EXPLORATORY RESEARCH AT THE CONVERGENCE OF FRONTIER TECHNOLOGIES TO ADVANCE DIGITAL HEALTH AND TELEREHABILITATION WITH EMPHASIS ON APPLICATIONS FOR GERIATRIC POPULATION

DISSERTATION

Submitted in Partial Fulfillment of

the Requirements for

the Degree of

DOCTOR OF PHILOSOPHY (Mechanical Engineering)

at the

NEW YORK UNIVERSITY
TANDON SCHOOL OF ENGINEERING

by

Hassam Khan Wazir

May 2024

EXPLORATORY RESEARCH AT THE CONVERGENCE OF FRONTIER TECHNOLOGIES TO ADVANCE DIGITAL HEALTH AND TELEREHABILITATION WITH EMPHASIS ON APPLICATIONS FOR GERIATRIC POPULATION

DISSERTATION

Submitted in Partial Fulfillment of

the Requirements for

the Degree of

DOCTOR OF PHILOSOPHY (Mechanical Engineering)

at the

NEW YORK UNIVERSITY
TANDON SCHOOL OF ENGINEERING
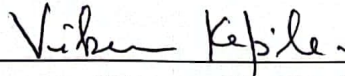
by

Hassam Khan Wazir

May 2024

Approved: *Katsuo Kurabayashi*

Department Chair Signature

04/26/2024

Date

Approved by the Guidance Committee:

Major: Mechanical Engineering

Vikram Kapila

Vikram Kapila, Ph.D., Chair
Professor, Mechanical and Aerospace Engineering Department
NYU Tandon School of Engineering

_04/23/2024_
Date

F. Atashzar

S. Farokh Atashzar, Ph.D.
Assistant Professor, Mechanical and Aerospace Engineering Department
NYU Tandon School of Engineering
04/23/2024
Date

Chen Feng, Ph.D.
Assistant Professor, Mechanical and Aerospace Engineering Department
NYU Tandon School of Engineering
04/23/2024
Date

Siddharth Garg, Ph.D.
Associate Professor, Electrical and Computer Engineering Department
NYU Tandon School of Engineering
04/24/2024
Date

Microfilm or other copies of this dissertation are obtainable from

Vita

Hassam Khan Wazir was born in Bannu, a city in the beautiful Khyber Pakhtunkhwa province of Pakistan. He received his Bachelor of Engineering degree in Electrical and Communication Engineering from Universiti Teknologi Brunei (formerly Institut Teknologi Brunei) in 2014. He completed his Master of Science degree in Mechatronics and Robotics at New York University (NYU), Tandon School of Engineering in the spring of 2018. In September 2018, he began pursuing a Ph.D. in Mechanical Engineering at the Mechatronics, Controls, and Robotics Laboratory (MCRL) at NYU Tandon School of Engineering under the supervision of Prof. Vikram Kapila. His research interests lie at the intersection of healthcare, mechatronics, and robotics. During his doctoral studies, he was a graduate instructor and taught the Measurement Systems Laboratory course and Automatic Control Systems Laboratory course. He was also an instructor at the NYU Summer Program in Automation, Robotics, and Coding (SPARC). His doctoral studies were supported by the NYU Tandon School of Engineering Fellowship.

Acknowledgements

To every kind soul who taught me something

ABSTRACT

EXPLORATORY RESEARCH AT THE CONVERGENCE OF FRONTIER TECHNOLOGIES TO ADVANCE DIGITAL HEALTH AND TELEREHABILITATION WITH EMPHASIS ON APPLICATIONS FOR GERIATRIC POPULATION

by

Hassam Khan Wazir

Advisor: Prof. Vikram Kapila, Ph.D.

Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy (Mechanical Engineering)

May 2024

This dissertation examines the implications of an aging global population on the healthcare system, highlighting the increasing preference for aging-in-place and the consequent burden on healthcare services and caregivers. With advancements in healthcare enabling more individuals to live longer, the demand for supportive care and rehabilitation services that allow older adults to remain in their homes

is also increasing. This scenario emphasizes the critical role of ambient assistive technology in facilitating aging-in-place, offering relief to caregivers and healthcare providers who are under increasing pressure. Therefore, a strong case is made to introduce modern, data-driven, approaches to facilitate more optimized strategies for delivering at-home treatment and enable older individuals to monitor their health and seek treatment before further health complications arise. Recognizing the vast landscape of caregiving, digital health, and telerehabilitation, this study focuses on three important areas: caregiving in the hospital setting, caregiving at home, and therapy compliance monitoring. More specifically, the dissertation discusses the implementation of teleoperated robots, controlled through smart devices in healthcare settings, to offer care while reducing the risk of patient contagion. It also examines the application of digital voice assistants and vision-based methods for assessing range of motion. Furthermore, it investigates the use of smart devices to monitor therapy, focusing on breathing and physical exercises. Finally, the document explores how consumer wireless devices and smartphones can be utilized for the indoor localization of elderly individuals, aiming to improve their safety and autonomy. By investigating these domains, the dissertation aims to shed light on innovative solutions that utilize consumer devices coupled with mechatronic and robotic systems to support caregivers across settings and enhance the effectiveness of therapy adherence among the elderly. Through a detailed exploration of these topics, the research contributes to the broader discourse on leveraging technology to meet the evolving needs of an aging population and the entities that provide them with care.

List of Publications

Journal Publications

[J1] A. RajKumar, F. Vulpi, S. R. Bethi, **H. K. Wazir**, P. Raghavan, and V. Kapila, "Wearable inertial sensors for range of motion assessment," *in IEEE Sensors Journal*, 20(7):3777–3787, 2020.

[J2] **H. K. Wazir**, C. Lourido, S. M. Chacko, and V. Kapila, "A COVID-19 emergency response for remote control of a dialysis machine with mobile HRI," *Frontiers in Robotics and AI*, vol. 8, 2021.

[J3] **H. K. Wazir**, A. Addagada, M. Bhadsorawala, and V. Kapila, "Indoor localization of elderly using ambient Wi-Fi," *in IEEE Sensors Journal*, manuscript in preparation.

[J4] **H. K. Wazir**, Z. Waghoo, and V. Kapila, "Acoustic and vision-based therapy compliance monitoring for lymphedema prevention," *in Scientific Reports*, manuscript in preparation.

Conference Publications

[C1] **H. K. Wazir**, S. R. Bethi, A. R. Kumar, F. Caruso, and V. Kapila, "A wearable pendant sensor to monitor compliance with range of motion lymphatic health exercise," in *IEEE International Conference on Engineering in Medicine & Biology Society*, 2020, pp. 4588–4591.

[C2] **H. K. Wazir**, K. Gaikwad, and V. Kapila, "Range of motion assessment using

a digital voice assistant," in *IEEE International Conference on Engineering in Medicine & Biology Society*, 2022, pp. 2577–2580.

[C3] M. Pawar, **H. K. Wazir**, and V. Kapila, "A lymphatic drainage robot for lymphedema rehabilitation," in *IEEE International Conference on Engineering in Medicine & Biology Society*, 2022, pp. 2598–2601.

[C4] **H. K. Wazir**, Z. Waghoo, and V. Kapila, "Wireless earphone-based real-time monitoring of breathing exercises: A deep learning approach," in *IEEE International Conference on Engineering in Medicine & Biology Society*, 2024, under review.

[C5] C. Lourido, Z. Waghoo, **H. K. Wazir**, N. Bhagat, and V. Kapila, "VR game for upper arm range of motion evaluation and rehabilitation using capability maps," in *IEEE International Conference on Engineering in Medicine & Biology Society*, 2024, under review.

Patent Applications

[P1] Vikram Kapila, Valentin Siderskiy, Armando Granado, Sahil Kumar, Tom Sowers, **Hassam Khan Wazir**, Miles Quinn Kilcourse, Sai Prasanth Krishnamoothy, Ryan Gonzalez, "Containment, treatment, and removal of aerosolized viral contamination," U.S. Patent Appl. US20230218463A1, Jun. 17, 2021.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The advent of digital health marks a significant turning point in the evolution of healthcare, representing a transformative shift that redefines the boundaries of traditional healthcare systems, patient care, and health management. This shift is characterized by a duality that benefits both patients and healthcare providers. For patients, digital health offers alternatives that can rival conventional treatments in efficacy, thereby promoting a sense of autonomy and empowerment. For healthcare professionals, digital health presents an opportunity to refine and accelerate the healthcare delivery process, optimizing treatment protocols, and ultimately improving patient outcomes. Despite being relatively nascent, digital health has the potential to unlock new possibilities for improving health outcomes, enhancing patient experiences, and making healthcare systems more resilient. Thus, digital health can not only address immediate healthcare challenges but also lay the groundwork for a future where healthcare is more accessible, personalized, and efficient.

This transition is further facilitated by the consistent reduction in the cost

of technology and the miniaturization of electronic devices, coupled with an unprecedented pace of innovation in new sensing techniques, user interfaces (UI), and human-robot interaction (HRI). The advent of smaller, more powerful single-board computers and the development of wearable devices and monitoring systems underscore the onset of a new era in digital health and general well-being. The increasing adoption of smartphone-based sensors and digital voice assistant (DVA) devices, along with the exploration of these technologies beyond their initial applications, as evidenced in recent studies [1], [2], indicates a future where digital health interventions become an integral part of our daily lives. At the heart of this transformation is the integration of traditional medical practices with advanced technologies such as artificial intelligence (AI), machine learning, wearable devices, and the Internet of Things (IoT). These technologies enable the development and deployment of novel, data-driven therapeutic approaches in ways that were unimaginable just a few decades ago. With the rapid development of new sensing modalities, data science empowers healthcare providers to make informed, evidence-based clinical decisions and generate personalized treatment plans. These technological strides are particularly pertinent in addressing the challenges posed by an aging global population, offering scalable solutions to improve the quality of life, and healthcare delivery for the elderly.

1.1   Aging global population

With rapid advancements in modern medicine and improved standards of healthcare, most of which were just achieved in the last century, the global average human life expectancy is at an all-time high. This means that people, on average,

tend to live longer now than at any time in recorded history. Although a major achievement for humanity in terms of public health and economic development, an aging global population also brings challenges. An increase in global average life expectancy directly correlates with an increase in the global elderly population. According to the United Nations [3], by the year 2050, almost 16% of the world population will be above the age of 65 which is a 6% increase from the statistics in 2022. This major shift in the demographics, along with global population growth, are termed two of the four global demographic "megatrends" and they pose challenges related to healthcare and wellness, among others.

The structure of the healthcare system in many developed countries, including the United States, is not designed to deal with such a significant shift in the demographic. Broadly speaking, healthcare systems in all countries rely on the younger population to take care of the older population. However, this begs the question, what happens when the older population surpasses the younger population in society? In the United States for example, by the year 2035 for the first time in history, the number of adults aged 65 years and above is projected to surpass the number of children under 18, resulting in older adults making up 21% of the population [4]. Dealing with such an increase would require fundamental changes in the healthcare system in the long term and technology can play a vital role in creating such a system.

To relieve some of the strain on an already overloaded healthcare system, e.g., that of the United States, recent advancements in sensing and connectivity can be utilized for wellness, remote monitoring, and patient care. These sensors and connected systems for healthcare applications enable patient-centric and individualized treatment, and such systems are broadly classified under the umbrella term

"Health-IoT" [5]. Such devices promote long-term patient assessment and the data they gather can potentially be used for early diagnosis of several diseases and health complications. Such a preemptive approach to healthcare can lead to more optimized strategies for delivering at-home treatment, promote aging in place, and enable older individuals to monitor their health and seek treatment before further health complications arise.

1.2   Aging in place

Also called community-dwelling, aging in place refers to older adults living in their homes with some level of independence. Living at home is a fundamental right of every individual and the primary choice of most older adults [6, 7]. Aside from relieving some burden from the healthcare system, it also helps the elderly maintain their psychological well-being through social connectedness [8], which can mitigate the detrimental physical and psychological impact of isolation [9]. To achieve this, however, effective systems must be in place to ensure that the quality of care and attention given to the elderly is not compromised in their homes. Professionally trained caregivers can ensure the physical and mental well-being of elderly and help them carry out activities of daily living. While the benefits of caregivers cannot be overstated, their scarce availability is another problem entirely.

1.3   Caregiver shortage

The enduring problem of professional caregiver shortage, both globally and in the US, has been studied extensively [10, 11, 12, 13, 14]. Despite the large projected demand for healthcare support occupations for the period 2021–2031 [15],

young caregivers remain a scarce human resource, as evident from the prevailing healthcare support staffing shortages [14]. One alternative to professional caregivers is the often-overlooked workforce of informal caregivers, comprising largely of family and friends [16]. Although not professionally trained, such individuals provide a valuable service by offsetting the burden from the healthcare system without a monetary cost to the government. The primary tasks performed by caregivers include: assisting the elderly with medication management, facilitating doctor visits, aiding with activities of daily living, and monitoring overall well-being to identify potential problems and track any disease progression [17]. These tasks necessitate human intervention and supervision, yet several of these tasks, notably monitoring the elderly, can be greatly supplemented with technologies that facilitate ambient assisted living (AAL). In the face of a growing caregiver shortage, digital health technologies, especially in the realm of Health-IoT and AAL, emerge as vital tools to bridge the gap, ensuring continuous and personalized care.

1.4   Digital health in a post COVID-19 world

The importance of digital health has been further accentuated by recent global health crises, such as the Coronavirus Disease 2019 (COVID-19) pandemic, which accelerated the adoption of digital and robotic solutions for safer and more efficient healthcare delivery [18, 19]. The deployment of digital tools for contact tracing, outbreak surveillance, and the delivery of healthcare services has been instrumental in curbing the spread of infections and in orchestrating efficient public health responses. Moreover, the adoption of mechatronics and telerobotics for facilitating remote monitoring and control of patients and medical devices alike can significantly

enhance patient outcomes as indicated by previous research [20]. While direct evidence pertaining to COVID-19 patients specifically is yet to be fully established, insights drawn from these studies suggest that such approaches could be beneficial in managing care for those affected by the virus, providing a promising direction for future healthcare strategies.

1.5   Human-robot interaction

The push towards digital innovation in healthcare is not limited to software and sensing technologies. The field of HRI represents a frontier where technology meets tactile care, offering a glimpse into a future where robots augment human healthcare providers. The last few decades have witnessed widespread adoption of robotic solutions by several industries for operations that are considered difficult or dangerous for humans to perform [21]. In the automotive industry, for example, heavy-duty industrial manipulators form an integral part of the assembly line [22], and one would be hard-pressed to find an automotive manufacturing facility that does not employ some sort of robotic assistance. Moreover, robots are actively being developed, examined, and used for inspection, decontamination, and decommissioning of nuclear plants [23, 24]; search and rescue operations following natural, industrial, and man-made disasters [25]; and exploration in outer space [26]. The above examples have one common thread, i.e., obviating the exposure to harm and risk to human safety. Thus, when operating in hazardous environments, in most cases the robots act as a physical extension of their human operators to enhance their dexterity, sensory experience, and cognition [27]. Endowing a human operator with the ability to utilize the robot to its maximum potential

requires the development of intuitive user interfaces for HRI. In recent years, several advancements have been made to render the HRI as seamless as it can be.

HRI is a rapidly advancing research field with several active areas of application that include human-supervised control of robots, autonomous robot control, and human-robot social interaction [28]. The human supervised control can be further divided into proximal *versus* remote control, the latter of which includes teleoperation and telerobotics [29]. In a hazard-prone, high-risk environment, the use of remotely controlled robots is preferable over proximally controlled robots because the human operator can perform the required tasks from a safe remote location. Varied HRI modalities for telerobotics have been developed over the years and each approach achieves a particular objective. Some early examples of HRI for telerobotics include using a joystick for teleoperation [30], performing stroke gestures on a touchscreen [31], and pointing gestures using a camera [32] and using a wearable sleeve [33].

The journey through digital health advancements, from wearable technologies to AI and robotics, circles back to the core aspiration of transforming healthcare into a more accessible, personalized, and efficient system. Each technological stride, be it in response to a pandemic or the integration of robots in care, layers onto the foundation of digital health, promising a future where healthcare challenges are met with innovative solutions.

1.6 Document organization

This dissertation is structured as follows. Chapters 2 to 5 present the core findings from the research conducted. These chapters detail four key research

initiatives: (1) remote control of a dialysis machine with mobile HRI (Chapter 2); (2) range of motion (ROM) assessment using a digital voice assistant (Chapter 3); (3) wireless earphone-based real-time monitoring of breathing exercises (Chapter 4); and (4) indoor localization of the elderly using ambient Wi-Fi (Chapter 5). The dissertation concludes in Chapter 6 that summarizes the outcomes of aforementioned studies and explores potential avenues for future investigation. Further details are provided in the subsequent sections.

