust and find better treatment plans based on trends over time. Moreover, the contributed technologies could be extended to veterinary contexts, such as monitoring atopic dermatitis in dogs or cats, where objective symptom tracking is equally lacking and could improve quality of care and determination of treatment efficacy.

For wearable assistive interfaces, there are two avenues of future work. The first is on active and shared control interfaces like HAT where future work should primarily involve more testing with individuals with motor impairments in home

settings and for additional caregiving tasks. This insight will help identify real-world usability challenges and reveal patterns in user preferences and adaptation over time, allowing further improvement of the interfaces. The second avenue is expanding on wearable interfaces that enable passive control of caregiving robots. Our results in Chapter 7, showed huge potential for reduction of physical and cognitive effort while enhancing users' sense of control and agency using passive sensing. Signals from multiple wearable devices including smartglasses, rings, and watch form factors consisting of various sensing modalities including IMUs, contact microphones, and cameras could enable more appropriate robot behavior without explicit commands in additional scenarios and settings. Future work should focus on using passive sensing for other activities of daily living such as robot-assisted dressing. As these systems become more mature, they should also be tested through long-term studies in the home with individuals with motor impairments.

More broadly, future research can explore how these sensing technologies generalize to other contexts of chronic care, how they integrate into healthcare workflows, and how they can be scaled across diverse populations.

Bibliography

[1] Obi feeding robot. URL https://meetobi.com/. 6.2, 6.3.2, 7.2, 7.1, 7.4.1, 7.4.4

[2] Oura ring. URL https://ouraring.com/. 2.5

[3] Sensel morph. URL https://morph.sensel.com/. 2.3.2

[4] Atlas.ti | the #1 software for qualitative data analysis. URL https://atlasti.com/. 6.7.2

[5] Openai api - gpt-3.5 turbo. URL https://platform.openai.com/docs/models/gpt-3-5-turbo. 6.2

[6] Openai api - whisper. URL https://platform.openai.com/docs/models/whisper. 6.4.2

[7] Alexander C Abad and Anuradha Ranasinghe. Visuotactile sensors with emphasis on gelsight sensor: A review. *IEEE Sensors Journal*, 20(14):7628–7638, 2020. 3.2

[8] Evan Ackerman. This robot could be the key to helping people with disabilities, Oct 2023. URL https://spectrum.ieee.org/stretch-assistive-robot. 1.3, 5.4

[9] Arpit Agarwal, Abhiroop Ajith, Chengtao Wen, Veniamin Stryzheus, Brian Miller, Matthew Chen, Micah K Johnson, Jose Luis Susa Rincon, Justinian Rosca, and Wenzhen Yuan. Robotic defect inspection with visual and tactile perception for large-scale components. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 10110–10116. IEEE, 2023. 3.2

[10] Arpit Agarwal, Achu Wilson, Timothy Man, Edward Adelson, Ioannis Gkioulekas, and Wenzhen Yuan. Vision-based tactile sensor design using physically based rendering. *Communications Engineering*, 4(1):21, 2025. 3.3.6

[11] Michael Ahn, Debidatta Dwibedi, Chelsea Finn, Montse Gonzalez Arenas, Keerthana Gopalakrishnan, Karol Hausman, Brian Ichter, Alex Irpan, Nikhil Joshi, Ryan Julian, et al. Autort: Embodied foundation models for large scale

orchestration of robotic agents. *arXiv preprint arXiv:2401.12963*, 2024. 6.2, 6.3.1

[12] Marco Altini and Hannu Kinnunen. The promise of sleep: A multi-sensor approach for accurate sleep stage detection using the oura ring. *Sensors*, 21 (13):4302, 2021. 2.5

[13] Leire Ambrosio, Juana Maria Senosiain García, Mario Riverol Fernández, Sagrario Anaut Bravo, Sara Díaz De Cerio Ayesa, María Eugenia Ursúa Sesma, Neus Caparrós, and Mari Carmen Portillo. Living with chronic illness in adults: a concept analysis. *Journal of clinical nursing*, 24(17-18):2357–2367, 2015. 1

[14] Anthony Wertz. Sml posey motion sensors github organization. https://github.com/SML-Posey. Accessed: 2025-05-08. 7.4.1

[15] David F. Apple, Carol A. Anson, John D. Hunter, and R. Bryan Bell. Spinal cord injury in youth. *Clinical Pediatrics*, 34(2):90–95, 1995. doi: 10.1177/000992289503400205. 4.2

[16] Montserrat Gonzalez Arenas, Ted Xiao, Sumeet Singh, Vidhi Jain, Allen Z Ren, Quan Vuong, Jake Varley, Alexander Herzog, Isabel Leal, Sean Kirmani, et al. How to prompt your robot: A promptbook for manipulation skills with code as policies. In *Towards Generalist Robots: Learning Paradigms for Scalable Skill Acquisition@ CoRL2023*, 2023. 6.3.1, 6.4.1

[17] Brenna D. Argall. Autonomy in rehabilitation robotics: An intersection. *Annual Review of Control, Robotics, and Autonomous Systems*, 1(1):441–463, 2018. doi: 10.1146/annurev-control-061417-041727. URL https://doi.org/10.1146/annurev-control-061417-041727. 5.3.3

[18] Brian S Armour, Elizabeth A Courtney-Long, Michael H Fox, Heidi Fredine, and Anthony Cahill. Prevalence and causes of paralysis—united states, 2013. *American journal of public health*, 106(10):1855–1857, 2016. 1.1, 4.2

[19] Indranil Bardhan, Hsinchun Chen, and Elena Karahanna. Connecting systems, data, and people: A multidisciplinary research roadmap for chronic disease management. *MIS Quarterly*, 44(1), 2020. 1

[20] Cristian Barrué, Alejandro Suárez, Marco Inzitari, Aida Ribera, and Guillem Alenyà. Nyam: The role of configurable engagement strategies in robotic-assisted feeding. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 228–232, 2024. 7.3.1

[21] Brian Lee Beatty, Shani Kahan, Burcak Bas, Bettina Zou, and Nicole Werpachowski. Contrasting results of surface metrology techniques for three-dimensional human fingerprints. *arXiv preprint arXiv:2410.12648*, 2024. 3.2

[22] Abdelkareem Bedri, Diana Li, Rushil Khurana, Kunal Bhuwalka, and Mayank

Goel. Fitbyte: Automatic diet monitoring in unconstrained situations using multimodal sensing on eyeglasses. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2020. 7.3.2, 7.4.2

[23] Suneel Belkhale, Ethan K Gordon, Yuxiao Chen, Siddhartha Srinivasa, Tapomayukh Bhattacharjee, and Dorsa Sadigh. Balancing efficiency and comfort in robot-assisted bite transfer. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 4757–4763. IEEE, 2022. 6.3.2, 7.3.1

[24] Margrit Betke, James Gips, and Peter Fleming. The camera mouse: visual tracking of body features to provide computer access for people with severe disabilities. *IEEE Transactions on neural systems and Rehabilitation Engineering*, 10(1):1–10, 2002. 4.3, 4.5, 5.3.1

[25] Tapomayukh Bhattacharjee, Maria E Cabrera, Anat Caspi, Maya Cakmak, and Siddhartha S Srinivasa. A community-centered design framework for robot-assisted feeding systems. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pages 482–494, 2019. 6.3.2, 7.2, 7.3.1

[26] Tapomayukh Bhattacharjee, Gilwoo Lee, Hanjun Song, and Siddhartha S Srinivasa. Towards robotic feeding: Role of haptics in fork-based food manipulation. *IEEE Robotics and Automation Letters*, 4(2):1485–1492, 2019. 7.3.1

[27] Tapomayukh Bhattacharjee, Ethan K. Gordon, Rosario Scalise, Maria E. Cabrera, Anat Caspi, Maya Cakmak, and Siddhartha S. Srinivasa. Is More Autonomy Always Better?: Exploring Preferences of Users with Mobility Impairments in Robot-assisted Feeding. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 181–190, Cambridge United Kingdom, March 2020. ACM. ISBN 978-1-4503-6746-2. doi: 10.1145/3319502.3374818. URL https://dl.acm.org/doi/10.1145/3319502.3374818. 5.3.3, 5.6.2, 7.5.3

[28] Tapomayukh Bhattacharjee, Ethan K Gordon, Rosario Scalise, Maria E Cabrera, Anat Caspi, Maya Cakmak, and Siddhartha S Srinivasa. Is more autonomy always better? exploring preferences of users with mobility impairments in robot-assisted feeding. In *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*, pages 181–190, 2020. 6.3.2, 6.6.1, 7.3.1, 7.4.4

[29] Monica CT Bloemen, Maaike S van Gerven, Martijn BA van der Wal, Pauline DHM Verhaegen, and Esther Middelkoop. An objective device for measuring surface roughness of skin and scars. *Journal of the American Academy of Dermatology*, 64(4):706–715, 2011. 3.2, 3.5

[30] Anthony Brohan, Noah Brown, Justice Carbajal, Yevgen Chebotar, Xi Chen,

Krzysztof Choromanski, Tianli Ding, Danny Driess, Avinava Dubey, Chelsea Finn, et al. Rt-2: Vision-language-action models transfer web knowledge to robotic control. *arXiv preprint arXiv:2307.15818*, 2023. 6.2, 6.3.1

[31] He-Thong Bui, Le-Van Nguyen, Thanh-Nghi Ngo, Tuan-Sinh V Nguyen, Anh-Ngoc T Ho, and Qui-Tra Phan. Improved electric wheelchair controlled by head motion. In *Research in Intelligent and Computing in Engineering*, pages 121–129. Springer, 2021. 4.3, 5.3.2, 5.4.5

[32] Lina Bunketorp Käll, Johanna Wangdell, and Carina Reinholdt. Surgical restoration of hand function in tetraplegia. *Spinal Cord Series and Cases*, 7(1), 2021. doi: 10.1038/s41394-021-00387-5. 1.1, 4.2

[33] Maria E Cabrera, Tapomayukh Bhattacharjee, Kavi Dey, and Maya Cakmak. An exploration of accessible remote tele-operation for assistive mobile manipulators in the home. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, pages 1202–1209. IEEE, 2021. 4.3, 4.5, 5.3.1, 6.2

[34] Danny Camfferman, John D Kennedy, Michael Gold, Alfred J Martin, and Kurt Lushington. Eczema and sleep and its relationship to daytime functioning in children. *Sleep medicine reviews*, 14(6):359–369, 2010. 1.1, 2.2

[35] Gustavo Carlos da Silva, Miguel B Barbosa, Flávio B Camargo Júnior, Patrícia L Moreira, Rejane Werka, and Airton A Martin. Detection of skin wrinkles and quantification of roughness using a novel image processing technique from a dermatoscope device. *Skin Research and Technology*, 29(6):e13335, 2023. 3.2, 3.5

[36] Sarah L Chamlin, Christine L Mattson, Ilona J Frieden, Mary L Williams, Anthony J Mancini, David Cella, and Mary-Margaret Chren. The price of pruritus: sleep disturbance and cosleeping in atopic dermatitis. *Archives of pediatrics & adolescent medicine*, 159(8):745–750, 2005. 1.1, 2.2

[37] Tiffany L Chen, Matei Ciocarlie, Steve Cousins, Phillip M Grice, Kelsey Hawkins, Kaijen Hsiao, Charles C Kemp, Chih-Hung King, Daniel A Lazewatsky, Adam E Leeper, et al. Robots for humanity: using assistive robotics to empower people with disabilities. *IEEE Robotics & Automation Magazine*, 20(1):30–39, 2013. 1.1, 1.3, 4.2, 4.3, 5.2, 5.3.1, 5.4, 5.5, 6.2, 6.7.3

[38] Y-H Choi, D Kim, E Hwang, and BJ Kim. Skin texture aging trend analysis using dermoscopy images. *Skin Research and Technology*, 20(4):486–497, 2014. 3.2

[39] Keum San Chun, Youn J Kang, Jong Yoon Lee, Morgan Nguyen, Brad Lee, Rachel Lee, Han Heul Jo, Emily Allen, Hope Chen, Jungwoo Kim, et al. A skin-conformable wireless sensor to objectively quantify symptoms of pruritus.

Science advances, 7(18):eabf9405, 2021. 2.2

[40] Matei Ciocarlie, Kaijen Hsiao, Adam Leeper, and David Gossow. Mobile manipulation through an assistive home robot. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5313–5320. IEEE, 2012. 1.1, 1.3, 4.2, 4.3, 5.3.1, 5.4, 5.5

[41] Jennifer L Collinger, Brian Wodlinger, John E Downey, Wei Wang, Elizabeth C Tyler-Kabara, Douglas J Weber, Angus JC McMorland, Meel Velliste, Michael L Boninger, Andrew B Schwartz, and et al. High-performance neuroprosthetic control by an individual with tetraplegia. *The Lancet*, 381(9866):557–564, 2013. doi: 10.1016/s0140-6736(12)61816-9. 1.1, 4.2

[42] Thomas H Cook, Trelia J Craft, Rocco L Brunelle, Frank Norris, and William A Griffin. Quantification of the skin's topography by skin profilometry. *International Journal of Cosmetic Science*, 4(5):195–205, 1982. 3.5

[43] Wenceslao Shaw Cortez, Denny Oetomo, Chris Manzie, and Peter Choong. Control barrier functions for mechanical systems: Theory and application to robotic grasping. *IEEE Transactions on Control Systems Technology*, 29(2): 530–545, 2019. 6.4.1

[44] Courage+Khazaka electronic GmbH. Visioscan® vc 20plus – skin topography measured with the multi-talent. https://www.courage-khazaka.com/en/scientific-products/visioscan-vc-20plus, September 2024. Last updated 25 September 2024. 3.5

[45] Courage+Khazaka electronic GmbH. Visioline® vl 650 (quantirides®) – assessment of deep facial lines. https://www.courage-khazaka.com/en/scientific-products/visioline-vl-650, 2025. Accessed: 2025-06-08. 3.5

[46] Neil H Cox. Palpation of the skin–an important issue, 2006. 3.2

[47] Neil H Cox. A literally blinded trial of palpation in dermatologic diagnosis. *Journal of the American Academy of Dermatology*, 56(6):949–951, 2007. 3.2

[48] Kathrin Cresswell, Sarah Cunningham-Burley, and Aziz Sheikh. Health care robotics: qualitative exploration of key challenges and future directions. *Journal of medical Internet research*, 20(7):e10410, 2018. 6.7.2

[49] Gabriela O Cula, Paulo R Bargo, A Nkengne, and N Kollias. Assessing facial wrinkles: automatic detection and quantification. *Skin Research and Technology*, 19(1):e243–e251, 2013. 3.2, 3.5

[50] Nisha S Desai, Gabriele B Poindexter, Yvette Miller Monthrope, Sandra E Bendeck, Robert A Swerlick, and Suephy C Chen. A pilot quality-of-life instrument for pruritus. *Journal of the American Academy of Dermatology*, 59 (2):234–244, 2008. 2.2

[51] Michael J. DeVivo, Bette K. Go, and Amie B. Jackson. Overview of the national spinal cord injury statistical center database. *The Journal of Spinal Cord Medicine*, 25(4):335–338, 2002. doi: 10.1080/10790268.2002.11753637. 4.2

[52] Julia Di, Zdravko Dugonjic, Will Fu, Tingfan Wu, Romeo Mercado, Kevin Sawyer, Victoria Rose Most, Gregg Kammerer, Stefanie Speidel, Richard E Fan, et al. Using fiber optic bundles to miniaturize vision-based tactile sensors. *arXiv preprint arXiv:2403.05500*, 2024. 3.2

[53] Aline Darc Piculo Dos Santos, Ana Lya Moya Ferrari, Fausto Orsi Medola, and Frode Eika Sandnes. Aesthetics and the perceived stigma of assistive technology for visual impairment. *Disability and Rehabilitation: Assistive Technology*, 17 (2):152–158, 2022. 2.5

[54] Anca D Dragan and Siddhartha S Srinivasa. A policy-blending formalism for shared control. *The International Journal of Robotics Research*, 32(7):790–805, June 2013. ISSN 0278-3649, 1741-3176. doi: 10.1177/0278364913490324. URL http://journals.sagepub.com/doi/10.1177/0278364913490324. 5.4.6

[55] Pan Du, Warren A Kibbe, and Simon M Lin. Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *bioinformatics*, 22(17):2059–2065, 2006. 2.3.2, 3.3.7, 3.3.8

[56] A Dzwigałowska, A Sołyga-Żurek, RM Dębowska, and I Eris. Preliminary study in the evaluation of anti-aging cosmetic treatment using two complementary methods for assessing skin surface. *Skin Research and Technology*, 19(2):155–161, 2013. 3.2

[57] Peter F Edemekong, Deb Bomgaars, Sukesh Sukumaran, and Shoshana B Levy. Activities of daily living. 2019. 1, 7.2

[58] S Elman, LS Hynan, V Gabriel, and MJ Mayo. The 5-d itch scale: a new measure of pruritus. *British Journal of Dermatology*, 162(3):587–593, 2010. 2.2

[59] Ahmetcan Erdogan and Brenna D. Argall. The effect of robotic wheelchair control paradigm and interface on user performance, effort and preference: An experimental assessment. *Robotics and Autonomous Systems*, 94:282–297, 2017. ISSN 0921-8890. doi: https://doi.org/10.1016/j.robot.2017.04.013. URL https://www.sciencedirect.com/science/article/pii/S0921889016303670. 5.3.3

[60] Stephen Erickson and Brian S Kim. Research techniques made simple: itch measurement in clinical trials. *Journal of Investigative Dermatology*, 139(2): 264–269, 2019. 2.2

[61] Enzo Errichetti and Giuseppe Stinco. Dermoscopy in general dermatology: a practical overview. *Dermatology and therapy*, 6:471–507, 2016. 3.2