1.7 Collaborator contributions

The research presented in this dissertation was enriched by the contributions of students and researchers from the Mechatronics, Controls, and Robotics Laboratory (MCRL), including Anirudh Addagada, Mustafa Bhadsorawala, Dr. Sonia Mary Chacko, Kshitij Gaikwad, Christian Lourido, and Zaid Waghoo.

- In Chapter 2, Dr. Sonia Mary Chacko, a recent NYU Tandon graduate and former colleague, and fellow doctoral candidate Christian Lourido supported the author in exploring mobile HRI for dialysis machine operation during COVID-19. Lourido contributed to the programming and control of the mobile manipulator, as well as user data collection. Dr. Chacko assisted in organizing user studies and analyzing the data collected.

- In Chapter 3, Kshitij Gaikwad, a Master's student at MCRL, aided in the development of software for extracting 2D joint coordinates for human pose detection and authored the initial program for exercise inference.

- In Chapter 4, Zaid Waghoo, a former Master's student at MCRL now serving

as an adjunct faculty member at NYU, played a crucial role in generating synthetic data, training models, and model inference. His expertise was also vital in network parameter tuning and model testing.

- In Chapter 5, Anirudh Addagada and Mustafa Bhadsorawala, both Master's students at MCRL, contributed to the early design and testing of the Wi-Fi localization system, including data collection. Bhadsorawala also developed the filtering algorithm for the system using inertial measurement units (IMUs). Both students laid the foundational work for this research during their Master's projects.

Chapter 2

Remote control of a dialysis machine with mobile HRI

2.1 Introduction

In recent years, as mobile devices (e.g., smartphones and tablets) have become ubiquitous in our personal and work settings, users have gained increased comfort in utilizing the rear-facing cameras of mobile devices to interact with their environments. Since mobile devices with well-endowed sensing, interaction, communication, and computing functionality are readily available to the common user, mobile mixed-reality interfaces have become greatly accessible and do not require research-grade devices to implement algorithms that were previously thought to be computationally expensive. Recent implementations of augmented reality (AR) based approaches include tracking single or multiple fiducial markers on the robot [34, 35] or its surroundings [36, 37] to determine the pose of the robot or objects in its workspace. Other studies have used this approach for multi-robot tracking

and control [38]. Although marker-based tracking has its merits, with the advent of markerless technologies, e.g., Google's AR Core [39, 40], the tracking can be performed in even unstructured environments while using highly intuitive user interfaces. Studies such as [36] and [37] have explored the potential of directing a robot manipulator to perform pick-and-place tasks using virtual elements in a semi-autonomous manner with the aid of a human collaborator. Another study suggests the use of virtual waypoints to guide a robot along a path [41]. With tele-robotics and human-robot interaction (HRI) being used for myriad applications, we propose to use these approaches in a healthcare setting and show that telerobotics and intuitive HRI can obviate the need for patients and healthcare workers to be exposed to high-risk interactions during a pandemic.

Medical caregivers such as doctors and nurses share physical space and interact with patients routinely. These shared spaces have a higher concentration of pathogens, making their occupants particularly susceptible to contracting bacterial and viral infections. The situation is exacerbated in the case of an epidemic, or more importantly, a pandemic, which can lead to a widespread shortage of personal protective equipment (PPE) and increase the risk of contagion for both caregivers and patients in a medical facility. A recent example of this situation was the spread of the COVID-19 pandemic across the world, including in the United States. In 2020, there was a massive global shortage of PPE, including face masks, eye protection, respirators, gloves, and gowns [42]. This shortage became a major barrier to responding effectively to the pandemic and in mitigating the resulting spread of COVID-19. Thus, essential healthcare workers, such as first responders, nurses, and doctors were forced to forgo or reuse PPE when working with patients with or without COVID-19 to preserve their limited stocks. Additionally, the novel

coronavirus was found to transmit asymptomatically, i.e., through infected patients who do not yet display any symptoms [43], at a significant rate, thus markedly increasing the likelihood of cross-contamination during the treatment and care of all patients. Healthcare workers were additionally exposed to the risk of infection through interaction and contact with fomites, including medical devices or instrument panels, and subsequently transmitting the disease to coworkers [44]. Many healthcare providers caring for COVID-19 patients were infected and even lost their lives due to a lack of sufficient access to PPE [45]. In addition to increasing the strain on an already overloaded healthcare system, such a lack of protection poses a significant threat to the morale of healthcare workers and their families.

With the shortage of PPE, patients without COVID-19 who needed critical and/or life-saving treatments also faced increased risk in healthcare facilities [46], including patients on dialysis, who predominantly comprise the elderly [47]. Such patients tend to be severely immunocompromised and are at a high risk of suffering serious complications if infected by the virus, as reported in China [46]. To minimize the risk of cross-contamination and infection, hospitals and dialysis centers implemented strict protocols with multiple additional precautions in dialysis units for staff members, patients, and their family members [46]. However, during the COVID-19 pandemic, dialysis centers were plagued by staff, equipment, and PPE shortages. In fact, at the peak of the pandemic in New York City, a headline in the city's paper of record *The New York Times* declared that "Dialysis Patients Face Close-Up Risk From Coronavirus" [48]. During this period, healthcare workers sought to minimize visits with dialysis patients by using baby monitors and performing physical interaction with dialysis machines without fully entering the patient rooms (see Figure 2.1A). To mitigate the plight of these patients and avoid

healthcare worker exposure, concerned authorities, such as the Food and Drug Administration (FDA), encouraged expanding the non-invasive remote monitoring of such patients [49]. To remotely determine whether a specific patient requires help, many healthcare device manufacturers started rolling out IoT devices to remotely monitor bio-signals relating to their temperature, heart rates, respiration rates, etc., [50]. These remote systems are important tools for avoiding the overcrowding of emergency rooms and hospitals and reducing the unnecessary exposure of vulnerable people to pathogens during situations such as pandemics.

Historically, most of the research around medical robotics has concentrated on surgical teleoperation robots such as the DaVinci robot (Intuitive Surgical, Mountain View, CA), and is more focused on patient safety during surgical procedures by mitigating human error and promoting minimally invasive procedures. Other medical robotic approaches focus on augmenting the doctor's vision with virtual overlays to provide additional information [51, 52]. Some social and companion robots are available that target the elderly [53] or serve as emotional support [54], however there is a dearth of examples of telerobots that can be used to manipulate medical devices using intuitive HRI. There are autonomous robots that can deliver medications throughout hospitals [55], and a study explored the development of a tele-nursing robot [56] that can navigate and interact with objects in the environment, but these solutions are either not relevant to this work or are cost prohibitive to be rapidly deployed in case of a pandemic.

In this work, we propose to create an emergency, non-invasive remote monitoring and control response system that addresses the needs of a highly vulnerable population: patients with severe kidney diseases. A viable solution for remotely monitoring and controlling a dialysis machine's instrumentation panel poses several

design challenges. Typically, dialysis centers consist of multiple reclining chairs or beds with attendant dialysis machines placed next to them (see Figure 2.1B). Potential solutions for remotely manipulating the dialysis machine's instrument panel include: (1) accessing embedded firmware of medical devices and (2) retrofitting the machine with a teleoperated robotic manipulator. As medical devices are sensitive instruments with proprietary firmware, varied software architectures, and individualized system requirements, it is not feasible to create a generalized framework to access the embedded firmware for remotely monitoring and controlling different medical instruments using smartphone/tablet-based third-party apps, especially as expeditiously as a pandemic emergency demands. Thus, retrofitting dialysis machines with teleoperated robotic arms, which can be easily mounted or removed as needed, is deemed the most viable option. We envision a remote-monitoring-and-control framework wherein a camera-equipped robotic manipulator interacts with the instrument control panel of the dialysis machine, thus reducing the risk of COVID-19 exposure for both patients and healthcare providers. Our proposed solution can address the shortage of PPE in healthcare facilities during a pandemic, enabling patients who require dialysis to continue receiving the life-saving treatment in isolation. At the same time, staff members in dialysis units can continue to provide high-quality care with a relatively low risk of cross-contamination. This work's engineering merits involve piloting a framework to quickly retrofit available dialysis machines with robust off-the-shelf four degrees-of-freedom (DoF) robotic manipulators and supporting remote management of the device instrumentation panel with high fidelity. Thus, in the proof-of-concept study of this work, we recreate and live-stream the instrument control panel touchscreen (ICPT) of a commonly used dialysis machine, the Gambro X-36 Phoenix (Baxter International

Inc., Deerfield, IL) (see Figure 2.2), to replicate and access it on a remote user's tablet computer touchscreen (TCT). Moreover, we develop the control framework for the robot manipulator to achieve precise and accurate remote manipulation of the dialysis machine's ICPT. We test our intuitive smartphone/tablet-based interface with over 30 users.

The chapter is organized as follows. Section 2.2 elaborates on the materials and methods used in the study. This section provides details on the design of the robot manipulator and the user interface, the development of the communication architecture and marker detection, and the robot operation. Section 2.3 explains the system evaluation metrics used in this study. These metrics include a quantitative study about the accuracy of the robot and the user interaction, as well as qualitative studies about the user experience while operating the robot remotely. Following this, the results of the system evaluation are provided and discussed in section 2.4, and an improvement is suggested to render a distortion-free perception of the ICPT on the user TCT. Finally, section 2.5 provides concluding statements.

2.2   Materials and methods

The method proposed in this work uses an off-the-shelf four-DoF robotic manipulator equipped with a USB camera. The robot base and camera stand are fixed on a board, making the system installation and operation simple, just requiring the user to properly locate the robot in front of its workspace and point the camera to a touchscreen (representing a dialysis machine ICPT) with which the robot manipulator is required to interact. The HRI UI consisting of a mobile application (App) is connected to the same wireless network as the robot manipulator system.
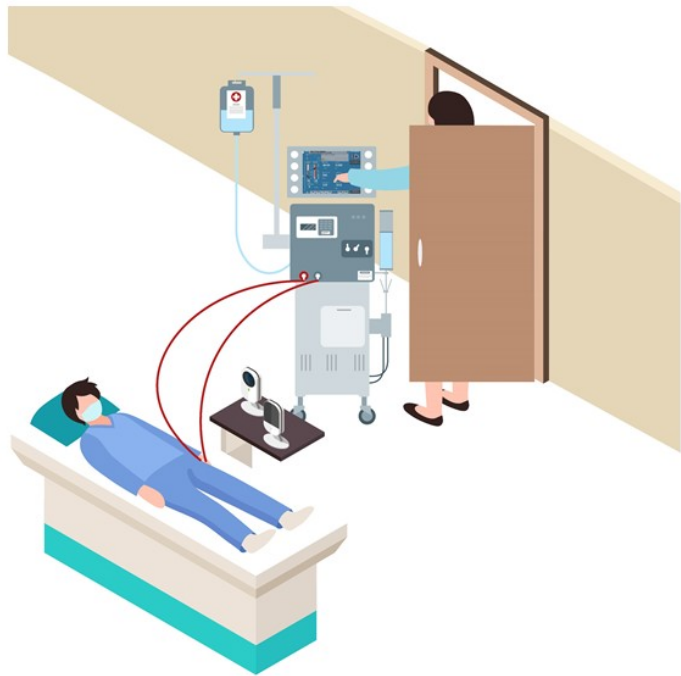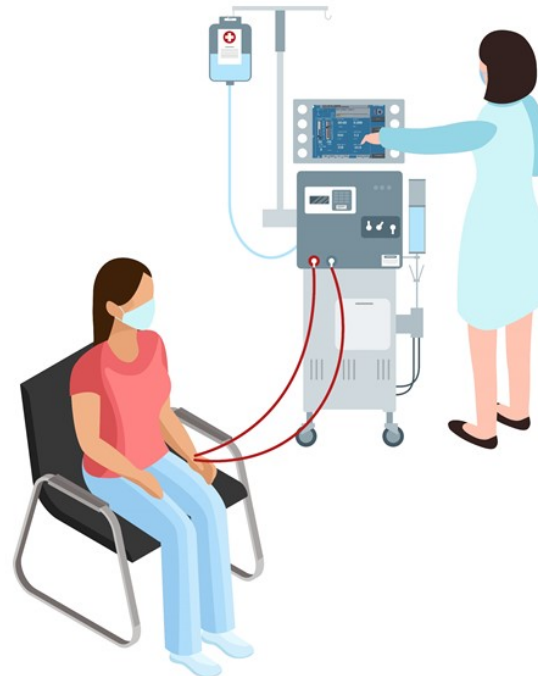
Figure 2.1: **(A)** Schematic representation of a dialysis patient receiving treatment at a hospital during the COVID-19 pandemic. The image shows the use of baby monitors and the reluctance of healthcare workers to enter the patient room. **(B)** A typical schematic representation of a patient receiving treatment at a dialysis center.
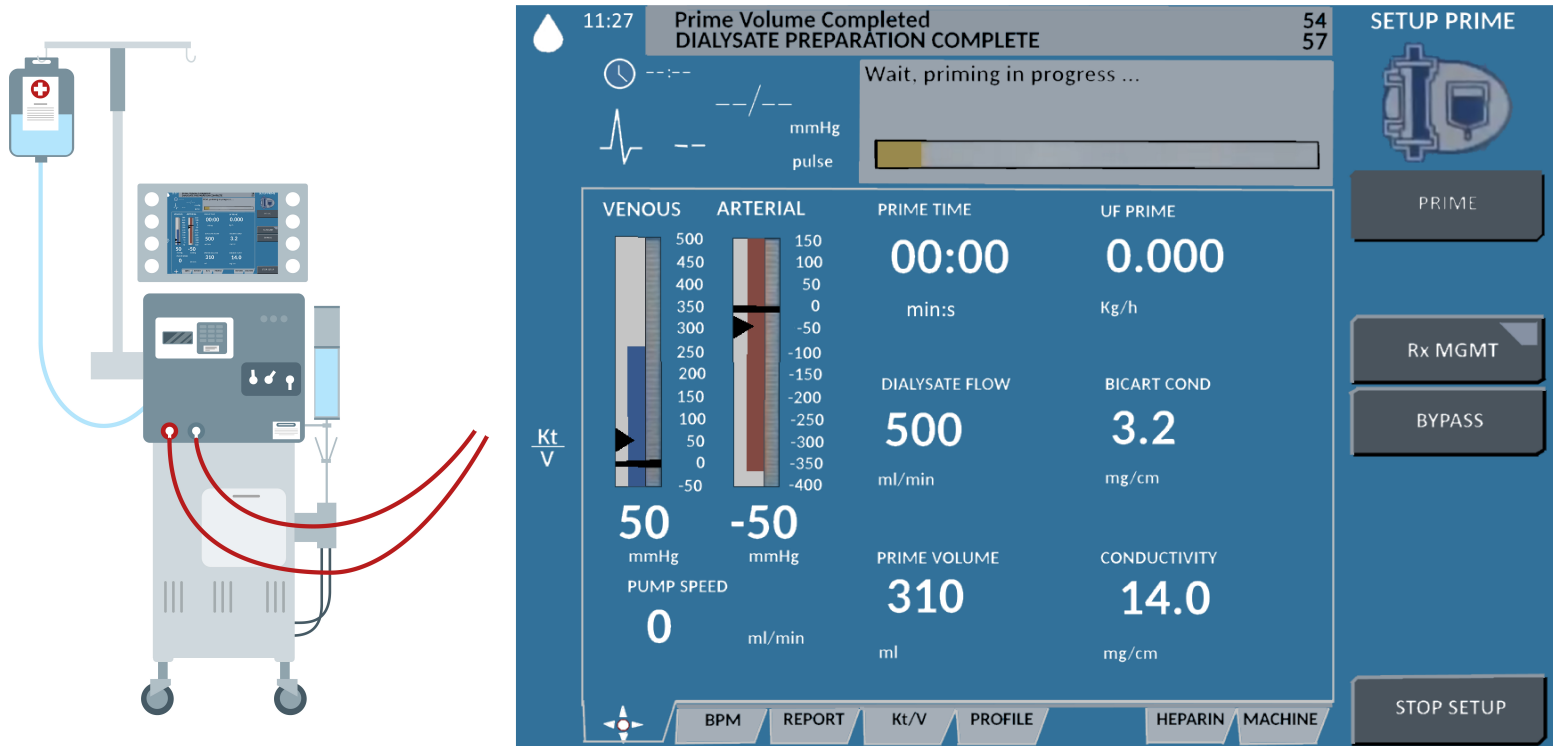
Figure 2.2: Schematic representation of Gambro X-36 Phoenix dialysis machine and its touchscreen interface.