[62] Bin Fang, Fuchun Sun, Huaping Liu, and Chunfang Liu. 3d human gesture

capturing and recognition by the immu-based data glove. *Neurocomputing*, 277: 198–207, 2018. 4.3, 5.3.2

[63] Linda Fehr, W Edwin Langbein, and Steven B Skaar. Adequacy of power wheelchair control interfaces for persons with severe disabilities: A clinical survey. *Journal of rehabilitation research and development*, 37(3):353–360, 2000. 4.3, 5.3.1

[64] Ryan Feng, Youngsun Kim, Gilwoo Lee, Ethan K Gordon, Matt Schmittle, Shivaum Kumar, Tapomayukh Bhattacharjee, and Siddhartha S Srinivasa. Robot-assisted feeding: Generalizing skewering strategies across food items on a plate. In *The International Symposium of Robotics Research*, pages 427–442. Springer, 2019. 7.3.1

[65] Bo Fernhall, Kevin S Heffernan, Sae Young Jae, and Bradley N. Hedrick. Health implications of physical activity in individuals with spinal cord injury: a literature review. *Journal of health and human services administration*, 30 4: 468–502, 2008. 4.2

[66] Andre Ferreira, Wanderley C Celeste, Fernando A Cheein, Teodiano F Bastos-Filho, Mario Sarcinelli-Filho, and Ricardo Carelli. Human-machine interfaces based on emg and eeg applied to robotic systems. *Journal of NeuroEngineering and Rehabilitation*, 5(1):1–15, 2008. 4.3

[67] Johanna Feuerstein, Daniel Austin, Robert Sack, and Tamara L Hayes. Wrist actigraphy for scratch detection in the presence of confounding activities. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3652–3655. IEEE, 2011. 2.2

[68] Roya Firoozi, Johnathan Tucker, Stephen Tian, Anirudha Majumdar, Jiankai Sun, Weiyu Liu, Yuke Zhu, Shuran Song, Ashish Kapoor, Karol Hausman, et al. Foundation models in robotics: Applications, challenges, and the future. *arXiv preprint arXiv:2312.07843*, 2023. 6.3.1

[69] Tobias W Fischer, Walter Wigger-Alberti, and Peter Elsner. Direct and non-direct measurement techniques for analysis of skin surface topography. *Skin Pharmacology and Physiology*, 12(1-2):1–11, 1999. 3.5

[70] Kevin C Fleming, Jonathan M Evans, and Darryl S Chutka. Caregiver and clinician shortages in an aging nation. In *Mayo Clinic Proceedings*, volume 78, pages 1026–1040. Elsevier, 2003. 7.2

[71] Déborah A Foloppe, Paul Richard, Takehiko Yamaguchi, Frédérique Etcharry-Bouyx, and Philippe Allain. The potential of virtual reality-based training to enhance the functional autonomy of alzheimer's disease patients in cooking activities: A single case study. *Neuropsychological rehabilitation*, 28(5):709–733, 2018. 5.2

[72] Daniel Gallenberger, Tapomayukh Bhattacharjee, Youngsun Kim, and Siddhartha S Srinivasa. Transfer depends on acquisition: Analyzing manipulation strategies for robotic feeding. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 267–276. IEEE, 2019. 7.3.1

[73] Aura Ximena González-Cely, Mauro Callejas-Cuervo, and Teodiano Bastos-Filho. Wheelchair prototype controlled by position, speed and orientation using head movement. *HardwareX*, 11:e00306, 2022. 4.3, 5.3.2, 5.4.5

[74] Daniel Gopher and Emanuel Donchin. Workload: An examination of the concept. 1986. 4.6.5

[75] Deepak Gopinath, Siddarth Jain, and Brenna D. Argall. Human-in-the-loop optimization of shared autonomy in assistive robotics. *IEEE Robotics and Automation Letters*, 2(1):247–254, 2017. doi: 10.1109/LRA.2016.2593928. 5.3.3, 5.4.6

[76] Ethan K Gordon, Xiang Meng, Tapomayukh Bhattacharjee, Matt Barnes, and Siddhartha S Srinivasa. Adaptive robot-assisted feeding: An online learning framework for acquiring previously unseen food items. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9659–9666. IEEE, 2020. 7.3.1

[77] Ethan K Gordon, Rajat Kumar Jenamani, Amal Nanavati, Ziang Liu, Daniel Stabile, Xilai Dai, Tapomayukh Bhattacharjee, Tyler Schrenk, Jonathan Ko, Haya Bolotski, et al. An adaptable, safe, and portable robot-assisted feeding system. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 74–76, 2024. 6.3.2, 7.3.1

[78] Ethan Kroll Gordon, Amal Nanavati, Ramya Challa, Bernie Hao Zhu, Taylor Annette Kessler Faulkner, and Siddhartha Srinivasa. Towards general single-utensil food acquisition with human-informed actions. In *Conference on Robot Learning*, pages 2414–2428. PMLR, 2023. 7.3.1

[79] Phillip M. Grice and Charles C. Kemp. In-home and remote use of robotic body surrogates by people with profound motor deficits. *PLOS ONE*, 14(3), 2019. doi: 10.1371/journal.pone.0212904. 1.3, 4.3, 5.3.1, 5.4, 5.5

[80] Luca Gualtieri, Erwin Rauch, and Renato Vidoni. Emerging research fields in safety and ergonomics in industrial collaborative robotics: A systematic literature review. *Robotics and Computer-Integrated Manufacturing*, 67:101998, 2021. 6.4.1

[81] Yao Guo, Xiangyu Liu, Shun Peng, Xinyu Jiang, Ke Xu, Chen Chen, Zeyu Wang, Chenyun Dai, and Wei Chen. A review of wearable and unobtrusive sensing technologies for chronic disease management. *Computers in Biology and Medicine*, 129:104163, 2021. 1

[82] Hyowon Ha, Tae-Hyun Oh, and In So Kweon. A multi-view structured-light system for highly accurate 3d modeling. In *2015 International Conference on 3D Vision*, pages 118–126. IEEE, 2015. 3.5

[83] Ayesha Hameed, Andrzej Ordys, Jakub Możaryn, and Anna Sibilska-Mroziewicz. Control system design and methods for collaborative robots: Review. *Applied Sciences*, 13(1), 2023. ISSN 2076-3417. doi: 10.3390/app13010675. URL https://www.mdpi.com/2076-3417/13/1/675. 5.3.3

[84] JM Hanifin and ML Reed. Eczema prevalence and impact working group. *A population-based survey of eczema prevalence in the United States. Dermatitis*, 18(2):82–91, 2007. 1.1, 2.2

[85] Sandra G Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50, pages 904–908. Sage publications Sage CA: Los Angeles, CA, 2006. 4.6.5

[86] Kelsey P Hawkins, Phillip M Grice, Tiffany L Chen, Chih-Hung King, and Charles C Kemp. Assistive mobile manipulation for self-care tasks around the head. In *2014 IEEE Symposium on computational intelligence in robotic rehabilitation and assistive technologies (CIR2AT)*, pages 16–25. IEEE, 2014. 1.1, 1.3, 4.2, 4.3, 5.3.1, 5.4, 5.5

[87] Jajack Heikenfeld, Andrew Jajack, Jim Rogers, Philipp Gutruf, Lei Tian, Tingrui Pan, Ruya Li, Michelle Khine, Jintae Kim, and Juanhong Wang. Wearable sensors: modalities, challenges, and prospects. *Lab on a Chip*, 18(2): 217–248, 2018. 7.3.2

[88] Halli Heimbuch, Yeong Rhee, Marty Douglas, Kirsten Juhl, Kelly Knoll, Sherri Stastny, and Ryan McGrath. Prevalence and trends of basic activities of daily living limitations in middle-aged and older adults in the united states. *Epidemiologia*, 4(4):483–491, 2023. 7.2

[89] Hiroki Higa, Kei Kurisu, and Hideyuki Uehara. A vision-based assistive robotic arm for people with severe disabilities. *Transactions on Machine Learning and Artificial Intelligence*, 2(4):12–23, 2014. 7.3.1

[90] Leigh R Hochberg, Daniel Bacher, Beata Jarosiewicz, Nicolas Y Masse, John D Simeral, Joern Vogel, Sami Haddadin, Jie Liu, Sydney S Cash, Patrick Van Der Smagt, et al. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature*, 485(7398):372–375, 2012. 4.3

[91] Delwar Hossain, Tonmoy Ghosh, and Edward Sazonov. Automatic count of bites and chews from videos of eating episodes. *Ieee Access*, 8:101934–101945, 2020. 7.2

[92] Brandi House, Jonathan Malkin, and Jeff Bilmes. The voicebot: a voice controlled robot arm. In *Proceedings of the SIGCHI Conference on Human*

Factors in Computing Systems, pages 183–192, 2009. 6.2

[93] Zichao Hu, Francesca Lucchetti, Claire Schlesinger, Yash Saxena, Anders Freeman, Sadanand Modak, Arjun Guha, and Joydeep Biswas. Deploying and evaluating llms to program service mobile robots. *IEEE Robotics and Automation Letters*, 2024. 6.3.1

[94] Shujun Huang, Zonghua Zhang, Yan Zhao, Jie Dai, Chao Chen, Yongjia Xu, E Zhang, and Lili Xie. 3d fingerprint imaging system based on full-field fringe projection profilometry. *Optics and Lasers in Engineering*, 52:123–130, 2014. 3.5

[95] Robert JK Jacob. Eye tracking in advanced interface design. *Virtual environments and advanced interface design*, 258:288, 1995. 4.3, 5.3.1

[96] Catrine Jacobsson, Karin Axelsson, Per Olov Österlind, and Astrid Norberg. How people with stroke and healthy older people experience the eating process. *Journal of Clinical Nursing (Wiley-Blackwell)*, 9(2), 2000. 7.2

[97] Siddarth Jain and Brenna Argall. Probabilistic human intent recognition for shared autonomy in assistive robotics. *J. Hum.-Robot Interact.*, 9(1), dec 2019. doi: 10.1145/3359614. URL https://doi.org/10.1145/3359614. 5.3.3, 5.4.6

[98] Shervin Javdani, Henny Admoni, Stefania Pellegrinelli, Siddhartha S. Srinivasa, and J. Andrew Bagnell. Shared autonomy via hindsight optimization for teleoperation and teaming. *The International Journal of Robotics Research*, 37 (7):717–742, June 2018. ISSN 0278-3649. doi: 10.1177/0278364918776060. URL https://doi.org/10.1177/0278364918776060. 5.3.3, 5.6.2, 7.5.3

[99] Shervin Javdani, Henny Admoni, Stefania Pellegrinelli, Siddhartha S Srinivasa, and J Andrew Bagnell. Shared autonomy via hindsight optimization for teleoperation and teaming. *The International Journal of Robotics Research*, 37(7): 717–742, 2018. 6.6.1

[100] Rajat Kumar Jenamani, Daniel Stabile, Ziang Liu, Abrar Anwar, Katherine Dimitropoulou, and Tapomayukh Bhattacharjee. Feel the bite: Robot-assisted inside-mouth bite transfer using robust mouth perception and physical interaction-aware control. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 313–322, 2024. 6.3.2, 7.2, 7.3.1

[101] Rajat Kumar Jenamani, Priya Sundaresan, Maram Sakr, Tapomayukh Bhattacharjee, and Dorsa Sadigh. Flair: Feeding via long-horizon acquisition of realistic dishes. *arXiv preprint arXiv:2407.07561*, 2024. 7.3.1

[102] Rajat Kumar Jenamani, Tom Silver, Ben Dodson, Shiqin Tong, Anthony Song, Yuting Yang, Ziang Liu, Benjamin Howe, Aimee Whitneck, and Tapomayukh Bhattacharjee. Feast: A flexible mealtime-assistance system towards in-the-wild personalization. In *Robotics: Science and Systems (RSS)*, 2025. 7.2, 7.3.1, 7.2,

7.1, 7.4.5

[103] Ju Ji, Jordan Venderley, Hui Zhang, Mengjue Lei, Guangchen Ruan, Neel Patel, Yu-Min Chung, Regan Giesting, and Leah Miller. Assessing nocturnal scratch with actigraphy in atopic dermatitis patients. *NPJ Digital Medicine*, 6(1):72, 2023. 2.2

[104] Xiaodan Jia, Rui Li, Mandayam A Srinivasan, and Edward H Adelson. Lump detection with a gelsight sensor. In *2013 World Haptics Conference (WHC)*, pages 175–179. IEEE, 2013. 3.2

[105] Micah K Johnson and Edward H Adelson. Retrographic sensing for the measurement of surface texture and shape. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1070–1077. IEEE, 2009. 3.2

[106] Ariel S Kapusta, Phillip M Grice, Henry M Clever, Yash Chitalia, Daehyung Park, and Charles C Kemp. A system for bedside assistance that integrates a robotic bed and a mobile manipulator. *Plos one*, 14(10):e0221854, 2019. 1.3, 5.4

[107] Arie E Kaufman, Amit Bandopadhay, and Bernard D Shaviv. An eye tracking computer user interface. In *Proceedings of 1993 IEEE Research Properties in Virtual Reality Symposium*, pages 120–121. IEEE, 1993. 4.3, 5.3.1

[108] Kento Kawaharazuka, Tatsuya Matsushima, Andrew Gambardella, Jiaxian Guo, Chris Paxton, and Andy Zeng. Real-world robot applications of foundation models: A review. *arXiv preprint arXiv:2402.05741*, 2024. 6.3.1

[109] Maya N Keely, Heramb Nemlekar, and Dylan P Losey. Kiri-spoon: A soft shape-changing utensil for robot-assisted feeding. *arXiv preprint arXiv:2403.05784*, 2024. 7.3.1

[110] Charles C Kemp, Aaron Edsinger, Henry M Clever, and Blaine Matulevich. The design of stretch: A compact, lightweight mobile manipulator for indoor human environments. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 3150–3157. IEEE, 2022. 5.2

[111] Mina J Kian, Mingyu Zong, Katrin Fischer, Abhyuday Singh, Anna-Maria Velentza, Pau Sang, Shriya Upadhyay, Anika Gupta, Misha A Faruki, Wallace Browning, et al. Can an llm-powered socially assistive robot effectively and safely deliver cognitive behavioral therapy? a study with university students. *arXiv preprint arXiv:2402.17937*, 2024. 6.3.1

[112] Callie Y Kim, Christine P Lee, and Bilge Mutlu. Understanding large-language model (llm)-powered human-robot interaction. *arXiv preprint arXiv:2401.03217*, 2024. 6.3.1, 6.7.2, 6.7.3, 6.7.4

[113] Jee-Eun Kim, Jongin Lee, Masahiro Bessho, and Ken Sakamura. Field trial of

a wearable itch monitoring system for the self-management of atopic dermatitis. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2457–2463. IEEE, 2021. 2.5

[114] Suhan Kim, Sukjun Kim, Houriyeh Majditehran, Dinesh K. Patel, Carmel Majidi, and Sarah Bergbreiter. Electromechanical characterization of 3d printable conductive elastomer for soft robotics. In *2020 3rd IEEE International Conference on Soft Robotics (RoboSoft)*, pages 318–324, 2020. doi: 10.1109/RoboSoft48309.2020.9116027. 3.3.7

[115] AB Kimball, AN Naegeli, E Edson-Heredia, C-Y Lin, C Gaich, E Nikaï, K Wyrwich, and G Yosipovitch. Psychometric properties of the itch numeric rating scale in patients with moderate-to-severe plaque psoriasis. *British Journal of Dermatology*, 175(1):157–162, 2016. 2.5

[116] Chih-Hung King, Tiffany L Chen, Zhengqin Fan, Jonathan D Glass, and Charles C Kemp. Dusty: an assistive mobile manipulator that retrieves dropped objects for people with motor impairments. *Disability and Rehabilitation: Assistive Technology*, 7(2):168–179, 2012. 1.1, 4.2

[117] Seema P Kini, Laura K DeLong, Emir Veledar, Anne Marie McKenzie-Brown, Michael Schaufele, and Suephy C Chen. The impact of pruritus on quality of life: the skin equivalent of pain. *Archives of dermatology*, 147(10):1153–1156, 2011. 1.1, 2.2

[118] NJ Korman, Y Zhao, J Pike, J Roberts, and E Sullivan. Increased severity of itching, pain, and scaling in psoriasis patients is associated with increased disease severity, reduced quality of life, and reduced work productivity. *Dermatology online journal*, 21(10), 2015. 1.1, 2.2

[119] Jan Kottner, Marianne Schario, Natalie Garcia Bartels, Ellina Pantchechnikova, Kathrin Hillmann, and Ulrike Blume-Peytavi. Comparison of two in vivo measurements for skin surface topography. *Skin research and technology*, 19(2): 84–90, 2013. 3.2

[120] Mike Lambeta, Po-Wei Chou, Stephen Tian, Brian Yang, Benjamin Maloon, Victoria Rose Most, Dave Stroud, Raymond Santos, Ahmad Byagowi, Gregg Kammerer, et al. Digit: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation. *IEEE Robotics and Automation Letters*, 5(3):3838–3845, 2020. 3.2

[121] Mirte Langeveld, Lara S van de Lande, Eimear O'Sullivan, Berend van der Lei, and Joris A van Dongen. Skin measurement devices to assess skin quality: A systematic review on reliability and validity. *Skin Research and Technology*, 28 (2):212–224, 2022. 3.2, 3.5

[122] Clemente Lauretti, Francesca Cordella, Francesco Scotto di Luzio, Stefano

Saccucci, Angelo Davalli, Rinaldo Sacchetti, and Loredana Zollo. Comparative performance analysis of m-imu/emg and voice user interfaces for assistive robots. In *2017 International Conference on Rehabilitation Robotics (ICORR)*, pages 1001–1006. IEEE, 2017. 6.2