To identify the surface plane of action of the robot, the mobile App uses the camera's video feed which includes a 2D image marker located in the plane of the ICPT, in front of the robot manipulator. The mobile App determines this plane of action (i.e., robot workspace) based on the dimensions of the robot and its *computed* position relative to the image marker. With the mobile App executing on a hand-held smartphone or tablet, when a user taps on the TCT at any location of the displayed surface of operation, an algorithm transforms the tapped location's pixel coordinates to a corresponding location coordinate in the workspace and frame of reference of the robot and sends it to the robot manipulator controller. Given this *commanded* position, another algorithm on the robot manipulator controller uses inverse kinematics to calculate a set of joint angles that can be used to attain the given position and orientation of the robot end effector and provides a solution to reach the specified location in space [57]. Then, in a sequence of steps, the system plans a path, moves the robot manipulator to go to the desired location on the ICPT, the robot taps on the desired location, and returns to its home position to wait for the next instruction. Figure 2.3 illustrates the components and interconnections of the proposed dialysis machine HRI environment.

2.2.1   Robot hardware

The robotic platform used in this study is a modified version of the Robotis OpenManipulator-X [58]. Based on the Robot Operating System (ROS) framework, this system is open-source and open-hardware, i.e., its controllers and CAD models of most of its components are accessible and free to use (see Figure 2.4A). This robot platform's system configuration is a four DoF arrangement, with a pen holder tool holding a stylus pen (see Figure 2.4B), which interacts with the ICPT during
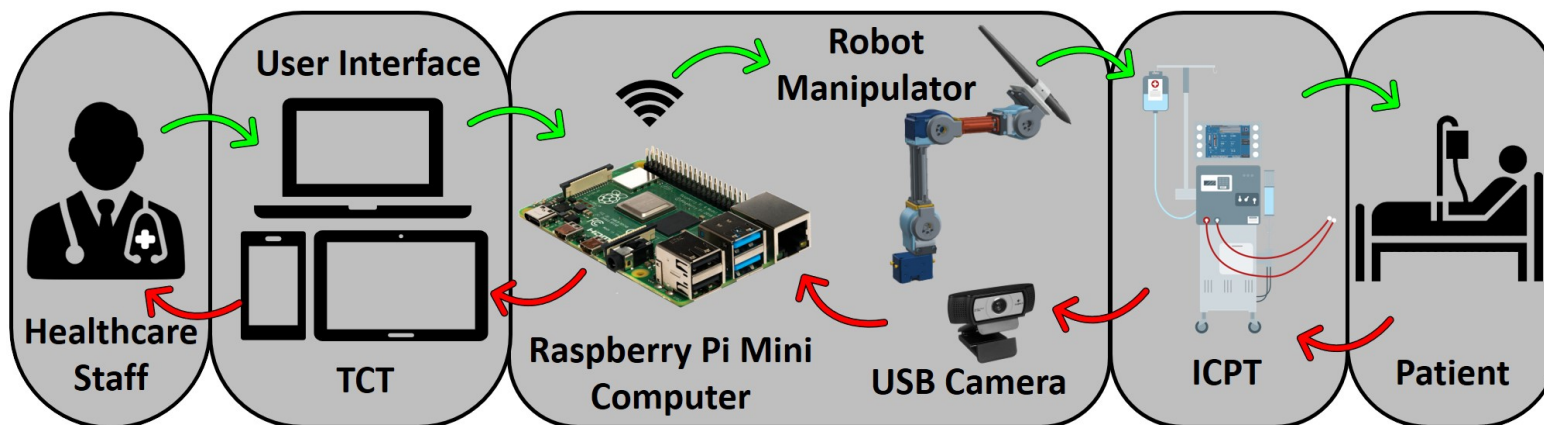
Figure 2.3: Schematic of a remote monitoring and control system for medical instruments. A healthcare worker interacts with the video feed from a camera on a user interface (UI) hosted on a tablet computer touchscreen (TCT). The user commands are processed to control a robot manipulator to interact with the instrument control panel touchscreen (ICPT) of a dialysis machine serving a patient.
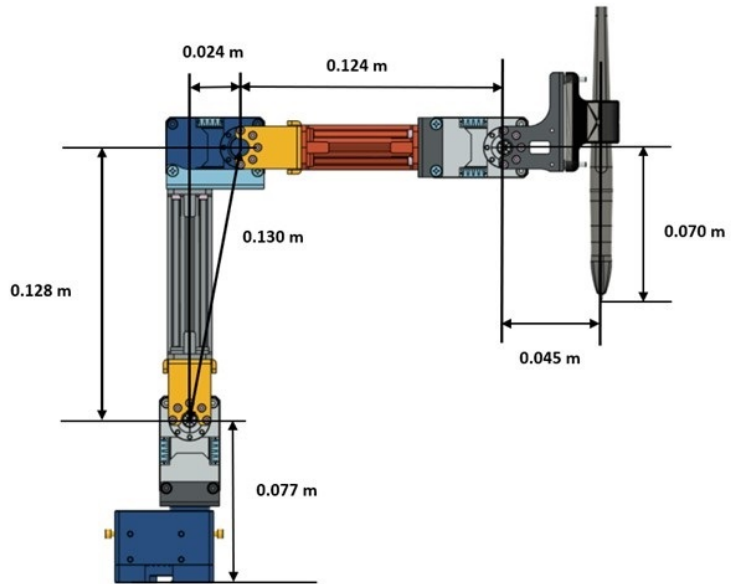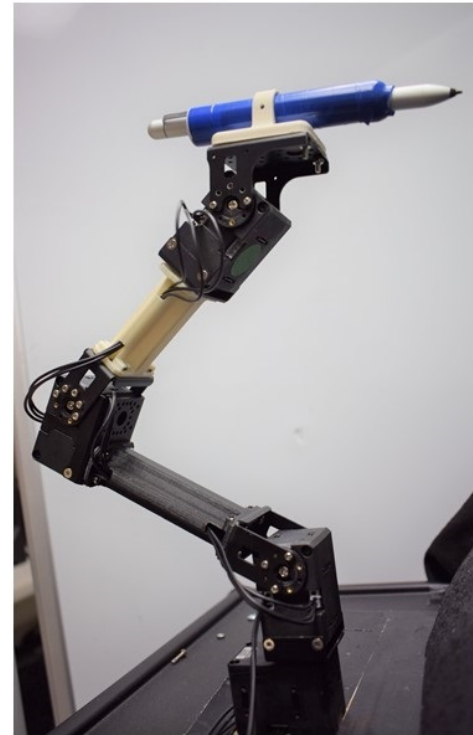
Figure 2.4: Robot manipulator prototype: **(A)** prototype CAD model and **(B)** built prototype.

operation. For the controller to function correctly, its program has been altered to account for the modified end effector, the number of actuators used, and each link's dimensions to accurately calculate the forward and inverse kinematics. The modified manipulator consists of four Dynamixel XL430-W250-T servomotors and two 3D-printed links made of polylactic acid (PLA) that are connected using metal brackets (see Figure 2.4A). The end effector is a PLA 3D-printed pen holder that holds the stylus pen to interact with the screen. The load capacity of the modified manipulator is conservatively estimated to be 160 g which can easily accommodate the 15 g end effector and 20 g stylus pen. A Raspberry Pi 4 (RPi4), with 4GB of RAM and with ROS Melodic installed on Raspbian-Buster OS, controls the robot manipulator using the ROS packages executing on it. Using this powerful and cost-effective single-board microcomputer gives the system sufficient capacity to control the robot and run computer vision algorithms without compromising the system's memory. Its small dimensions also make it simple to install and locate it near the system without interfering with the robot manipulator workspace.

To determine the workspace of the robot manipulator, the forward kinematics are first determined using the Denavit-Hartenberg (D-H) convention [59]. Then, the Monte Carlo method is employed to generate the manipulator's work envelope using the forward kinematics equations along with a random sampling of permissible joint angles [60]. This method produces a graphical representation of the manipulator workspace [61] that in turn is used to determine the range of ideal positions to install the robot relative to the medical device ICPT monitor. The allowable maximum and minimum distances between the robot and the medical device ICPT are determined to be 0.27 m and 0.20 m, respectively. The maximum distance is determined as the maximum distance between the robot and the ICPT which

ensures that the entirety of the ICPT lies within the estimated workspace of the robot. The minimum distance is obtained by placing the ICPT as close to the robot as possible while ensuring that all of the interactions and the fiducial marker on the ICPT remain visible to the camera (see Figure 2.5).

To establish the achievable accuracy and repeatability of the robot, tests are conducted by commanding it to move the end effector from its home position of $(x = 0.09, y = 0.0, z = 0.284)$ m to a test position and then returning the end effector to its home position. This test is conducted for five test positions, one at each corner of the ICPT and one at the center, with the position of each test point measured relative to the lower left corner of the ICPT. Moreover, the process is repeated 50 times for each test point and the computed accuracy and repeatability are provided in Table 2.1. Note that the accuracy represents the distance between the desired test position and the average of the achieved positions. Moreover, the repeatability represents the radius of the smallest circle that encompasses all of the achieved positions corresponding to a desired test position [62].

2.2.2 Camera calibration

The USB camera used in this setup is a C920 HD Pro Webcam, configured to capture a 640×480 image. By executing the camera driver on the ROS System, the webcam capture is made available as a ROS topic and becomes accessible to any subscribing program. Next, we perform a one-time geometric camera calibration using a pattern on a planar surface [63], allowing the system to correct the image for lens distortion and to detect and measure objects in world units by determining camera location in the scene. These calibration parameters are estimated using an available ROS package for camera calibration and are stored as a file, to be later
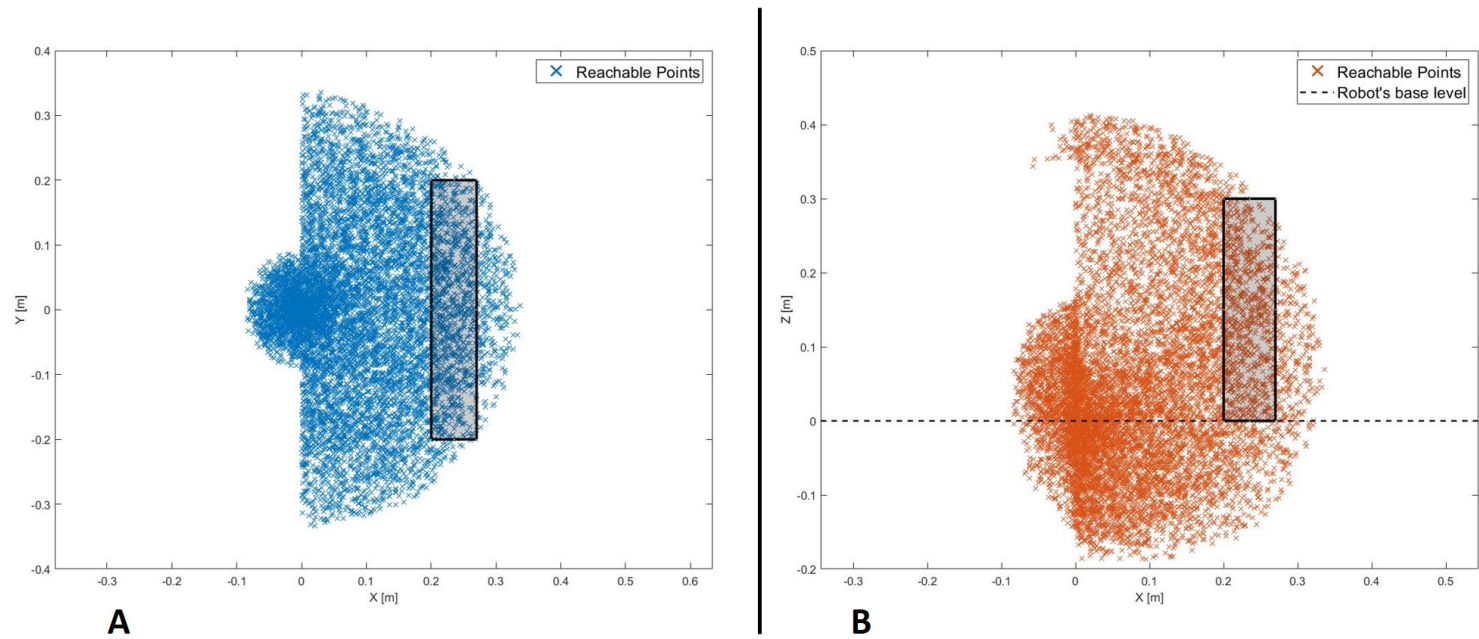
Figure 2.5: Workspace of the robot with the rectangular regions showing the range of allowable positions for the ICPT: **(A)** top-view of the workspace and **(B)** side-view of the workspace.

Table 2.1: Robot accuracy and repeatability test results

	$P_1(u,v)$	$P_2(u,v)$	$P_3(u,v)$	$P_4(u,v)$	$P_5(u,v)$
Ideal (mm)	$(128.5, 84.3)$	$(55.1, 151.6)$	$(206, 151)$	$(204.5, 18.1)$	$(56, 20)$
Accuracy (mm)	0.55	0.06	0.21	0.19	1.00
Repeatability (mm)	1.29	1.13	0.88	1.76	1.03

used during operation by the HRI interface and estimate spatial coordinates.

2.2.3   Communication with HRI interface

Using the built-in Wi-Fi adapter of the RPi4, the information generated and published by the nodes running on the ROS system is made accessible to all members of the network on which the microcomputer is connected. Using a WebSocket server node on ROS establishes a communication bridge and allows web interaction with the ROS topics using an IP address and a port number. Upon joining as a client, the mobile App used for the HRI interface communicates with the RPi4 server and accesses the information running on the ROS system. This mobile HRI interface, developed using the Unity Engine (Unity Technologies, San Francisco, CA) and a freely available ROS asset, lets the App publish and subscribe to ROS topics (see Figure 2.6). When the application first starts on the mobile device, it immediately looks for the IP address and port to establish communication with the RPi4 microcomputer. The RPi4 and the mobile HRI interface are connected to an *ad hoc* wireless network created using a Netgear Nighthawk X10 AD7200 Wi-Fi router. For the laboratory environment of this study, the maximum range of the wireless network is experimentally obtained to be 27 m.
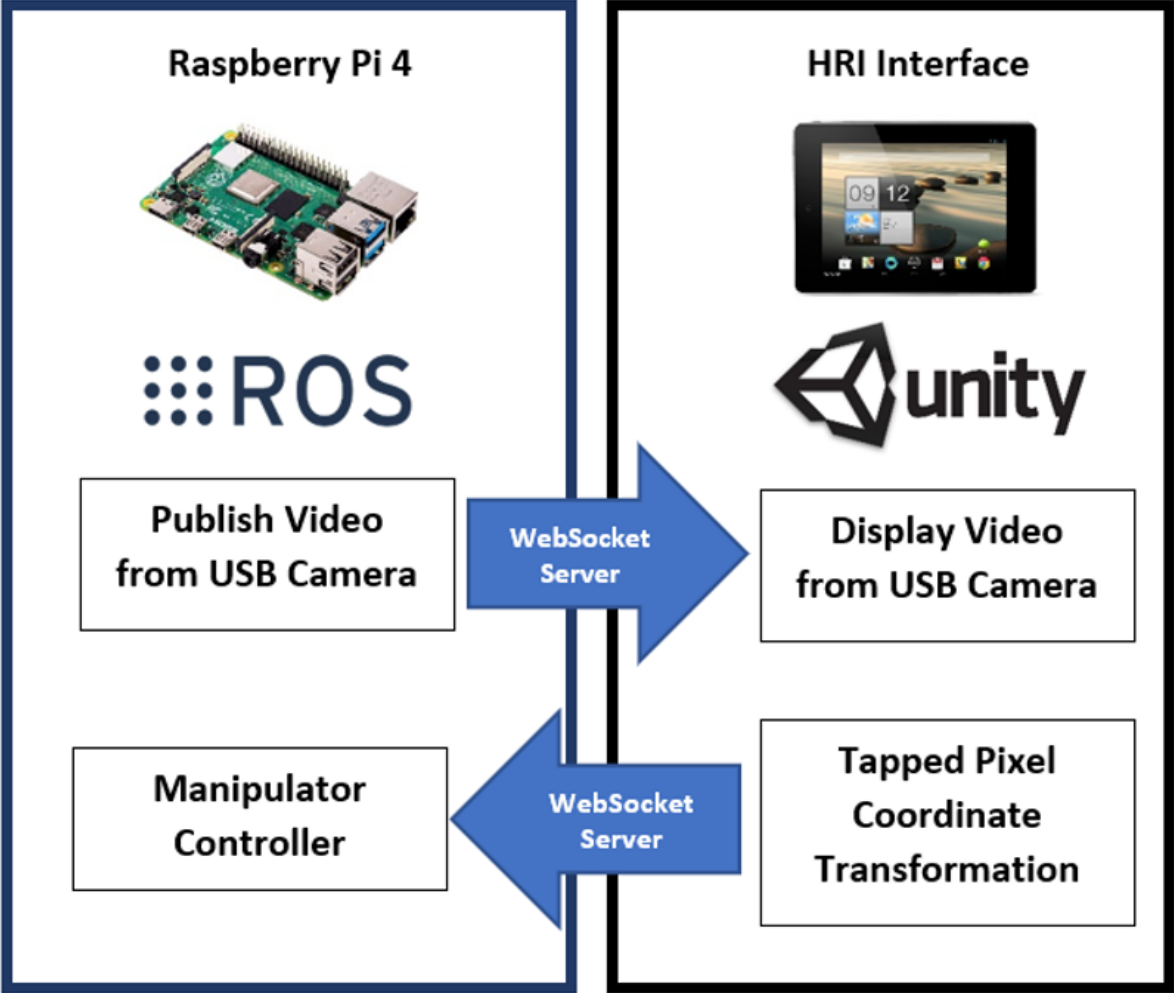
Figure 2.6: Communication between RPi4 and App.