[123] Michael Joseph Lavery, Carolyn Stull, Leigh A Nattkemper, Kristen M Sanders, Helen Lee, Shwetapadma Sahu, Rodrigo Valdes-Rodriguez, Yiong-Huak Chan, and Gil Yosipovitch. Nocturnal pruritus: prevalence, characteristics, and impact on itchyqol in a chronic itch population. *Acta dermato-venereologica*, 97(4): 513–515, 2017. 2.2

[124] Jongin Lee, Dae-ki Cho, Seokwoo Song, SeungHo Kim, Eunji Im, and John Kim. Mobile system design for scratch recognition. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1567–1572, 2015. 2.5

[125] Jongin Lee, Daeki Cho, Junhong Kim, Eunji Im, JinYeong Bak, Kyung Ho Lee, Kwan Hong Lee, and John Kim. Itchtector: A wearable-based mobile system for managing itching conditions. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 893–905, 2017. 2.2, 2.5

[126] Shuang Li, Jiaxi Jiang, Philipp Ruppel, Hongzhuo Liang, Xiaojian Ma, Norman Hendrich, Fuchun Sun, and Jianwei Zhang. A mobile robot hand-arm teleoperation system by vision and imu. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 10900–10906. IEEE, 2020. 4.3, 5.3.2

[127] Zefang Li, Gina Olson, Dinesh K. Patel, Lining Yao, and Carmel Majidi. Electrically controlled liquid crystal elastomer surfaces for dynamic wrinkling. *Advanced Intelligent Systems*, 6(2):2200402, 2024. doi: https://doi.org/10.1002/aisy.202200402. URL https://advanced.onlinelibrary.wiley.com/doi/abs/10.1002/aisy.202200402. 3.3.7

[128] Jacky Liang, Wenlong Huang, Fei Xia, Peng Xu, Karol Hausman, Brian Ichter, Pete Florence, and Andy Zeng. Code as policies: Language model programs for embodied control. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9493–9500. IEEE, 2023. 6.2, 6.3.1

[129] Ryan Lilien. Applied research and development of a three-dimensional topography system for imaging and analysis of striated and impressed tool marks for firearm identification using gelsight. In *Forensic Science Seminar ISSN*, 2015. 3.2

[130] Abby M Lindberg, F Garrett Reis, and Micah K Johnson. A handheld 3d surface measurement system. In *2021 IEEE 8th International Workshop on Metrology for AeroSpace (MetroAeroSpace)*, pages 149–153. IEEE, 2021. 3.2

[131] Rui Liu, Amisha Bhaskar, and Pratap Tokekar. Adaptive visual imitation learning for robotic assisted feeding across varied bowl configurations and food types. *arXiv preprint arXiv:2403.12891*, 2024. 7.3.1

[132] Dylan P. Losey, Craig G. McDonald, Edoardo Battaglia, and Marcia K. O'Malley. A Review of Intent Detection, Arbitration, and Communication Aspects of Shared Control for Physical Human–Robot Interaction. *Applied Mechanics Reviews*, 70(1):010804, 02 2018. ISSN 0003-6900. doi: 10.1115/1.4039145. URL https://doi.org/10.1115/1.4039145. 5.3.3

[133] Joanne Lynn, David M Adamson, et al. *Living well at the end of life: adapting health care to serious chronic illness in old age*. Rand Santa Monica, 2003. 1

[134] Jacob P Macdonald, Rohit Mallick, Allan B Wollaber, Jaime D Peña, Nathan McNeese, and Ho Chit Siu. Language, camera, autonomy! prompt-engineered robot control for rapidly evolving deployment. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 717–721, 2024. 6.3.1

[135] Pablo Maceira-Elvira, Traian Popa, Anne-Christine Schmid, and Friedhelm C. Hummel. Wearable technology in stroke rehabilitation: Towards improved diagnosis and treatment of upper-limb motor impairment. *Journal of Neuro-Engineering and Rehabilitation*, 16(1), 2019. doi: 10.1186/s12984-019-0612-y. 1.1, 4.2

[136] Rishabh Madan, Rajat Kumar Jenamani, Vy Thuy Nguyen, Ahmed Moustafa, Xuefeng Hu, Katherine Dimitropoulou, and Tapomayukh Bhattacharjee. Sparcs: Structuring physically assistive robotics for caregiving with stakeholders-in-the-loop. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 641–648. IEEE, 2022. 7.2

[137] Karthik Mahadevan, Jonathan Chien, Noah Brown, Zhuo Xu, Carolina Parada, Fei Xia, Andy Zeng, Leila Takayama, and Dorsa Sadigh. Generative expressive robot behaviors using large language models. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 482–491, 2024. 6.2, 6.3.1

[138] Nikhil Mahadevan, Yiorgos Christakis, Junrui Di, Jonathan Bruno, Yao Zhang, E Dorsey, Wilfred R Pigeon, Lisa A Beck, Kevin Thomas, Yaqi Liu, et al. Development of digital measures for nighttime scratch and sleep using wrist-worn wearable devices. *NPJ digital medicine*, 4(1):1–10, 2021. 2.2

[139] Nikola Marangunić and Andrina Granić. Technology acceptance model: a literature review from 1986 to 2013. *Universal access in the information society*, 14(1):81–95, 2015. 4.6.3

[140] Guilherme Marins, Diego Carvalho, André Marcato, and Ivo Junior. Develop-

ment of a control system for electric wheelchairs based on head movements. In *2017 Intelligent Systems Conference (IntelliSys)*, pages 996–1001. IEEE, 2017. 4.3, 5.3.2, 5.4.5

[141] Jorge S Marques, Catarina Barata, and Teresa Mendonça. On the role of texture and color in the classification of dermoscopy images. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4402–4405. IEEE, 2012. 3.2

[142] Ashley E Mason, Frederick M Hecht, Shakti K Davis, Joseph L Natale, Wendy Hartogensis, Natalie Damaso, Kajal T Claypool, Stephan Dilchert, Subhasis Dasgupta, Shweta Purawat, et al. Detection of covid-19 using multimodal data from a wearable device: results from the first tempredict study. *Scientific reports*, 12(1):1–15, 2022. 2.5

[143] Elizabeth N. Matheis, David S. Tulsky, and Robert J. Matheis. The relation between spirituality and quality of life among individuals with spinal cord injury. *Rehabilitation Psychology*, 51(3):265–271, 2006. doi: 10.1037/0090-5550.51.3.265. 4.2

[144] Thea F Meijman and Gijsbertus Mulder. Psychological aspects of workload. In *A handbook of work and organizational psychology*, pages 15–44. Psychology Press, 2013. 4.6.5

[145] Wolfgang Merkt, Yiming Yang, Theodoros Stouraitis, Christopher E. Mower, Maurice Fallon, and Sethu Vijayakumar. Robust shared autonomy for mobile manipulation with continuous scene monitoring. In *2017 13th IEEE Conference on Automation Science and Engineering (CASE)*, August 2017. doi: 10.1109/COASE.2017.8256092. 5.3.3

[146] Sean L Metzger, Kaylo T Littlejohn, Alexander B Silva, David A Moses, Margaret P Seaton, Ran Wang, Maximilian E Dougherty, Jessie R Liu, Peter Wu, Michael A Berger, et al. A high-performance neuroprosthesis for speech decoding and avatar control. *Nature*, pages 1–10, 2023. 5.2

[147] M Minderer, A Gritsenko, A Stone, M Neumann, D Weissenborn, A Dosovitskiy, A Mahendran, A Arnab, M Dehghani, Z Shen, et al. Simple open-vocabulary object detection with vision transformers. arxiv 2022. *arXiv preprint arXiv:2205.06230*. 5.4.6

[148] Raymond T Moore. An analysis of ridge-to-ridge distance on fingerprints. *J Forensic Ident*, 39(4):231–238, 1989. 3.3.8

[149] Arnaud Moreau, Peter Anderer, Marco Ross, Andreas Cerny, Timothy H Almazan, and Barry Peterson. Detection of nocturnal scratching movements in patients with atopic dermatitis using accelerometers and recurrent neural networks. *IEEE journal of biomedical and health informatics*, 22(4):1011–1018,

2017. 2.2

[150] Caroline Sian Murray and Jonathan L Rees. Are subjective accounts of itch to be relied on? the lack of relation between visual analogue itch scores and actigraphic measures of scratch. *Acta dermato-venereologica*, 91(1):18–23, 2011. 2.2

[151] Amal Nanavati, Patricia Alves-Oliveira, Tyler Schrenk, Ethan K Gordon, Maya Cakmak, and Siddhartha S Srinivasa. Design principles for robot-assisted feeding in social contexts. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, pages 24–33, 2023. 6.3.2, 7.2, 7.3.1, 7.4.4

[152] Amal Nanavati, Vinitha Ranganeni, and Maya Cakmak. Physically assistive robots: A systematic review of mobile and manipulator robots that physically assist people with disabilities. *Annual Review of Control, Robotics, and Autonomous Systems*, 7, 2023. 7.2

[153] Amal Nanavati, Ethan K Gordon, Taylor A Kessler Faulkner, Yuxin Ray Song, Jonathan Ko, Tyler Schrenk, Vy Nguyen, Bernie Hao Zhu, Haya Bolotski, Atharva Kashyap, et al. Lessons learned from designing and evaluating a robot-assisted feeding system for out-of-lab use. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 696–707. IEEE, 2025. 7.2, 7.3.1

[154] Ali Bou Nassif, Manar Abu Talib, Qassim Nasir, and Fatima Mohamad Dakalbab. Machine learning for anomaly detection: A systematic review. *Ieee Access*, 9:78658–78700, 2021. 5.6.5

[155] Mitsuru Nemoto, Eiju Uchinuma, and Shohei Yamashina. Three-dimensional analysis of forehead wrinkles. *Aesthetic plastic surgery*, 26:10–16, 2002. 3.2, 3.5

[156] Christa Neuper, Gernot R Müller, Andrea Kübler, Niels Birbaumer, and Gert Pfurtscheller. Clinical application of an eeg-based brain–computer interface: a case study in a patient with severe motor impairment. *Clinical neurophysiology*, 114(3):399–409, 2003. 5.2

[157] Mpho Ngoepe, Yahya E Choonara, Charu Tyagi, Lomas Kumar Tomar, Lisa C Du Toit, Pradeep Kumar, Valence MK Ndesendo, and Viness Pillay. Integration of biosensors and drug delivery technologies for early detection and chronic management of illness. *Sensors*, 13(6):7680–7713, 2013. 1

[158] Vy Nguyen. Increasing independence with stretch: A mobile robot enabling functional performance in daily activities. 1.3, 5.4, 5.5

[159] Alicja Ograczyk, Joanna Miniszewska, Anna Kepska, and Anna Zalewska-Janowska. Itch, disease coping strategies and quality of life in psoriasis patients. *Advances in Dermatology and Allergology*, 31(5):299–304, 2014. 1.1, 2.2

[160] Takeshi Okuyama, Kazuki Hatakeyama, and Mami Tanaka. Measurement of human scratch behavior using compact microphone. *International Journal of Applied Electromagnetics and Mechanics*, 45(1-4):731–737, 2014. 2.2

[161] Clara Oliveira and Tiago Torres. More than skin deep: the systemic nature of atopic dermatitis. *European Journal of Dermatology*, 29(3):250–258, 2019. 1.1, 2.2

[162] Jan Ondras, Abrar Anwar, Tong Wu, Fanjun Bu, Malte Jung, Jorge Jose Ortiz, and Tapomayukh Bhattacharjee. Human-robot commensality: Bite timing prediction for robot-assisted feeding in groups. In *6th Annual Conference on Robot Learning*, 2022. 7.2, 7.3.1, 7.4.4

[163] World Health Organization and International Spinal Cord Society. *International perspectives on spinal cord injury*. World Health Organization, 2013. 4.2, 5.2

[164] Bethel A. Osuagwu, Sarah Timms, Ruth Peachment, Sarah Dowie, Helen Thrussell, Susan Cross, Tony Heywood, Rebecca Shirley, and Julian Taylor. Clinical trial of the soft extra muscle glove to assess orthotic and long-term functional gain following chronic incomplete tetraplegia: Preliminary functional results. *Biosystems & Biorobotics*, page 385–389, 2018. doi: 10.1007/978-3-030-01845-0_77. 1.1, 4.2

[165] Akhil Padmanabha, Frederik Ebert, Stephen Tian, Roberto Calandra, Chelsea Finn, and Sergey Levine. Omnitact: A multi-directional high-resolution touch sensor. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 618–624. IEEE, 2020. 3.2

[166] Akhil Padmanabha, Sonal Choudhary, Carmel Majidi, and Zackory Erickson. A multimodal sensing ring for quantification of scratch intensity. *Communications Medicine*, 3(1):115, 2023. I, 2.3.4, 2.4.3, 7.3.2

[167] Akhil Padmanabha, Qin Wang, Daphne Han, Jashkumar Diyora, Kriti Kacker, Hamza Khalid, Liang-Jung Chen, Carmel Majidi, and Zackory Erickson. Hat: Head-worn assistive teleoperation of mobile manipulators. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 12542–12548. IEEE, 2023. II, 5.2, 5.3.2, 5.4.5, 6.2, 7.2, 7.3.2, 7.5.6

[168] Akhil Padmanabha, Janavi Gupta, Chen Chen, Jehan Yang, Vy Nguyen, Douglas J Weber, Carmel Majidi, and Zackory Erickson. Independence in the home: A wearable interface for a person with quadriplegia to teleoperate a mobile manipulator. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 542–551, 2024. II, 5.4.2, 5.4.3, 5.4.4, 5.5, 5.6, 6.2, 6.6.1, 6.7.3, 6.7.3, 7.2, 7.3.2, 7.5.6

[169] Akhil Padmanabha, Jessie Yuan, Janavi Gupta, Zulekha Karachiwalla, Carmel Majidi, Henny Admoni, and Zackory Erickson. Voicepilot: Harnessing llms as

speech interfaces for physically assistive robots. *arXiv preprint arXiv:2404.04066*, 2024. II, 6.4.2, 6.4.3, 6.5.1, 6.5.2, 6.6.1, 6.6.2, 7.2, 7.3.1

[170] Akhil Padmanabha, Saravanan Govindarajan, Hwanmun Kim, Sergio Ortiz, Rahul Rajan, Doruk Senkal, and Sneha Kadetotad. Egocharm: Resource-efficient hierarchical activity recognition using an egocentric imu sensor. *arXiv preprint arXiv:2504.17735*, 2025. 7.4.2

[171] Akhil Padmanabha, Carmel Majidi, and Zackory Erickson. Towards wearable interfaces for robotic caregiving. *arXiv preprint arXiv:2502.05343*, 2025. 7.3.2, 7.5.6

[172] C Pailler-Mattei, S Bec, and H Zahouani. In vivo measurements of the elastic mechanical properties of human skin by indentation tests. *Medical engineering & physics*, 30(5):599–606, 2008. 3.5

[173] Abhilash Pandya. Chatgpt-enabled davinci surgical robot prototype: advancements and limitations. *Robotics*, 12(4):97, 2023. 6.3.1

[174] Daehyung Park, Zackory Erickson, Tapomayukh Bhattacharjee, and Charles C Kemp. Multimodal execution monitoring for anomaly detection during robot manipulation. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 407–414. IEEE, 2016. 5.6.5

[175] Daehyung Park, Yuuna Hoshi, and Charles C Kemp. A multimodal anomaly detector for robot-assisted feeding using an lstm-based variational autoencoder. *IEEE Robotics and Automation Letters*, 3(3):1544–1551, 2018. 6.3.2

[176] Daehyung Park, Yuuna Hoshi, Harshal P Mahajan, Ho Keun Kim, Zackory Erickson, Wendy A Rogers, and Charles C Kemp. Active robot-assisted feeding with a general-purpose mobile manipulator: Design, evaluation, and lessons learned. *Robotics and Autonomous Systems*, 124:103344, 2020. 7.3.1, 7.4.4

[177] Dinesh K. Patel, Amir Hosein Sakhaei, Michael Layani, Biao Zhang, Qi Ge, and Shlomo Magdassi. Highly stretchable and uv curable elastomers for digital light processing based 3d printing. *Advanced Materials*, 29(15):1606000, 2017. doi: https://doi.org/10.1002/adma.201606000. URL https://advanced.onlinelibrary.wiley.com/doi/abs/10.1002/adma.201606000. 3.3.7

[178] Mariella Pazzaglia and Marco Molinari. The embodiment of assistive devices—from wheelchair to exoskeleton. *Physics of life reviews*, 16:163–175, 2016. 5.3.1

[179] Chamika Janith Perera, Thilina Dulantha Lalitharatne, and Kazuo Kiguchi. Eeg-controlled meal assistance robot with camera-based automatic mouth position tracking and mouth open detection. In *2017 IEEE international conference on robotics and automation (ICRA)*, pages 1760–1765. IEEE, 2017. 7.3.1, 7.4.4

[180] Emanuele Perini, Simone Soria, Andrea Prati, and Rita Cucchiara. Facemouse: A human-computer interface for tetraplegic people. In *European Conference on Computer Vision*, pages 99–108. Springer, 2006. 4.3, 4.5, 5.3.1

[181] Indrashis Podder, Himel Mondal, and George Kroumpouzos. Nocturnal pruritus and sleep disturbance associated with dermatologic disorders in adult patients. *International Journal of Women's Dermatology*, 7(4):403–410, 2021. 1.1, 2.2

[182] Samuel Poirier, François Routhier, and Alexandre Campeau-Lecours. Voice control interface prototype for assistive robots for people living with upper limb disabilities. In *2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR)*, pages 46–52. IEEE, 2019. 6.2

[183] Alexandra Price and David E Cohen. Assessment of pruritus in patients with psoriasis and atopic dermatitis: subjective and objective tools. *Dermatitis*, 25 (6):334–344, 2014. 2.2