2.2.4 Reference marker detection

The approaches initially considered for the design of the HRI user interface in this work can be distinguished by the number of reference markers affixed on the medical device ICPT monitor, i.e., (1) four markers (projective transformation) approach and (2) single marker approach.

2.2.4.1 Four markers approach

Using the projective transformation technique [64], this approach allows the estimation of any location on the instrumentation control panel displayed in the video feed on the touchscreen monitor. The four markers are placed on each corner of the ICPT monitor (see Figure 2.7A) and detected from the USB camera capture. The video feed of the camera is used to estimate its real-world 3D pose (relative to the plane formed by the four markers) and subsequently to compute the pose of any point on the ICPT monitor relative to the camera's coordinate frame. In this approach, the user can select each button of the ICPT by touching the corresponding location of the button on the streaming video image shown on the UI of the TCT (see Figure 2.7A, top panel). Moreover, the markers' detected points are used to correct the perspective distortion caused by the placement of the camera relative to the ICPT monitor and to scale the image to fit it on the UI of the TCT display. This method relies on two assumptions: (1) visual markers affixed to the ICPT monitor and interactive control elements (buttons and sliders) of the instrument control panel are on the same plane (coplanar points) and (2) the base location of the robot relative to the camera position can be estimated (see subsection 2.2.5). Even though this approach can allow our system to interact with any medical machine with an ICPT, regardless of the ICPT function arrangements,

placement of four markers on the same plane as the machine screen, in some cases, may block portions of the display containing important information for the machine functionality.

2.2.4.2 Single marker approach

This approach uses only one reference marker (see Figure 2.7B). The system localizes the robot relative to the marker's position using marker corners as correspondences to perform a projective transformation, but reducing the accuracy of the estimation (compared to the four markers approach) due to the lower number of correspondences detected. With this in mind, the robot control needs to be pre-programmed using the *a priori* knowledge about the locations of the on-screen control elements (buttons and sliders) relative to the attached marker to establish a one-to-one correspondence. For example, when the user touches button A on the UI of the TCT, the corresponding location (u_1, v_1) for the ICPT needs to be assigned automatically as the intended location. The UI executing on the TCT consists of a streaming video panel and a button panel. For each button on the medical device ICPT monitor, a corresponding button is available on the button panel of the UI on the TCT. This approach also assumes that the location of the robot relative to the camera position can be estimated. However, requiring information about the arrangement of control elements on the ICPT to pre-program the UI of the TCT will limit the usability of this arrangement since the on-screen layouts of control panels may vary between machines of different manufacturers and especially for different medical machines. Moreover, not having the well-defined four corners of the surface plane of action (as in the four markers approach) limits the system's ability to accurately correct perspective distortion (see Figure 2.7B, top panel).
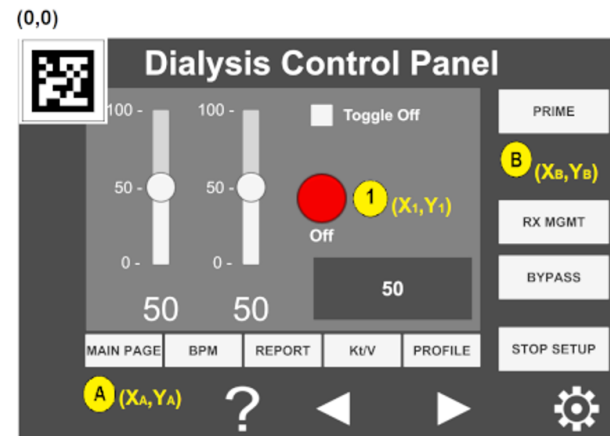
Figure 2.7: Reference marker detection: (**A**) four marker approach and (**B**) single marker approach.

2.2.4.3   Hybrid approach

In this work, we present an early *proof-of-concept* that employs a hybrid approach by building on the two methods discussed above (see Figure 2.8). By subscribing to the image published by the camera driver node on the ROS system, the mobile App gains access to its video feed that contains a single ArUco marker [65] placed on the top-left corner of the screen and detects it using the open-source ArUco module [66] of the Open Source Computer Vision Library (OpenCV). Instead of requiring a pre-programmed control panel on the UI of the TCT with known locations of the control elements on the ICPT (as in the single marker approach), the UI now detects the reference marker's corners, and an algorithm estimates the homography [67] between the camera image to the surface plane of the reference marker. With this transformation and the information from the camera calibration file, the mobile App maps coordinates of a user-selected pixel on the the video streamed image on the UI of the TCT to a spatial coordinate on the ICPT in world units, relative to the camera's reference frame. As in the previous approaches, this approach assumes that the robot base location relative to the camera can be estimated. Its functionality is similar to the four markers approach, letting the user select a control element (button or slider) of the ICPT by touching its corresponding location on the UI's image on the TCT. However, its accuracy may be compromised due to the limited number of correspondences detected.

As described above, the usability of the hybrid approach benefits the system by not relying on the *a priori* knowledge of the arrangement of the control elements on medical device ICPT or on risking portions of the ICPT being blocked by the placement of multiple markers, however, it has less accuracy than the four markers approach. The hybrid approach will also not correct the captured image's
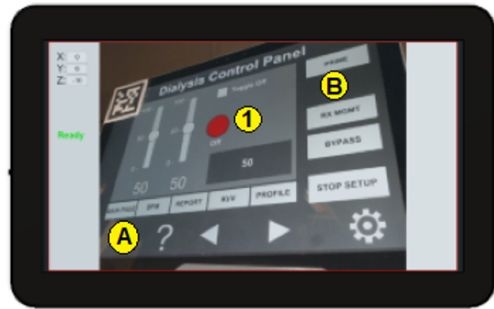
User's view through App

Live video (with perspective distortion)

A
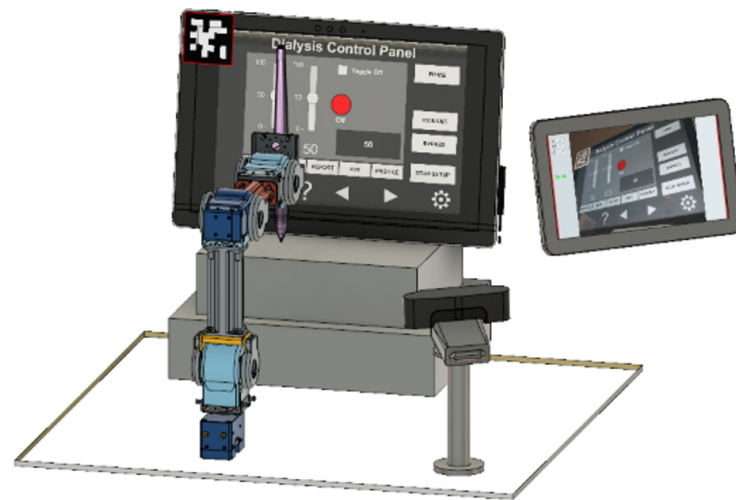
B

Figure 2.8: Reference marker detection: **(A)** hybrid marker detection approach and **(B)** complete setup with hybrid approach.

perspective of the USB camera for the UI displayed on the TCT.

2.2.5 Camera position and robot calibration

To allow the robot manipulator to interact with a point in its workspace (on ICPT) corresponding to any point selected by the user on the mobile App screen (on TCT), the robot controller requires the corresponding spatial coordinate specified in the robot's frame of reference (located on the center of the robot base). This necessitates imparting the system knowledge about the camera's pose relative to the robot frame of reference ($^{R}T_C$). Thus, a calibration routine is created and implemented before the system starts any HRI operations. That is, this routine is run immediately after the camera's orientation has been established to capture the robot's workspace surface (i.e., the ICPT monitor).

We first locate the ArUco marker in a predefined pose relative to the robot's reference frame (see Figure 2.9). With this known pose ($^{R}T_M$) and with the pose of the marker relative to the camera reference frame ($^{C}T_M$), estimated by the mobile App, the calibration routine computes $^{R}T_C$ as follows

$$^{R}T_C = {}^{R}T_M({}^{C}T_M)^{-1}. \tag{2.1}$$

Now $^{R}T_C$ is stored on and used by the mobile App to map the pixel location of any point tapped by the user on the TCT to a spatial coordinate on the ICPT in the robot's reference frame.

To achieve a mapping from the TCT to an ICPT of any size, the user enters, in millimeters, the width and height of the ICPT, and the u and v offsets of the top left corner of the ICPT from the center of the fiducial marker, into the App.

This creates an interactive region on the TCT that is the size of the ICPT as seen on the video feed on the TCT. Next, to map any desired point on the ICPT to the robot's workspace, we first locate the fiducial marker of known size (40 mm × 40 mm) on the ICPT surface. Based on the size and orientation of the marker obtained using computer vision, the App obtains the marker's pose relative to the camera position. It uses this information to map any pixel coordinate to a space coordinate relative to the camera frame. Finally, using the transformation matrix ($^{R}T_{C}$) obtained in the calibration step, the desired interaction point on the ICPT is mapped to spatial coordinates in the robot arm's coordinate frame. This coordinate serves as the input to command the robot to move to the desired position. As long as the ICPT is located within the robot workspace and its entire screen (with the fiducial marker located on it) is visible to the camera, the robot can reach any desired point. Finally, once the App maps the TCT coordinate into a spatial coordinate, it is published to a ROS topic, making it available to the robot manipulator controller. However, even if the calculation of $^{R}T_{C}$ is accurate, there may be slight residual errors in the end effector's final position. To compensate for this, the second part of the calibration routine consists of commanding the robot to go to the center of the reference marker multiple times. The user moves the end effector's final position by tapping on the UI screen at preprogrammed buttons, which are displayed during the calibration routine, to manipulate the stylus pen's tip in the X, Y, and Z directions until it matches the marker's center as precisely as possible. The offset values needed to reach the actual desired position are stored and used to increase accuracy during the operation.
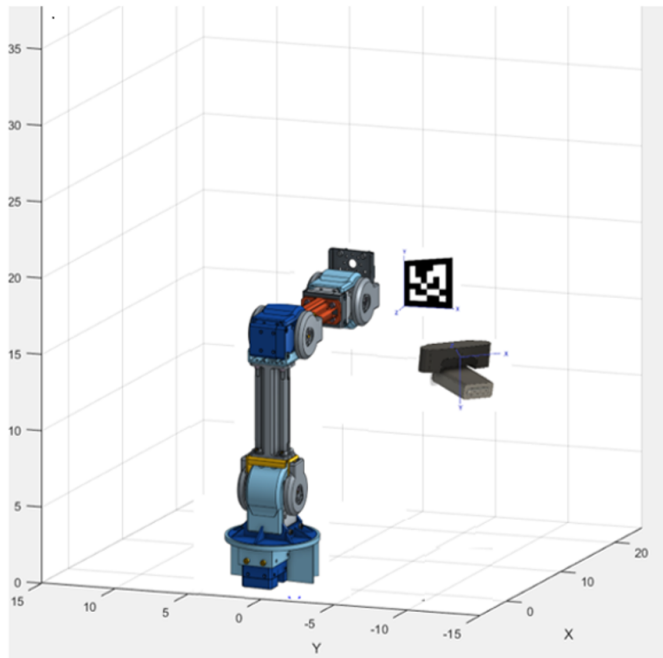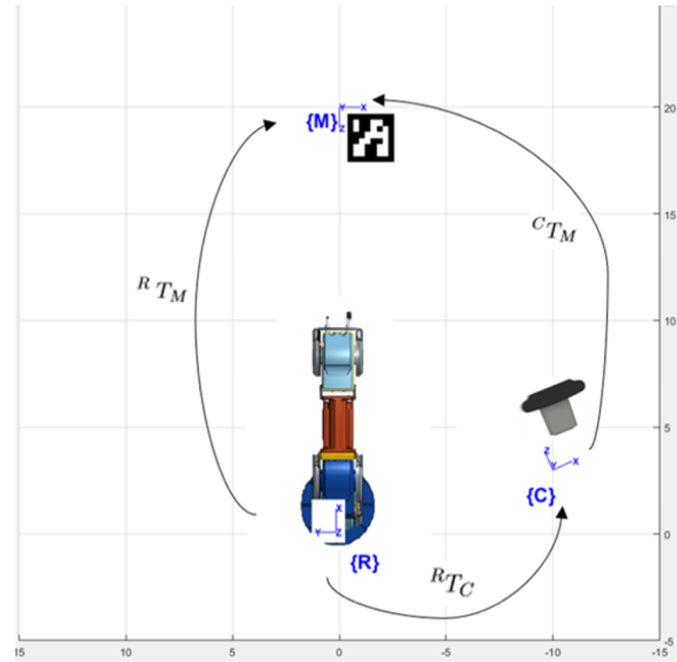
Figure 2.9: System setup to estimate camera pose: **(A)** isometric view and **(B)** top view.

2.2.6   Robot operation

A program on the RPi4 runs a ROS node that uses the information from the mobile App and uses the controller node of the manipulator robot to move it to the user-specified location. After performing the calibration routines, the system is ready to operate. The manipulator robot control program moves the robot to an initial position and waits for a user-specified coordinate to be available on the ROS topic where the mobile App publishes coordinates.

When the App starts, it immediately tries to communicate with the microcomputer. Once the communication is established, the touchscreen of the tablet device running the App will show the streaming video from the camera located next to the robot, capturing the images from the ICPT (see Figure 2.10A). With the detected 2D reference marker's information, the App will wait for the user to tap on the display of the TCT. The moment a new user-specified coordinate is received, using a sequence of events, the robot control program: (a) moves the robot manipulator to the desired location, just over the specified coordinate on the surface plane of the ICPT; (b) performs a tapping action that consists of moving slightly toward the ICPT until a contact occurs; and (c) returns to the initial position and waits for any new coordinates to be made available. This robot control program reads and responds to only one user-specified coordinate at a time and ignores any newly sent user coordinates while performing the sequence of operations for a previously received coordinate.

The complete system setup created for this proof-of-concept (see Figure 2.10B) uses a Microsoft Surface Pro 4 computer as the ICPT, running an application that mimics the functions of a dialysis machine instrument control panel.
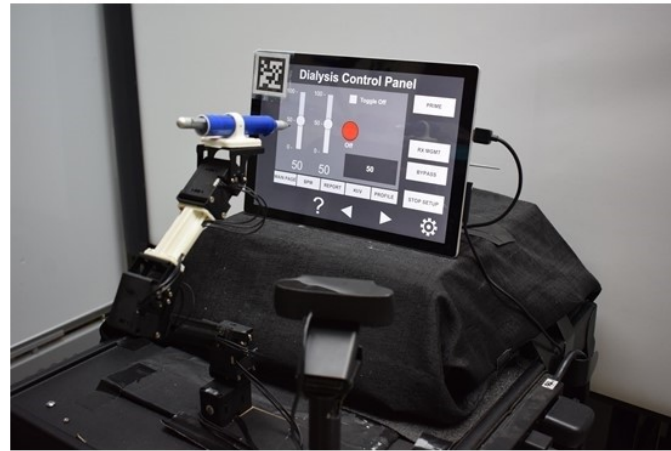
Figure 2.10: **(A)** HRI interface and **(B)** complete robot system setup.

2.3   System evaluation

An experimental study was conducted with participants to evaluate the performance and usability of the proposed system. The study was conducted with two groups of users, referred to hereafter as the in-person and remote groups. In the in-person group, 17 participants performed the experiment in a room adjacent to the room housing the robot, camera, and ICPT monitor. Alternatively, in the remote group, 16 participants performed the experiment from a remote location via the internet. See http://engineering.nyu.edu/mechatronics/videos/mhrifordialysis.html for a video illustrating a user interacting with the prototype to complete a set of tasks. Before performing the experiment, participants in both groups were briefed individually on the purpose of the experiment, what it entails, and how the interactions take place. They were informed that when the "Ready" prompt is shown on the TCT, the user can issue a command to the robot and when the "Busy" prompt is shown, it means that the robot is executing a task and will not accept any user command until the task is completed. No pretrial was conducted and each participant performed the experiment for only one time. This was done to ensure that the participants did not have any prior knowledge about the capabilities and the overall responsiveness of the system.

The participants who performed the experiment in-person were asked to use an Android tablet device with a touchscreen and interact with its screen using a stylus. During the experiment, the tablet device was connected to the same dedicated wireless network that the robot was connected to, and each user performed the experiment by staying in the same location in the room.

To test whether controlling the robot from a remote location has any influence over the system usability, system performance, and the task load of the user, an

online study was conducted wherein the participants were asked to command the robot by assuming control of a computer connected to the dedicated wireless network shared by the robot. The participants were briefed similarly to those in the in-person experiment, and no pretrial was conducted for this group either. The only major difference between the two groups was that the remote group of participants interacted with the video feed using a mouse pointer on their computer, whereas participants in the in-person group interacted with a tablet device using a stylus to issue commands to the robot.

During the experiment, the participants were asked to read a set of instructions on a PDF document and perform the experiment accordingly. The PDF instruction document listed six numbered tasks and an accompanying annotated image of the user interface (see Figure 2.11), where the six tasks correspond to six different interactions that the users needed to perform. These tasks were designed to mimic a set of user interactions that a healthcare worker typically performs on a dialysis machine interface. The details of the interactions are as follows.

(1) Press the red ON/OFF button.

(2) Change the value of the left slider to '0' and the value of the right slider to '100'.

(3) Press the toggle button.

(4) Increase/decrease the value displayed in the gray box using the arrow buttons.

(5) Select the RX MGMT button.

(6) Return to the Main Display using the MAIN PAGE button.
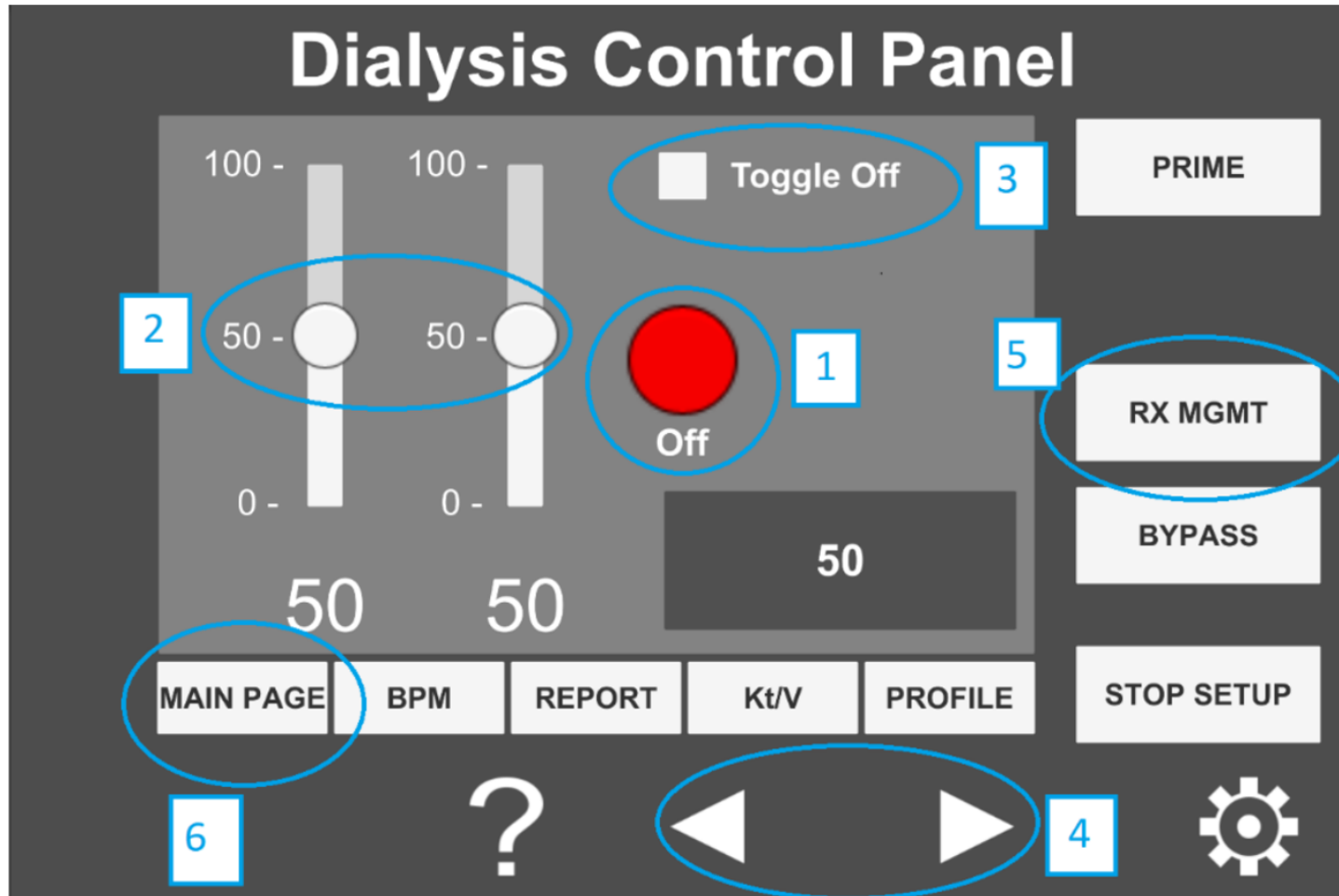
Figure 2.11: Annotated image of the user interface hosted on the tablet computer.

After the participants performed the six tasks, they were asked to respond to two questionnaires that assessed their experience for qualitative evaluation. The first part of the evaluation required the participant to respond to the NASA-Task Load indeX (NASA-TLX) [68] to assess the workload experienced by the participants while using the system. The NASA-TLX is used to rate the perceived workload of an individual while performing a task. It is divided into six categories that include physical workload, mental workload, temporal workload, effort, frustration, and performance. In this study, the Raw TLX (RTLX) assessment was performed in which the TLX scores are unweighted and the overall load of the task is calculated as the average score of the six categories in the NASA-TLX. In the second part of the evaluation, the participants were asked to express their level of agreement on a System Usability Scale (SUS) [69] questionnaire. The questionnaire consists of the following five positive and five negative statements with responses on a 5-point scale (1: strongly agree and 5: strongly disagree).

(1) I think that I would like to use this system frequently.

(2) I found the system unnecessarily complex.

(3) I thought the system was easy to use.

(4) I think that I would need the support of a technical person to be able to use this system.

(5) I found the various functions in this system to be well integrated.

(6) I thought there was too much inconsistency in this system.

(7) I imagine that most people would learn to use this system very quickly.

(8) I found the system to be very cumbersome to use.

(9) I felt very confident using the system.

(10) I needed to learn a lot of things before I could get going with this system.

The participants were provided Uniform Resource Locators (URL) to the NASA-RTLX and the SUS questionnaires and were asked to complete them on the spot immediately after completing the six-step interactive tasks provided above. The questionnaires were kept anonymous and no personal information was asked from the participants except their age group and their gender.

2.4 Results

The performance and the user experience of the proposed system was evaluated by conducting a study with 33 participants, of whom 29 participants were either engineering students or professionals working in a STEM-related field and the remaining four were medical professionals. A majority of the participants (72.73%) had operated or programmed a robotic system while the rest 27.27% had neither operated nor programmed a robot prior to their participation in the study. Note that the four medical professionals were part of the remote group and only one of them reported to have programmed or operated a robotic system previously. Furthermore, qualitative data obtained from the SUS questionnaire contained two outliers and one participant from the remote group did not complete the NASA-RTLX self-assessment. Thus, the data obtained from these three participants was not used for system evaluation and a total of 30 participants' data, 15 from each group, was used for the results reported below.

2.4.1 System performance

First, the performance of the system was evaluated by validating the accuracy with which a user is able to select and interact with desired points on the ICPT monitor using the proposed HRI interface on the TCT. We considered five reference points on the ICPT. These points were located at the center $(P_1(u, v))$ and near the four corners $(P_i(u, v), i = 2, \ldots, 5)$ of the screen. The experiment was conducted 50 times by a single user for each of the five reference points. The user input when interacting with TCT was recorded as pixel coordinates along the u and v axes and referred to as the *commanded* value. The point at which the robot interacted with the ICPT in response to the commanded value is referred to as the *measured* value and it was also stored as pixel coordinates along the u and v axes. The pixel coordinates for the commanded values were scaled up to the screen resolution of the ICPT so that a direct comparison with the measured values could be made. The performance of the AR interface was evaluated by calculating the absolute difference between the commanded and measured values for each interaction. Then the average absolute error was calculated for both the u and v coordinates. This was done for all five reference points and the results are shown in Table 2.2. The results indicate that for all five reference points, the highest average absolute error was less than 18.54 pixels for the u coordinate and 26.98 pixels for the v coordinate. Given that the resolution of the screen used for the ICPT is 2736 × 1824 pixels, with a diagonal screen size of 12.3 inches (312.42 mm), the pixel-to-length ratio was found to be 10.5 pixels/mm. Thus, the maximum average absolute error was 2.56 mm in the v coordinate of the fifth reference point $P_5(u, v)$. It is important to note that there was a button located at each of the five reference points and all 50 tests conducted on each button were successful, i.e., the button was successfully

pressed each time. The diameter of the buttons is 90 pixels which is approximately equal to 8.6 mm. This particular size of buttons is chosen because it is considerably smaller than all interactive elements on the touchscreen and the touchpad of a dialysis machine and therefore proves to be a reliable indicator of the performance of the system.

Table 2.2: Performance test results

Values in pixels	$P_1(u, v)$	$P_2(u, v)$	$P_3(u, v)$	$P_4(u, v)$	$P_5(u, v)$
Ideal	$(1368, 912)$	$(568, 1612)$	$(2168, 1612)$	$(2168, 212)$	$(568, 212)$
Commanded (average)	$(1367.4, 912.1)$	$(569.7, 1601.6)$	$(2166.9, 1612.1)$	$(2163.2, 211.4)$	$(595.9, 215.9)$
Measured (average)	$(1385.9, 905.7)$	$(565.8, 1596.9)$	$(2162.6, 1598.0)$	$(2173.6, 215.9)$	$(599.8, 189.1)$

The time taken by the robot to complete an interaction is determined by the task time programmed for the robot. In experimentation, it is measured as the difference between the time when the robot receives a command and the time when the robot returns to its home position after performing the interaction. The robot took 12.036 s to complete an interaction, without any significant difference in the times spent for different interactions. Next, the time it takes for a user to complete an interaction on the *ad hoc* wireless network is calculated as the difference between the time when the command is sent by the TCT and the time when the user receives the "Ready" prompt again on the TCT. For each of the following three scenarios, 15 tests were performed to measure the user interaction completion time.

(1) The user holding the TCT and the robot are in the same room.

(2) The user holding the TCT and the robot are in different but adjacent rooms.

(3) The user holding the TCT is at the maximum working distance from the wireless router (27 meters), with multiple rooms in between the user and the robot.

In all three scenarios, the average time to complete the user interaction showed no significant difference and was found to be 12.077 s. Finally, in the last time measurement experiment, we sought to determine the user interaction time when performing interactions with the robot over the internet. With a user located at a distance of approximately 1.5 mi from the robot, the average interaction time for 15 tests was obtained to be 12.56 s. Note that while the task completion time for the robot remains constant, the task completion time for the user and the maximum allowable interaction distance from the robot can change depending on user location and Wi-Fi signal strength, respectively.

2.4.2   User experience

While the results obtained using the system performance test validated the utility of the proposed system from an accuracy and precision point of view, it is important to consider the overall user experience when the participants operate the system. Thus, three different methods were used to perform the qualitative analysis of the user experience. The first method involved measuring the task load of the experiment using the NASA-RTLX self-assessment. This was followed by administering a system usability questionnaire, and finally the verbal/written feedback given by the participants was reviewed.

2.4.2.1   Workload

The assessment of the workload was performed by analyzing the results obtained from the NASA-RTLX self-assessment for the in-person and the remote participant groups. All categories were scored on a scale of 0–100 and the overall score for each participant was computed as a mean of the score for the six categories.

The score for each category was averaged and these calculations were used to compute the mean overall workload for both groups. Since there was no overlap between both groups, and therefore both samples are independent, a Welch's unequal variances two-tailed t-test was performed on the individual categories of the NASA-RTLX scores from both groups and the tests yielded $p > 0.05$ for responses of in-person *versus* remote experimenters. Thus, it was concluded that there was no statistically significant difference between the task loads experienced by the participants in the two groups. The combined average task load for both groups is computed and reported in Figure 2.12A.

The collective task load values for two groups (Figure 2.12A) and the raw data indicate that the frustration score was the highest for the two groups. The high frustration value can be attributed to the downtime that the participants experienced during the "Busy" phase of the robot movement, when the participants could not issue new commands to the robot. When we consider this factor with the slow speed at which the robot moves, it is plausible that the frustration value would increase as a result. Upon further examination of the raw task load data, it was observed that among the six categories, the effort scores exhibited a relatively high inter-group difference (in-person effort = 11.67, remote effort = 5.8). The difference between the effort values of both groups can be explained as a result of the type of interaction method with the robot. Participants in the remote group issued commands to the robot via a mouse pointer on a computer. This gave them very fine control with pixel perfect accuracy and a large screen size that definitely helped in the experiment. On the other hand, participants in the in-person group were asked to use a tablet device and a stylus to interact with the video-feed. The stylus requires extra pressure to be applied on the tablet computer screen to register

a touch input and the smaller screen size required the users to pay more attention to where they were interacting with the screen of the tablet device.