[184] Temiloluwa Prioleau, Elliot Moore, and Maysam Ghovanloo. Unobtrusive and wearable systems for automatic dietary monitoring. *IEEE Transactions on Biomedical Engineering*, 64(9):2075–2089, 2017. 7.3.2

[185] Terrin Babu Pulikottil, Marco Caimmi, Maria Grazia D'Angelo, Emilia Biffi, Stefania Pellegrinelli, and Lorenzo Molinari Tosatti. A voice control system for assistive robotic arms: preliminary usability tests on patients. In *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*, pages 167–172. IEEE, 2018. 6.2

[186] Puvesh Punj, Peter G Devitt, Brendon J Coventry, and Robert J Whitfield. Palpation as a useful diagnostic tool for skin lesions. *Journal of Plastic, Reconstructive & Aesthetic Surgery*, 67(6):804–807, 2014. 3.2

[187] Daniel Rakita, Bilge Mutlu, Michael Gleicher, and Laura M. Hiatt. Shared control–based bimanual robot manipulation. *Science Robotics*, 4(30):eaaw0955, 2019. doi: 10.1126/scirobotics.aaw0955. URL https://www.science.org/doi/abs/10.1126/scirobotics.aaw0955. 5.3.3

[188] Vinitha Ranganeni. Customizing tele-operation interfaces of assistive robots at home with occupational therapists. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 142–144, 2024. 6.7.3

[189] Vinitha Ranganeni, Noah Ponto, and Maya Cakmak. Evaluating customization of remote tele-operation interfaces for assistive robots. *arXiv preprint arXiv:2304.02771*, 2023. 5.3.1, 6.7.3

[190] Manuel Rauscher, Melanie Kimmel, and Sandra Hirche. Constrained robot control using control barrier functions. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 279–285. IEEE, 2016.

6.4.1

[191] Linda Resnik, Susan Allen, Deborah Isenstadt, Melanie Wasserman, and Lisa Iezzoni. Perspectives on use of mobility aids in a diverse population of seniors: Implications for intervention. *Disability and health journal*, 2(2):77–85, 2009. 2.5

[192] Wendy A Rogers, Widya A Ramadhani, and Maurita T Harris. Defining aging in place: The intersectionality of space, person, and time. *Innovation in aging*, 4(4):igaa036, 2020. 6.7.2

[193] C Roques, L Téot, N Frasson, and S Meaume. Primos: an optical system that produces three-dimensional measurements of skin surfaces. *Journal of wound care*, 12(9):362–365, 2003. 3.5

[194] Eric Rosen and Doruk Senkal. Charm: A hierarchical deep learning model for classification of complex human activities using motion sensors. *arXiv preprint arXiv:2207.07806*, 2022. 7.4.2

[195] Renzo Rozzini, Giovanni B Frisoni, Luigi Ferrucci, Piera Barbisoni, Bruno Bertozzi, and Marco Trabucchi. The effect of chronic diseases on physical function. comparison between activities of daily living scales and the physical performance test. *Age and Ageing*, 26(4):281–287, 1997. 1

[196] Sohrab Saeb, Luca Lonini, Arun Jayaraman, David C Mohr, and Konrad P Kording. The need to approximate the use-case in clinical machine learning. *Gigascience*, 6(5):gix019, 2017. 2.3.4

[197] B M Sakakibara, W C Miller, S G Orenczuk, and D L Wolfe. A systematic review of depression and anxiety measures used with individuals with spinal cord injury. *Spinal Cord*, 47(12):841–851, 2009. doi: 10.1038/sc.2009.93. 4.2

[198] Kristen M Sanders and Tasuku Akiyama. The vicious cycle of itch and anxiety. *Neuroscience & Biobehavioral Reviews*, 87:17–26, 2018. 1.1, 2.2

[199] Edward Sazonov, Stephanie Schuckers, Paulo Lopez-Meyer, Oleksandr Makeyev, Nadezhda Sazonova, Edward L Melanson, and Michael Neuman. Non-invasive monitoring of chewing and swallowing for objective quantification of ingestive behavior. *Physiological measurement*, 29(5):525, 2008. 7.3.2

[200] Sebastian Schröer, Ingo Killmann, Barbara Frank, Martin Völker, Lukas Fiederer, Tonio Ball, and Wolfram Burgard. An autonomous robotic assistant for drinking. In *2015 IEEE international conference on robotics and automation (ICRA)*, pages 6482–6487. IEEE, 2015. 7.3.1, 7.4.4

[201] Mario Selvaggio, Marco Cognetti, Stefanos Nikolaidis, Serena Ivaldi, and Bruno Siciliano. Autonomy in Physical Human-Robot Interaction: A Brief Survey. *IEEE Robotics and Automation Letters*, 6(4):7989–7996, October 2021. ISSN

2377-3766, 2377-3774. doi: 10.1109/LRA.2021.3100603. 5.3.3

[202] Lorenzo Shaikewitz, Yilin Wu, Suneel Belkhale, Jennifer Grannen, Priya Sundaresan, and Dorsa Sadigh. In-mouth robotic bite transfer with visual and haptic sensing. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9885–9895. IEEE, 2023. 7.3.1

[203] Samantha E Shune. An altered eating experience: Attitudes toward feeding assistance among younger and older adults. *Rehabilitation Nursing Journal*, 45 (2):97–105, 2020. 7.2

[204] Jonathan I Silverberg, Keiki Hinami, William E Trick, and David Cella. Itch in the general internal medicine setting: a cross-sectional study of prevalence and quality-of-life effects. *American journal of clinical dermatology*, 17(6):681–690, 2016. 1.1, 2.2

[205] Sandra F Simmons and John F Schnelle. Feeding assistance needs of long-stay nursing home residents and staff time to provide care. *Journal of the American Geriatrics Society*, 54(6):919–924, 2006. 7.2

[206] Eric L Simpson, Thomas Bieber, Emma Guttman-Yassky, Lisa A Beck, Andrew Blauvelt, Michael J Cork, Jonathan I Silverberg, Mette Deleuran, Yoko Kataoka, Jean-Philippe Lacour, et al. Two phase 3 trials of dupilumab versus placebo in atopic dermatitis. *New England Journal of Medicine*, 375(24):2335–2348, 2016. 2.5

[207] Ishika Singh, Valts Blukis, Arsalan Mousavian, Ankit Goyal, Danfei Xu, Jonathan Tremblay, Dieter Fox, Jesse Thomason, and Animesh Garg. Progprompt: Generating situated robot task plans using large language models. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11523–11530. IEEE, 2023. 6.2, 6.3.1, 6.4.1

[208] Dmitry A. Sinyukov and Taskin Padir. A Novel Shared Position Control Method for Robot Navigation Via Low Throughput Human-Machine Interfaces. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3913–3920, Madrid, October 2018. IEEE. ISBN 978-1-5386-8094-0. doi: 10.1109/IROS.2018.8593921. 5.3.3

[209] Albert L Siu, Lynn H Spragens, Sharon K Inouye, R Sean Morrison, and Bruce Leff. The ironic business case for chronic care in the acute care setting. *Health Affairs*, 28(1):113–125, 2009. 1

[210] Jongsang Son, Jan Fridén, and Richard L. Lieber. Biomechanical modeling of brachialis-to-wrist extensor muscle transfer function for daily activities in tetraplegia. *JBJS Open Access*, 7(3), 2022. doi: 10.2106/jbjs.oa.22.00018. 1.1, 4.2

[211] Roger Andre Søraa, Gunhild Tøndel, Mark W Kharas, and J Artur Serrano.

What do older adults want from social robots? a qualitative research approach to human-robot interaction (hri) studies. *International Journal of Social Robotics*, 15(3):411–424, 2023. 6.7.2

[212] Priya Sundaresan, Suneel Belkhale, and Dorsa Sadigh. Learning visuo-haptic skewering strategies for robot-assisted feeding. In *6th Annual Conference on Robot Learning*, 2022. 7.3.1

[213] Priya Sundaresan, Jiajun Wu, and Dorsa Sadigh. Learning sequential acquisition policies for robot-assisted feeding. *arXiv preprint arXiv:2309.05197*, 2023. 7.2, 7.3.1

[214] Kelly Tai, Stefanie Blain, and Tom Chau. A review of emerging access technologies for individuals with severe motor impairments. *Assistive technology*, 20(4):204–221, 2008. 4.3, 4.5, 5.3.1

[215] Danielle M Taylor. Americans with disabilities: 2014. *US Census Bureau*, pages 1–32, 2018. 7.2

[216] Carmen Theek, Hagen Tronnier, Ulrike Heinrich, and Nicole Braun. Surface evaluation of living skin (sels) parameter correlation analysis using data taken from astronauts working under extreme conditions of microgravity. *Skin Research and Technology*, 26(1):105–111, 2020. 3.2

[217] James H Thrall. Prevalence and costs of chronic disease in a health care system structured for treatment of acute illness. *Radiology*, 235(1):9–12, 2005. 1