2.4.2.2   System usability

To gain an insight into the user experience of the participants, they were asked to complete a system usability questionnaire using a 5-point scale (1: strongly agree and 5: strongly disagree). The participants' individual responses were subjected to an unequal variances two-tailed t-test and the responses for the in-person group were compared with those of the remote group. Out of the 10 questions on the SUS questionnaire, three questions [(1), (5), and (6)] showed a statistically significant difference with $p = 0.03$, $p = 0.03$, and $p = 0.001$, respectively. Upon close examination of the data, two in-person group participants' responses were identified as outliers due to the large distance between their responses and the mean response for questions [(1), (3), (5), (7), and (9)]. Upon removing the outliers from the data, an unequal variances two-tailed t-test was performed again on the responses from the remaining 15 participants in the in-person and 15 participants in the remote groups. The results are shown in Figures 2.12B,C.
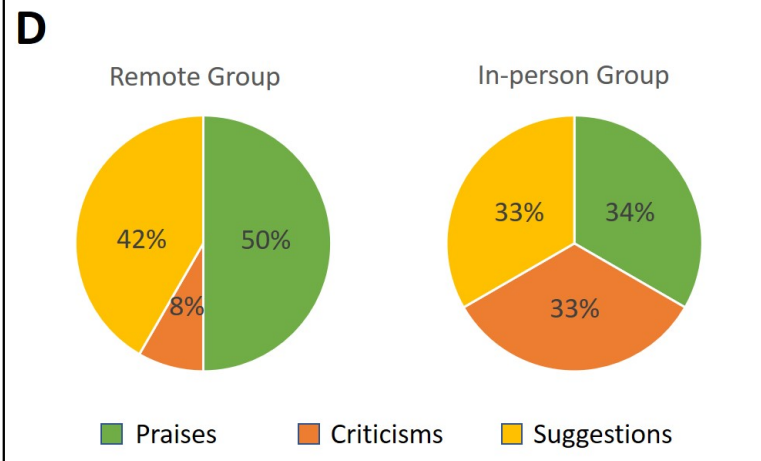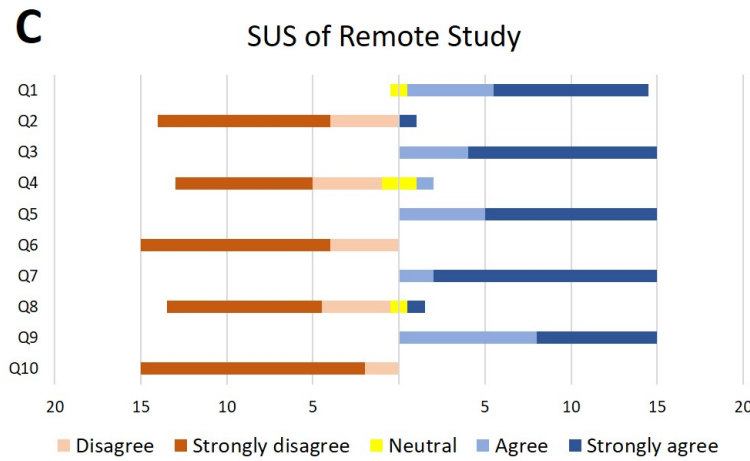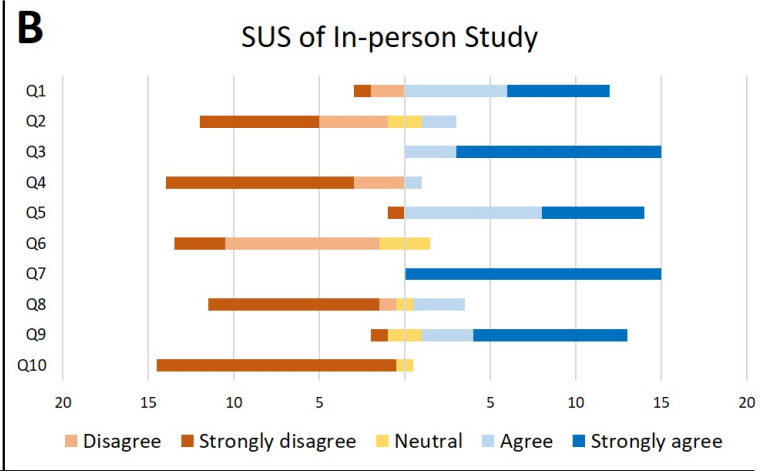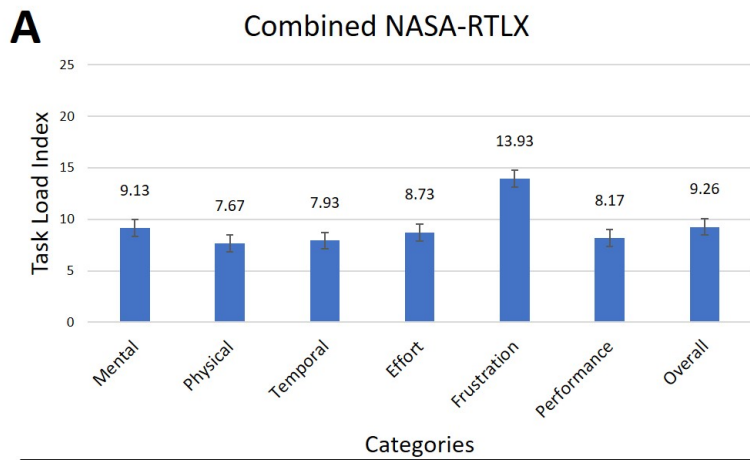
Figure 2.12: User study results (A) NASA-RTLX, (B) SUS in-person study, (C) SUS remote study, and (D) user opinion.

Out of the 10 questions, only question (6) showed a statistically significant difference with $p = 0.001$. Although all participants in the remote group disagreed or strongly disagreed that there was too much inconsistency in the system, some participants from the in-person group had neutral responses on this question. The neutral responses can be interpreted as participant reservations on the responsiveness of the tablet device when interacting with it using a stylus. Since the stylus used in this experiment had a relatively large tip, it is possible that some participants found inconsistencies when interacting with the tablet device if they did not pay close attention to where they touched the screen. This also explains why the participants in the remote group did not find any inconsistencies despite controlling the robot from a remote location. Since remote participants were using a mouse pointer on a comparatively larger screen (using a laptop or desktop computer), they could direct the robot more precisely, therefore reducing the human input error. A viable solution that would alleviate the problems faced by the in-person group would be to use a tablet device with a larger screen size, and/or use a stylus with a finer tip.

2.4.3 User comments

From the participants who tested the prototype, remotely and locally, we obtained different insights about their experiences interacting with the system by reviewing their comments and feedback. A total of 33 individuals participated in this study, out of which 21 provided comments and suggestions about their experience in controlling the manipulator. Some of the comments praised the system as evidenced by the use of terms such as "helpful," "efficient," "easy-to-use," "pretty good," "requires very little experience," among others. Although several other participants did not express negatively biased comments, they expressed

some reservation with the speed of the robot in executing the received commands, e.g., "the time taken by the robot to execute the command slows the process down." There were also criticisms from users who tested the prototype from a remote computer and on-site with a mobile device regarding the smoothness of the robot movements and the camera image shown in the HRI interface. Specifically, a participant who tested the prototype in-person using a tablet, suggested making the robot "more robust" and another participant who used the robot from a remote computer, advised "make the system more accurate and more stable [. . .] decrease the skew in the image from the camera."

Since this study proposes a solution to be used by heathcare workers, we also reached out to doctors who were willing to test the prototype and provide a review based on their experience working during the COVID-19 emergency. A total of four doctors remotely interacted with the proposed system and provided their feedback which included suggestions, criticisms, and compliments about the system and its utility as a viable solution for the control of dialysis machines during a pandemic. For example, one of the participant doctors, who used the prototype from a computer outside the United States, praised the system by commenting on its ease of understanding, use, sensitivity, absence of errors, ability to avoid contact with patient, etc. Another participant doctor offered insight into how this solution is perceived from a medical perspective, i.e., "interesting" and "of enormous use, especially when necessary to avoid physical contact." He also advised to improve the precision of robot because sometimes "it was necessary to select the same task until it was completed successfully." Additionally, another medical professional expressed his interest in how this system would perform in a real situation. This doctor provided a verbal review by stating that the system works very well but it

will require testing on a real dialysis machine to see how it controls it.

We found that the difference in these reviewer experiences is partially explained by the variations of internet connection speeds available on each participant's respective location (when controlling the system from a remote computer, off-site). In some cases, this variable added delay to video streaming, which did not let the users monitor how the robot was performing the tasks.

Figure 2.12D presents the percentages of each type of comments provided by the participants using the proposed system. We categorized the comments into "praise", "suggestion", and "criticism" categories. Positively biased comments were categorized as praises and accounted for 34% of the commenters from the in-person group and 50% of the commenters from the remote group. We categorized as criticisms the comments that identified a shortcoming without providing a recommendation. These accounted for 33% of the commenters from the in-person group and 8% of the commenters from the remote group. Finally, comments that provided recommendations to improve the system were categorized as suggestions and accounted for 33% of the commenters from the in-person group and 42% of the commenters from the remote group. It is seen that most participants were affected enough by their participation in the study to leave meaningful comments. Moreover, a sizable portion of participants was satisfied enough to praise their experience in writing. We took these praises to confirm our arguments for integrating mobile hardware and software as an effective way to interact with medical machines remotely using robot systems, such as the one proposed in this study. Many of the praises expressed the satisfaction of completing a set of tasks remotely, either using a "click" on a computer screen or tapping a location on a mobile device screen.

On the other hand, criticisms gave us areas of opportunity on which we can

focus to improve our prototype to deliver greater satisfaction in the use of HRI interfaces to control robots remotely and to meet the expectations of the system performance while executing a task. Observations regarding the smoothness of the robot movements and precision allow us to understand better how a system of this nature is perceived. Even if the users complete a set of tasks successfully, the speed while performing this task or lack of smoothness on the manipulator movements creates some distress. On the other hand, criticism about the skew of the camera view confirms that there is also some level of discomfort when a user perceives a distorted perspective of a surface (touch screen) with which interaction is required. While many of these suggestions for improvements will be considered in developing and testing future prototypes, below we offer one improvement to render a distortion-free perception of the ICPT on the TCT.

2.4.4 Suggested improvement based on user tests

On the SUS questionnaire and in the comment section, several participants provided written (and verbal) feedback concerning the skewed perception of the camera video-feed. In response, we have explored the potential of including an additional feature to the hybrid approach of this study, which uses a single reference marker, to correct and improve the video-feed displayed on the TCT interface. Specifically, as previously, when the users run the mobile UI App on the tablet device, the raw live-stream of the ICPT is displayed on the TCT with a distorted perception. Next, the App prompts the user to touch (from the mobile device) or click (from the remote computer) the four corners on the video-feed of the surface plane of action of the ICPT in a clockwise manner, starting from the corner closest to the fiducial marker (see Figure 2.13A). These user-selected pixel

coordinates, corresponding to the corners of the ICPT, are used to get a perspective transformation matrix and map the identified ICPT plane to fit the screen of the TCT by performing a perspective correction. This correction technique allows the user to be presented with a distortion-corrected view of the raw ICPT video-feed in the HRI interface (see Figure 2.13B). When the user interacts with the corrected image displayed on the TCT, the inverse of the perspective transformation matrix computed above can be used to map the pixel coordinates of the user interaction on the TCT to the original perspective view captured by the camera, allowing the application to work without any additional modifications. Figure 2.13 illustrates that it is feasible to implement such a perspective correction approach, however a complete set of user-tests with this improved approach is beyond the scope of this study and will be considered in a broader study with the two alternative approaches suggested in subsection 2.2.4.

2.5   Conclusion

This chapter presents a system for remote control of a dialysis machine with mobile HRI as part of the COVID-19 emergency response. The proposed approach utilizes the capabilities of a smartphone/tablet device as a mode of interaction with a 4-DOF robot and explores the possibility of manipulating the robot to remotely interact with the instrument panel of a dialysis machine. This allows medical professionals to maintain social distancing when treating dialysis patients, preventing potential exposure to pathogens for both the healthcare staff and the patients. Such a system will also help lower the use of PPE by doctors and nurses
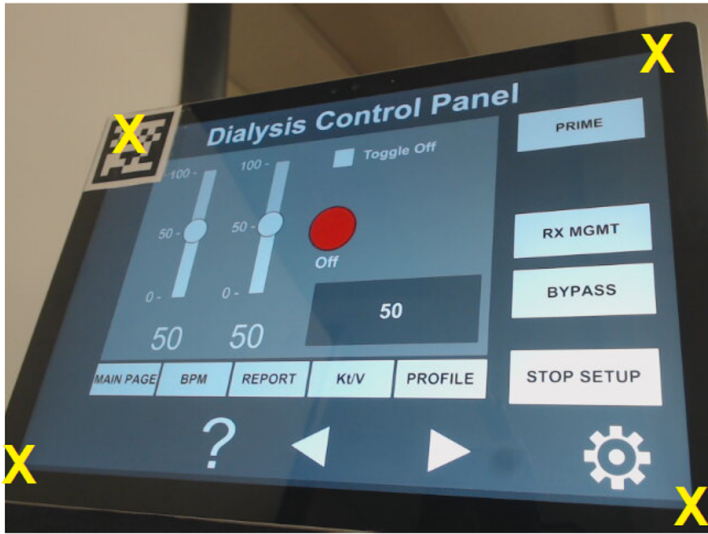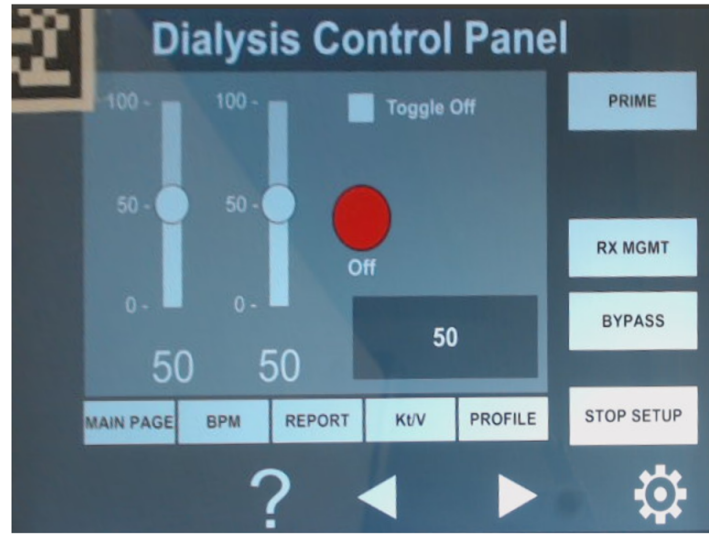
Figure 2.13: **(A)** Raw image with corners selected by the user and **(B)** image with distortion corrected.

while performing routine, simple procedures that could be performed by a robot. To evaluate the proposed system, its performance, and the user experience, a user study was conducted in which participants remotely issued commands to a robot via a tablet device or a computer. The participants received a live streaming video of a mock dialysis machine ICPT that allowed them to command the robot to manipulate the UI elements of the ICPT by touching those elements on the video feed on the TCT. Results of the study show that the participants were able to remotely access the UI elements of the ICPT and complete the tasks successfully. Based on the feedback received on the SUS questionnaire from the participants, an improvement to the proposed HRI interface was suggested and implemented which corrected the perspective distortion of the live stream of ICPT and allowed the user to interact with the corrected image for a more intuitive experience. Overall, the live-streaming video of the instrument panel provides a very natural and intuitive mode of interaction for the user and does not require prior experience in programming or operating robots. Most importantly, there is no need to develop a custom UI for the TCT since the user directly interacts with the video feed from the ICPT. This allows the proposed system to work with any touchscreen and the development of custom TCT interfaces that only work with their corresponding ICPT is not required. Finally, the proposed approach can be deployed very rapidly and requires minimum preparation work in case of an emergency, therefore saving valuable time and resources that can be directed elsewhere.

Chapter 3

Range of motion assessment using a digital voice assistant

3.1 Introduction

With the advent of smart devices, high-speed internet, and digital services, many hospitals and clinics have embraced telemedicine and digital healthcare as a viable service to the public. In fact, in the context of the COVID-19 pandemic, a large majority of people in the US have started to receive virtual clinical services [70]. Currently, such services entail teleconferencing software, making the interactions a virtual analog of an in-person visit to a doctor's office. However, other aspects of telehealth are yet to make their transition to the digital domain. One such aspect is data-driven telemedicine and therapy compliance monitoring in elderly care. A paradigm shift is waiting to unfold, since the technology to make it happen is already available, albeit used for other purposes. DVAs are one example of such a technology. The last decade has seen a growth in the adoption of DVAs,

with 46% of the US population already using some kind of DVAs as of 2017 [71]. Recently, DVAs have started to ship with built-in cameras, which greatly increases the potential of these devices. Yet, today the DVAs are used in a limited capacity, e.g., video conferencing, information retrieval, smart home control, etc.

As indicated previously, in recent years, discussions about digital healthcare and telehealth have come to the forefront [72] with an emphasis on geriatric health [73]. Even as the use of DVAs as a modality for telehealth and telemedicine remains a nascent concept, efforts are afoot to explore their utility and effectiveness for synchronous and asynchronous healthcare delivery [74], in remote healthcare monitoring [75], and as a conversational agent for the elderly [76]. The recent integration of built-in cameras in the DVAs offers yet another opportunity to embed novel modalities for therapy compliance into the DVA device ecosystem. Thus, this work proposes a DVA application for the ROM measurement, which is traditionally performed with sensor-based [77] or optical [78] methods. We demonstrate the use of the built-in camera of a DVA device for ROM monitoring. This proof-of-concept system can be explored further to develop more effective telehealth solutions involving DVA devices, e.g., for remote monitoring of compliance with physical therapy.

The chapter is organized as follows. Section 3.2 elaborates on the design and development of the system used for ROM tracking. Section 3.3 discusses the results based on the preliminary work and addresses the benefits and shortcomings of the approach used. Finally, section 3.4 provides concluding remarks.

3.2   Design and development

Our prototype is based on the *Google Nest Hub Max* DVA device. A computer acts as a local server and uses the Nest API to access the live stream of the device's camera through a Real-Time Streaming Protocol. An *if-this-then-that* applet accesses the Google Assistant in the DVA to respond to a user's voice commands. When the user issues a command, such as, "*OK Google, start ROM app*", the applet sends a GET request using the Flask web framework to the server. Once the command is acknowledged by the server, it starts receiving the live stream from the DVA camera, and the ROM measurement application starts on the server. A confirmation is sent to the user as a voice response by the Google Assistant and the user can start performing the exercise. Figure 3.1 shows the overview of the proposed system and various alternative ways in which an overall system can be constructed.

We consider a 3D pose estimation problem with a subject standing at a fixed distance from the DVA camera. It is assumed that the subject's frontal plane (see Figure 3.2) is parallel to the camera image plane (see Figure 3.1). To determine the 3D pose, 33 image coordinates (i.e., landmarks) in $\mathbb{R}^2$ corresponding to the body joints are detected using the MediaPipe library that utilizes the BlazePose [79] pose detection model. To study the upper extremity ROM, we use eight of the 33 landmarks from the MediaPipe. Although the model provides the landmarks in $\mathbb{R}^3$, the z-coordinate (corresponding to depth) is discarded as it has a high variance in error when repeated tests are carried out for the same pose.
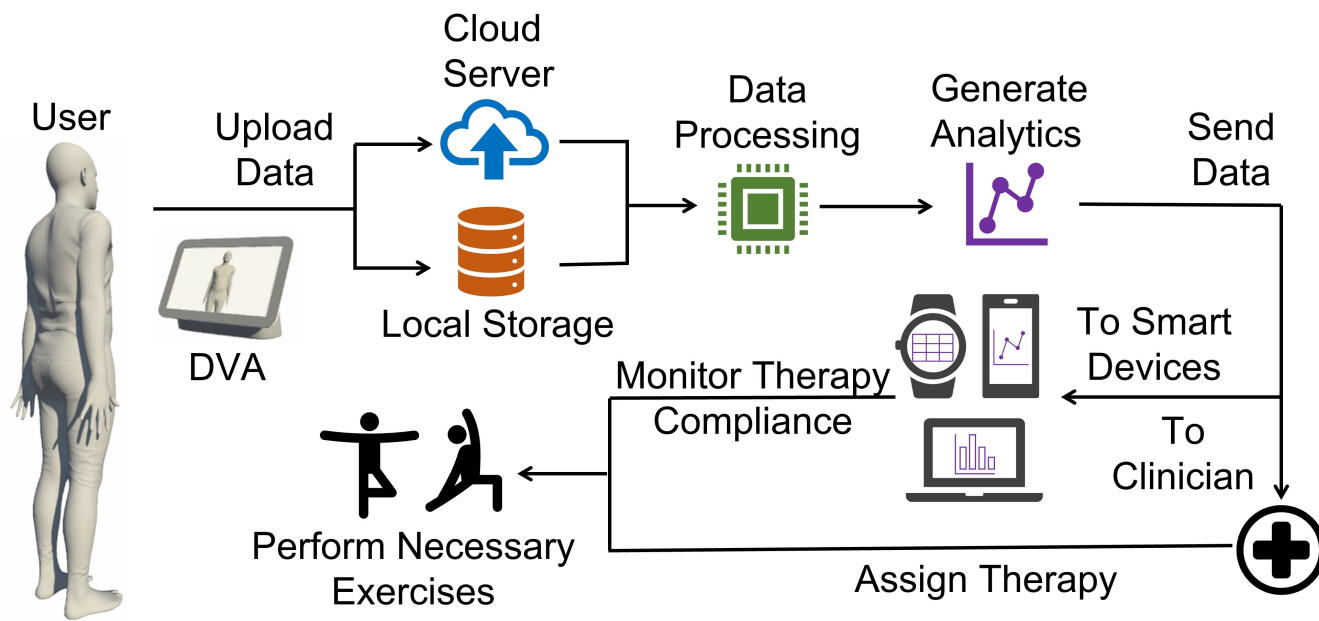
Figure 3.1: Overview of the proposed system.
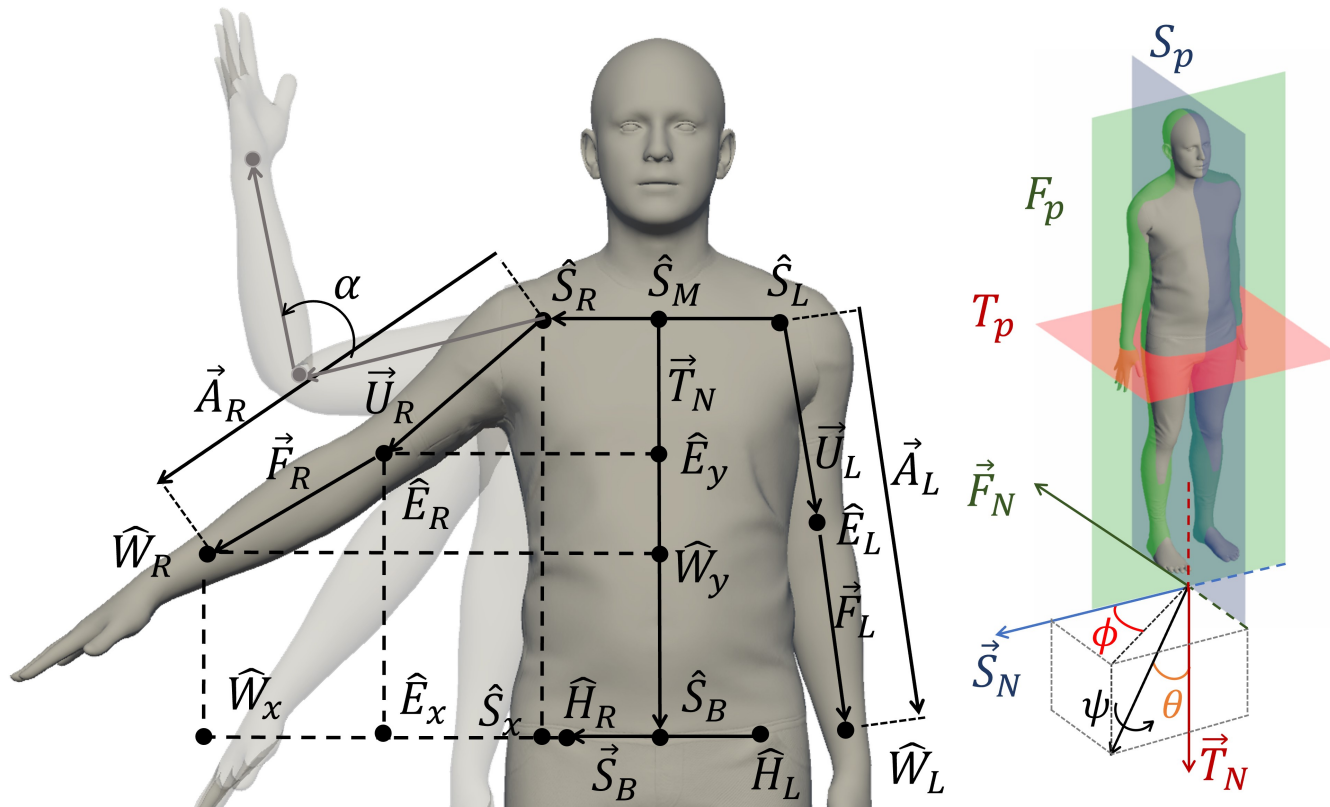
Figure 3.2: Joint positions obtained from the MediaPipe and the vectors used to compute the joint angles: $\hat{}$ and $\vec{}$ denote points and vectors, respectively.

3.2.1 Shoulder ROM measurement using camera image

Glenohumeral joint motion is modeled using spherical coordinates [80] where the joint is at the center. We consider that the wrist skirts the periphery of a virtual sphere. The three angles of interest are the thoracohumeral angle, i.e., the shoulder elevation angle θ, the clavicular-humeral angle projected on the transverse plane, i.e., the shoulder plane angle ϕ, and the humeral rotation around the long axis, i.e., the internal-external rotation angle ψ. The notation used and measurements performed for the right arm in this work can be analogously extended to the left arm. Figure 3.2 depicts the angles and their relation to the cardinal planes.

3.2.1.1 Calibration Step

First, the user stands in front of the DVA camera with arms down to the side in a "rest pose". The system calculates the length of the trunk vector $\vec{T}_N \triangleq \hat{S}_B - \hat{S}_M$, the maximum length of the right arm vector $\vec{A}_R \triangleq \hat{W}_R - \hat{S}_R$, and the maximum length of the right forearm vector $\vec{F}_R \triangleq \hat{W}_R - \hat{E}_R$ in pixel space. These calculations are performed for $N = 50$ samples and the length averages $T_{N_{\text{avg}}}$, $A_{R_{\text{avg}}}$, and $F_{R_{\text{avg}}}$ are computed to mitigate any errors due to noise. The angle θ (or ψ) calculation is performed using the ratio of the length of the arm (or forearm) to the length of the trunk for the rest pose in which it is maximum and denoted as the *rest ratio* a_r (or f_r) shown below

$$a_r \;=\; \frac{A_{R_{\text{avg}}}}{T_{N_{\text{avg}}}} \quad \text{and} \quad f_r = \frac{F_{R_{\text{avg}}}}{T_{N_{\text{avg}}}}. \tag{3.1}$$

This step helps overcome any changes in the apparent arm (or forearm) length due

to a shift in perspective. Next, for calculating ϕ, we compute the *stretch ratio* a_s with the user arm stretched to the side, yielding $a_s = a_r|_{\theta=90°,\phi=0°}$.

3.2.1.2 Calculation step

We now determine the shoulder and elbow angles corresponding to various ROM exercises. We begin by noting that for two arbitrary vectors $\vec{v}_i = \hat{y}_i - \hat{x}_i$, $i = 1, 2$, the point of projection $\hat{p}$ of the vector $\vec{v}_2$ on the vector $\vec{v}_1$ is given by

$$\hat{p} = \text{proj}_{\vec{v}_1} \vec{v}_2 + \hat{x}_1 = \left(\frac{\vec{v}_1 \cdot \vec{v}_2}{\|\vec{v}_1\|^2} \right) \vec{v}_1 + \hat{x}_1. \tag{3.2}$$

Using (3.2), we can now determine the following quantities:

(*i*) the vertical and horizontal projections of the elbow $\hat{E}_R$ on the base vector $\vec{S}_B$ and the trunk vector $\vec{T}_N$, denoted as $\hat{E}_x$ and $\hat{E}_y$, respectively;

(*ii*) the vertical and horizontal projections of the wrist $\hat{W}_R$ on $\vec{S}_B$ and $\vec{T}_N$, denoted as $\hat{W}_x$ and $\hat{W}_y$, respectively; and

(*iii*) the vertical projection of the shoulder $\hat{S}_R$ on $\vec{S}_B$, denoted as $\hat{S}_x$. See Figure 3.2 that shows the projections $\hat{E}_x$, $\hat{E}_y$, $\hat{W}_x$, $\hat{W}_y$, and $\hat{S}_x$ for the right arm.

It can be shown that for a 2D projection of a vector rotating inside a sphere with a radius equal to the length of the rotating vector, the change in the distance from the vertical projection point of a vector ($\hat{E}_y$ or $\hat{W}_y$) to $\hat{S}_M$ is directly proportional to $\cos(\theta)$. Thus, θ can be calculated for the shoulder abduction-adduction and flexion-extension exercises as shown in (3.3). With the user at a constant distance from the DVA camera and when $\theta = 90°$, ϕ can similarly be calculated as shown in

(3.3). For the shoulder internal-external rotation exercise, let the upper arm point downwards (i.e., $\theta = 0°$), and flex the elbow to $90°$ (i.e., elbow angle $\alpha = 90°$). Then the internal-external rotation (i.e., shoulder angle ψ) can be computed as shown in (3.3). Finally, for the elbow flexion-extension exercise, the arm is kept parallel to the camera image plane (i.e., $\phi = 0°$), and the elbow angle α, between the upper arm $\vec{U}_R$ and forearm $\vec{F}_R$ is calculated as shown in (3.3).

$$
\theta = \cos^{-1}\left(\frac{\hat{W}_y - \hat{S}_M}{T_{N_{\text{avg}}} a_r}\right), \; \phi = \cos^{-1}\left(\frac{\hat{W}_x - \hat{S}_x}{T_{N_{\text{avg}}} a_s}\right),
$$
$$
\psi = \cos^{-1}\left(\frac{\hat{W}_x - \hat{E}_x}{T_{N_{\text{avg}}} f_r}\right), \; \alpha = \cos^{-1}\left(\frac{\vec{U}_R \cdot \vec{F}_R}{\left|\vec{U}_R\right|\left|\vec{F}_R\right|}\right). \tag{3.3}
$$

3.2.2 Experiment design

A virtual environment is created using the Unity Engine (Unity Software Inc., San Francisco, CA, USA) to compare the proposed system with the synthetic ground truth data. The virtual environment comprises an SMPL-X [81] 3D human model standing in front of a virtual camera. The SMPL-X model undergoes various exercises for the right arm: (*i*) shoulder abduction-adduction ($\phi = 0°$, $\theta \in [0, 170]°$); (*ii*) shoulder flexion-extension ($\phi = 90°$, $\theta \in [0, 170]°$); (*iii*) shoulder plane angle ($\theta = 90°$, $\phi \in [0, 90]°$); (*iv*) shoulder internal-external rotation ($\theta = 0°, \alpha = 90°, \psi \in [-50, 50]°$); and (*v*) elbow flexion-extension ($\phi = 0, \alpha = [0, 130]°$). To further validate the system, the exercises are repeated by a neurologically intact individual with nominal upper extremity ROM while the data is collected and results are compared for the proposed approach *vs.* the Kinect. See Figure 3.3 for details.

For the ground truth data, measurements are timestamped and recorded in a
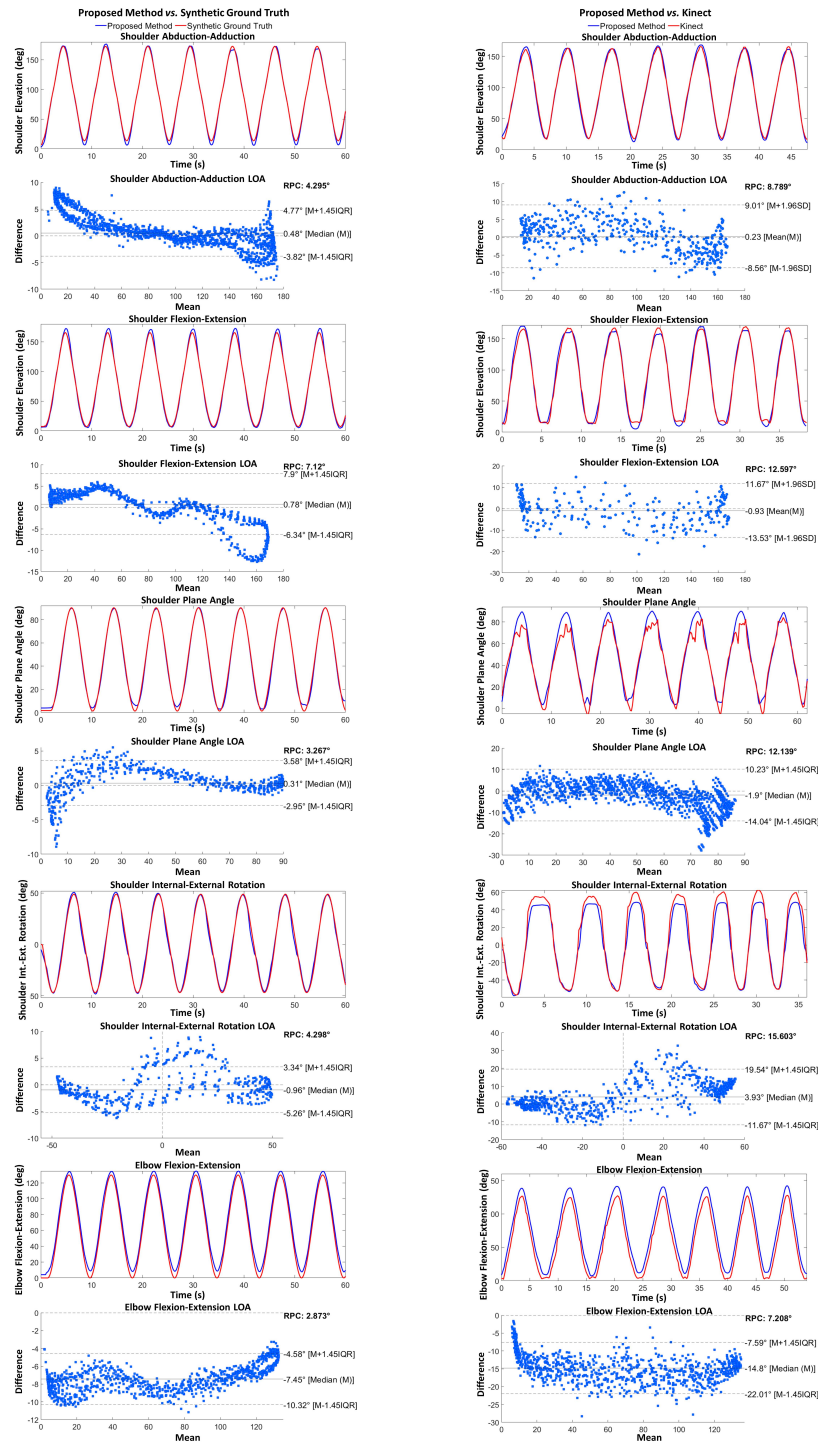
Figure 3.3: Comparison between the data from (a) the synthetic ground truth *vs.* the proposed method and (b) Kinect *vs.* the proposed method.

text file. Simultaneously, the video feed from the virtual camera is processed by the MediaPipe and the resulting 2D landmarks are also timestamped and recorded. Figure 3.3 shows the graphs of the computed ROM angles for the ground truth *vs.* the proposed method for various exercises. A similar data collection method is used when a user performs the exercises for comparing the proposed method *vs.* the Kinect-based measurements and these results are also shown in Figure 3.3. To mitigate noise effects, the data from the MediaPipe is filtered using a moving average filter with a window length of 8, which time-shifts the data. In post-processing, the data is time-aligned with the ground truth and Kinect data for analysis.