[218] C Trojahn, M Schario, G Dobos, U Blume-Peytavi, and J Kottner. Reliability and validity of two in vivo measurements for skin surface topography in aged adults. *Skin Research and Technology*, 21(1):54–60, 2015. 3.5

[219] Kazue Tsukahara, Yuichi Tamatsu, Yasushi Sugawara, and Kazuyuki Shimada. The relationship between wrinkle depth and dermal thickness in the forehead and lateral canthal region. *Archives of dermatology*, 147(7):822–828, 2011. 3.2, 3.5

[220] Michele Tufano, Marlou Lasschuijt, Aneesh Chauhan, Edith JM Feskens, and Guido Camps. Capturing eating behavior from video analysis: A systematic review. *Nutrients*, 14(22):4847, 2022. 7.2

[221] Julien Van Kuilenburg, Marc A Masen, and Emile Van Der Heide. Contact modelling of human skin: What value to use for the modulus of elasticity? *Proceedings of the institution of mechanical engineers, Part J: Journal of Engineering Tribology*, 227(4):349–361, 2013. 3.5

[222] Camilo Velez, Dinesh K. Patel, Sukjun Kim, Mahnoush Babaei, Cory R. Knick, Gabriel L. Smith, and Sarah Bergbreiter. Hierarchical integration of thin-film niti actuators using additive manufacturing for microrobotics. *Journal of*

Microelectromechanical Systems, 29(5):867–873, 2020. doi: 10.1109/JMEMS.2020.3019064. 3.3.7

[223] Sai Vemprala, Rogerio Bonatti, Arthur Bucker, and Ashish Kapoor. Chatgpt for robotics: Design principles and model abilities. *arXiv preprint arXiv:2306.17582*, 2023. 6.2, 6.3.1, 6.4.1

[224] Sai Vemprala, Shuhang Chen, Abhinav Shukla, Dinesh Narayanan, and Ashish Kapoor. Grid: A platform for general robot intelligence development. *arXiv preprint arXiv:2310.00887*, 2023. 6.3.1, 6.4.1, 6.7.4

[225] Viswanath Venkatesh. Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information systems research*, 11(4):342–365, 2000. 4.6.3

[226] Tri Vu, Feng Lin, Nabil Alshurafa, and Wenyao Xu. Wearable food intake monitoring technologies: A comprehensive review. *Computers*, 6(1):4, 2017. 7.3.2

[227] Edward H Wagner. Chronic disease management: what will it take to improve care for chronic illness? *Effective clinical practice*, 1(1), 1998. 1

[228] Edward H Wagner, Brian T Austin, and Michael Von Korff. Organizing care for patients with chronic illness. *The Milbank Quarterly*, pages 511–544, 1996. 1

[229] Chao Wang, Stephan Hasler, Daniel Tanneberg, Felix Ocker, Frank Joublin, Antonello Ceravola, Joerg Deigmoeller, and Michael Gienger. Lami: Large language models for multi-modal human-robot interaction. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–10, 2024. 6.2, 6.3.1

[230] Yongchang Wang, LG Hassebrook, and DL Lau. Noncontact, depth-detailed 3d fingerprinting. *SPIE Newsroom*, 98(100), 2009. 3.5

[231] Robin Whittemore and Jane Dixon. Chronic illness: the process of integration. *Journal of clinical nursing*, 17(7b):177–187, 2008. 1

[232] Jimmy Wu, Rika Antonova, Adam Kan, Marion Lepert, Andy Zeng, Shuran Song, Jeannette Bohg, Szymon Rusinkiewicz, and Thomas Funkhouser. Tidybot: Personalized robot assistance with large language models. *Autonomous Robots*, 47(8):1087–1102, 2023. 6.3.1, 6.7.3

[233] Bao Xi, Shuo Wang, Xuemei Ye, Yinghao Cai, Tao Lu, and Rui Wang. A robotic shared control teleoperation method based on learning from demonstrations. *International Journal of Advanced Robotic Systems*, 16(4):172988141985742, July 2019. ISSN 1729-8814, 1729-8814. doi: 10.1177/1729881419857428. 5.3.3

[234] Linfei Xiong, Chin Boon Chng, Chee Kong Chui, Peiwu Yu, and Yao Li. Shared control of a medical robot with haptic guidance. *International Journal of*

Computer Assisted Radiology and Surgery, 12(1):137–147, January 2017. ISSN 1861-6410, 1861-6429. doi: 10.1007/s11548-016-1425-0. 5.3.3

[235] Albert F Yang, Morgan Nguyen, Alvin W Li, Brad Lee, Keum San Chun, Ellen Wu, Anna B Fishbein, Amy S Paller, and Shuai Xu. Use of technology for the objective evaluation of scratching behavior: A systematic review. *JAAD international*, 5:19–32, 2021. 2.2

[236] Albert F Yang, Keum San Chun, Lian Yu, Jessica R Walter, Donghyun Kim, Jong Yoon Lee, Hyoyoung Jeong, Matthew C Keller, Dhruv R Seshadri, Matthew O Olagbenro, et al. Validation of a hand-mounted wearable sensor for scratching movements in adults with atopic dermatitis. *Journal of the American Academy of Dermatology*, 88(3):726–729, 2023. 2.2

[237] Chenguang Yang, Sai Chang, Peidong Liang, Zhijun Li, and Chun-Yi Su. Teleoperated robot writing using emg signals. In *2015 IEEE International Conference on Information and Automation*, pages 2264–2269. IEEE, 2015. 4.3

[238] Jehan Yang, Kent Shibata, Douglas Weber, and Zackory Erickson. High-density electromyography for effective gesture-based control of physically assistive mobile manipulators. *arXiv preprint arXiv:2312.07745*, 2023. 1.1, 7.2

[239] Ali K Yetisen, Juan Leonardo Martinez-Hurtado, Barış Ünal, Ali Khademhosseini, and Haider Butt. Wearables in medicine. *Advanced Materials*, 30(33): 1706910, 2018. 7.3.2

[240] G Yosipovitch, M Reaney, V Mastey, L Eckert, A Abbé, L Nelson, M Clark, N Williams, Z Chen, M Ardeleanu, et al. Peak pruritus numerical rating scale: psychometric validation and responder definition for assessing itch in moderate-to-severe atopic dermatitis. *British Journal of Dermatology*, 181(4): 761–769, 2019. 2.2

[241] Nuray Yozbatiran, Jeffrey Berliner, Marcia K O'Malley, Ali Utku Pehlivan, Zahra Kadivar, Corwin Boake, and Gerard E Francisco. Robotic training and clinical assessment of upper extremity movements after spinal cord injury: a single case report. *Journal of rehabilitation medicine*, 44(2):186, 2012. 5.2

[242] Jessie Yuan, Janavi Gupta, Akhil Padmanabha, Zulekha Karachiwalla, Carmel Majidi, Henny Admoni, and Zackory Erickson. Towards an llm-based speech interface for robot-assisted feeding. In *Adjunct Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, pages 1–4, 2024. II, 7.3.1

[243] Wenzhen Yuan, Siyuan Dong, and Edward H Adelson. Gelsight: High-resolution robot tactile sensors for estimating geometry and force. *Sensors*, 17(12):2762, 2017. 3.2, 3.3.6

[244] Wenzhen Yuan, Chenzhuo Zhu, Andrew Owens, Mandayam A Srinivasan, and

Edward H Adelson. Shape-independent hardness estimation using deep learning and a gelsight tactile sensor. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 951–958. IEEE, 2017. 3.2

[245] Bowen Zhang and Harold Soh. Large language models as zero-shot human models for human-robot interaction. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7961–7968. IEEE, 2023. 6.2, 6.3.1

[246] Dandan Zhang, Zicong Wu, Junhong Chen, Ruiqi Zhu, Adnan Munawar, Bo Xiao, Yuan Guan, Hang Su, Wuzhou Hong, Yao Guo, Gregory S. Fischer, Benny Lo, and Guang-Zhong Yang. Human-robot shared control for surgical robot based on context-aware sim-to-real adaptation. In *2022 International Conference on Robotics and Automation (ICRA)*, page 7694–7700. IEEE Press, 2022. doi: 10.1109/ICRA46639.2022.9812379. URL https://doi.org/10.1109/ICRA46639.2022.9812379. 5.3.3

[247] Tongjie Y Zhang and Ching Y. Suen. A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*, 27(3):236–239, 1984. 3.3.8

[248] Xufeng Zhao, Mengdi Li, Cornelius Weber, Muhammad Burhan Hafez, and Stefan Wermter. Chat with the environment: Interactive multimodal perception using large language models. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3590–3596. IEEE, 2023. 6.2, 6.3.1

[249] Yuanchao Zhu, Canjun Yang, Qianxiao Wei, Xin Wu, and Wei Yang. Human–robot shared control for humanoid manipulator trajectory planning. *Industrial Robot: the international journal of robotics research and application*, 47(3):395–407, February 2020. ISSN 0143-991X, 0143-991X. doi: 10.1108/IR-10-2019-0217. 5.3.3