3.3   Results and discussion

The results from the proposed method are compared to the ground truth using the Bland-Altman test [82] to find the limits of agreement (LOA). Figure 3.3 and Table 3.1 show the LOA for various exercises. The LOA for exercise 2 is seen to be higher than for other exercises. This is explained by the arm motion out of the frontal plane for exercise 2, where the lack of depth information causes the proposed approach to be degraded especially near the extreme angles ($\theta = 0°$ and $\theta = 180°$). In exercise 3, a similar effect is seen at $\phi = 0°$, but since the angle never reaches $\phi = 180°$, the LOA is lower in this case. In exercise 4, since only the forearm moves out of the frontal plane, the lack of depth estimation does not degrade the corresponding ROM estimate excessively.

Figure 3.3 and Table 3.1 also show the LOA between the proposed method *vs.* Kinect. The LOA data has larger values for comparison with Kinect in contrast

Table 3.1: Limits of agreement for shoulder (S) and elbow (E) exercises.

No.	Exercise	*vs.* Synthetic data	*vs.* Kinect
1	S: abduction-adduction	$\pm 4.3°$	$\pm 8.8°$
2	S: flexion-extension	$\pm 7.1°$	$\pm 12.6°$
3	S: plane angle	$\pm 3.3°$	$\pm 12.1°$
4	S: internal-external rotation	$\pm 4.3°$	$\pm 15.6°$
5	E: flexion-extension	$\pm 2.9°$	$\pm 7.2°$

to the ground truth. One reason for this may be that the camera sensor suffers from noise. Next, for exercise 1, the large LOA against Kinect may arise from the discrepancy in the data for the low values of shoulder elevation angles, which are normally the resting position of the arm and may not be significant for one's ability to perform activities of daily living. For the larger shoulder elevation angles, the data from the proposed system closely follows the data from the Kinect. For exercise 2, the large value of LOA can be attributed to the discrepancy in the data for the low and high values of shoulder elevation angles, where the Kinect measurements may suffer due to only small changes in depth. For exercise 3, the large value of LOA can be attributed to data discrepancy for $\phi = 90°$ where the Kinect sensor suffers from occlusion of the elbow joint. For exercise 4, the large LOA value may arise from the moving average process used in the proposed method that reduces the slope of the motion trajectory as the arm switches from internal to external rotation. While this does not affect the peak values, a reduction in the moving average window size is found to improve the LOA. Finally, for exercise 5, the large LOA against Kinect may be caused by the pose prediction uncertainty from the MediaPipe and the Kinect.

3.4   Conclusion

This chapter introduced a ROM assessment system leveraging a DVA device equipped with an integrated camera. By analyzing the real-time video feed, this system tracks 2D joint landmarks to monitor ROM, focusing on shoulder and elbow movements. The proposed system performs analytical computations of joint angles, which were validated against synthetic ground truth and Kinect sensor measurements. This method not only offers a novel avenue for ensuring therapy adherence but also enables at-home monitoring, which can potentially enhance patient engagement and compliance. Furthermore, the use of DVA devices with built-in cameras presents opportunities for future research, including long-term ROM assessment and the expansion of telehealth applications focused on therapy compliance.

Chapter 4

Wireless earphone-based real-time monitoring of breathing exercises

4.1 Introduction

Deep breathing exercises have been shown to reduce conditions such as hypertension [83] and anxiety [84]. Moreover, several therapy interventions for medical conditions such as lymphedema [85], asthma [86, 87], and chronic obstructive pulmonary disease [88] rely on breathing exercises as a primary component. Thus, tracking breathing using a monitoring device can support therapy compliance.

Respiratory rate, blood carbon dioxide level, and the volume of air inhaled and exhaled by the lungs serve as vital indicators of health. In a clinical setting, these indicators are measured using devices such as spirometers and capnographs [89]. However, for general diagnosis and therapy compliance applications, metrics such as the respiratory rate are often measured manually by counting the number of chest expansions over a minute [89]. For such non-acute applications, several

devices are commercially available including respiratory thoracic belts and pulse oximeters with integrated respiratory rate monitoring [89]. A major drawback of these devices is that they are considered specialized medical hardware that may not be readily available at home. To address these issues, one viable alternative is the use of commodity hardware to measure acoustic biomarkers for monitoring therapy compliance.

Prior research has explored the effectiveness of acoustic data for administering therapies such as breathing training [90] and for tracking metrics such as respiratory rate [91, 92]. These studies rely on a user breathing directly into a microphone, which limits their application for at-home therapy compliance monitoring, especially when the therapy requires physical movement with breathing exercises. These studies also do not distinguish between nasal and oral breathing, which is important for some therapy interventions [85].

Recent advances in acoustic hardware design have led to the development and, consequently, the widespread adoption of wireless earphones that are capable of connecting to smart mobile devices. Almost all wireless earphones have a built-in microphone and several latest models also include IMUs and proximity sensors. These developments open new sensing modalities and research avenues for digital healthcare and telerehabilitation. For example, recent research has used wireless earphones for respiratory rate monitoring using audio [93] and the built-in IMU [94] sensors of the earphone. Yet other studies have combined prior approaches to integrate the audio signals and IMU data from the earphones for determining the respiration rate and breathing channel (nasal *vs.* oral breathing) [95, 96]. All of these studies utilize deep-learning models to detect respiratory rate and breathing channel. It is pertinent to note that the aforementioned studies do not provide

the dataset used to train their underlying neural network models, which makes it impossible to reproduce their results.

To accurately monitor breathing exercises using wireless earphones, this work creates a framework that has the potential for assessing a patient's compliance with an at-home therapy. The study builds upon [97] that developed an at-home lymphedema therapy compliance monitoring system. Along with physical activity, such therapies entail breathing exercises wherein a user performs nasal inhales and oral exhales, which warrant accurate tracking of breathing phases (inhale/exhale) and channels (nasal/oral). One such therapy is the optimal lymph flow (TOLF) which is a therapeutic exercise program designed for post-operative patients to help them reduce the risk of lymphedema. It includes both physical exercises, where the user is required to move their limbs, and breathing exercises, where the user inhales through the nose and exhales through the mouth. To monitor a patient's compliance with the TOLF program, it is important to accurately detect these exercises as a patient performs them so that their compliance and therapy with the program can be monitored. While solutions for human pose estimation are discussed in Chapter 3, this chapter specifically focuses on detecting breathing exercises. Two main contributions of this study are:

(*i*) a system for real-time detection of breathing phases and channels when the user performs breathing exercises and

(*ii*) an annotated breathing audio dataset, created using wireless earphones, to facilitate the reproducibility of results and enable further research.

The chapter is organized as follows. Section 4.2 elaborates on the creation of the dataset and the architecture and training of the system used for breath

classification. Section 4.3 discusses the results based on the preliminary work and addresses the benefits and shortcomings of the approach used. Finally, Section 4.4 provides concluding remarks.

4.2   Design and development

The proposed system is envisioned as an application (App) running on a smart mobile device. Using a pair of wireless earphones connected to the smart device, the system will receive audio signals in real-time and use them to infer the breathing pattern of the user. Specifically, the App will detect (i) the breathing channel, i.e., nasal or oral, and (ii) the breathing phase, i.e., inhale or exhale. To begin using the system, the user will launch the App on the smart device and wear the wireless earphones. Next, the App will acquire the breathing audio recorded through the wireless earphones and pass it through a neural network classifier that will infer the channel and phase of breathing. Additionally, the classifier will detect any pause between breathing and, thus, can be used to calculate metrics such as the total number of breaths, respiratory rate, and breath-phase duration. Below we provide details about a proof-of-concept system where the data-gathering process and the architecture of the neural network have been implemented on an Apple iPhone and a high-performance computing system, respectively. Porting the designed neural network to the iOS environment will enable the App-based system.

4.2.1   Audio dataset creation

As of this writing, datasets of breathing audio sounds recorded with wireless earphones are not publicly available. Moreover, a majority of the publicly available

breathing audio datasets are either used for detecting pathologies of the respiratory tract [98] or were recorded using contact microphones [99]. Finally, the data quality of many public datasets is too low due to signal corruption from noise, distortions, and audio artifacts, making the datasets unsuitable for our application. In response, we sought to create a breathing audio dataset using wireless earphones.

To capture the audio data, a pair of wireless AirPods connected to an iPhone were chosen due to their broad availability and popularity. Airpods operate under three modes, i.e., noise cancellation, transparency, and off. To eliminate any unwanted audio interference or amplification caused by the noise cancellation and transparency modes, the off mode was used for audio recording. The recordings were stored in the M4A audio file format and converted to the WAV format.

For the dataset collection, eight healthy individuals, aged between 23 and 34 years, were invited for two recording sessions. During each session, they were asked to record two-minute-long audio clips each for nasal and oral breathing. The recordings were done in a quiet room and each participant was given a demonstration of the exercises before the recording session. Participants who owned AirPods were encouraged to use them, the rest of the participants were provided with AirPods and new ear tips. The data was obtained with the participants' signed consent and the research was approved by the NYU Institutional Review Board (IRB-FY2020-4198).

To label the audio data, an initial dataset was recorded with two participants who performed two sets of two-minute-long nasal and oral breathing exercises. After being converted to WAV format, the audio files were manually labeled using the open-source application "Audacity". The labeling was performed by visualizing the audio files as spectrograms and listening to the audio using earphones. The

following categories were used for systematic labeling: pause (0), nose-inhale (1), nose-exhale (2), mouth-inhale (3), and mouth-exhale (4). These labels were assigned to the corresponding sections of audio, providing a structured representation of the breathing patterns. The labels were exported as a text file where each row contains the start time, end time, and label of the audio segment. Figure 4.1 shows an example of annotated spectrograms for nasal and oral breathing.

To simplify the labeling for the recordings of the remaining six participants, the labeling process was streamlined by training a binary convolutional neural network (CNN) model to distinguish between two classes: pause (0) and breath (~0). This model achieved an accuracy of 90%, which significantly expedited the labeling process. The pre-labeled files obtained through CNN underwent a round of scrutiny to rectify any minor mistakes made by the model during inference, and to provide the correct classes $(1, \ldots, 4)$, for the data from the breath class (~0), based on the channel and phase of breathing. This classifier was re-trained after the addition of each new participant's data until an accuracy of 96% was achieved. The architecture of this labeling classifier is identical to the channel classifier model discussed in the next subsection with the exception that the labeling classification head was modified to output two classes instead of three given by the channel classifier. The dataset is made publicly available at: https://shorturl.at/jlrKU.

4.2.2 Training the model

The next steps entailed preprocessing the audio data, tuning model hyperparameters, and training the neural network to distinguish between channels and phases of breathing.
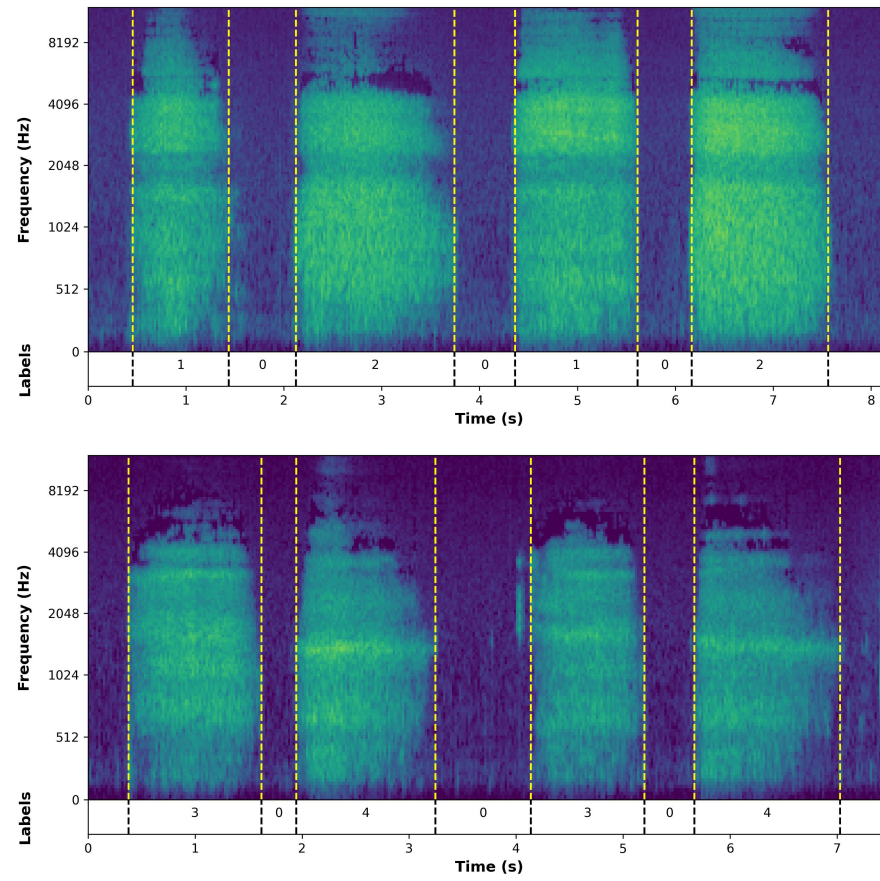
Figure 4.1: Audio spectrograms for (top) nasal and (bottom) oral breathing.

4.2.2.1 Audio preprocessing

The WAV audio clips underwent several preprocessing steps before being used to train the model. To begin with, the audio clips were resampled at 16 kHz to standardize the input data. Next, the audio clips were segmented and labeled into 500 ms long intervals with a frame stride of 250 ms to allow overlap between adjacent segments. The frame length of 500 ms was deemed to have enough information for model training without adding too much delay to the system during inference. To label each audio segment, the text file of labels for the corresponding audio clip was used. We employed a multilabeling approach wherein the audio segments were allowed to have more than one label. This is important because an audio segment that contains, e.g., a transition from a pause to a nasal inhale should be assigned labels for pause and nose-inhale. The labels were also one-hot encoded to prevent the model from learning the ordinal relationship between the categories.

To improve the generalizability of the model, the training dataset was augmented with noise taken from an audio clip that includes background noise from multiple people talking. The noise was uniformly sampled across a range of signal-to-noise ratios from 20 dB to 40 dB.

The segmented audio waveforms were converted to spectral features, i.e., mel-spectrograms and mel-frequency cepstral coefficients (MFCCs) of sizes 128×126 and 40×41, respectively. The mel-spectrograms were divided into 128 mel-filterbanks, with 2048 fast Fourier transform points, a window length of 2048, and a hop length of 64. To create the MFCC matrices, the number of MFCCs was chosen to be 40. The mel-spectrograms and MFCCs were generated using the TorchAudio [100] library from PyTorch.

4.2.2.2   Model training pipeline

The problem statement for model training was split into two tasks. The first task entailed classifying between pause, nasal breathing, and oral breathing. To do this, a CNN classifier was designed with three convolutional blocks having filter sizes of 8, 16, and 32, respectively. Each convolutional block's convolutional layer was followed by a batch normalization layer and a max pooling layer. Moreover, four fully connected layers were added after the last convolutional block. The neural network used a rectified linear unit as the activation function and each convolutional layer had a 3×3 kernel size. As noted in subsection 4.2.1, this model was also used to train the binary classifier for data labeling. The classifier's architecture is shown in Figure 4.2 and is denoted as 'channel classifier'.

Having passed through the channel classifier, the audio data was classified as either pause, nasal breathing, or oral breathing. If an audio segment was classified as a pause, the system displayed the result to the user and began processing the next audio segment. However, if the audio segment was classified as nasal or oral breathing, then the audio segment was passed to a second classifier that sought to determine whether it represented an inhale or exhale. This classifier had a single convolutional block with a filter size of 4 and a 3×3 kernel size. The convolutional block was followed by four fully connected layers. The architecture of the CNN is shown in Fig. 4.2 as 'phase classifier'.
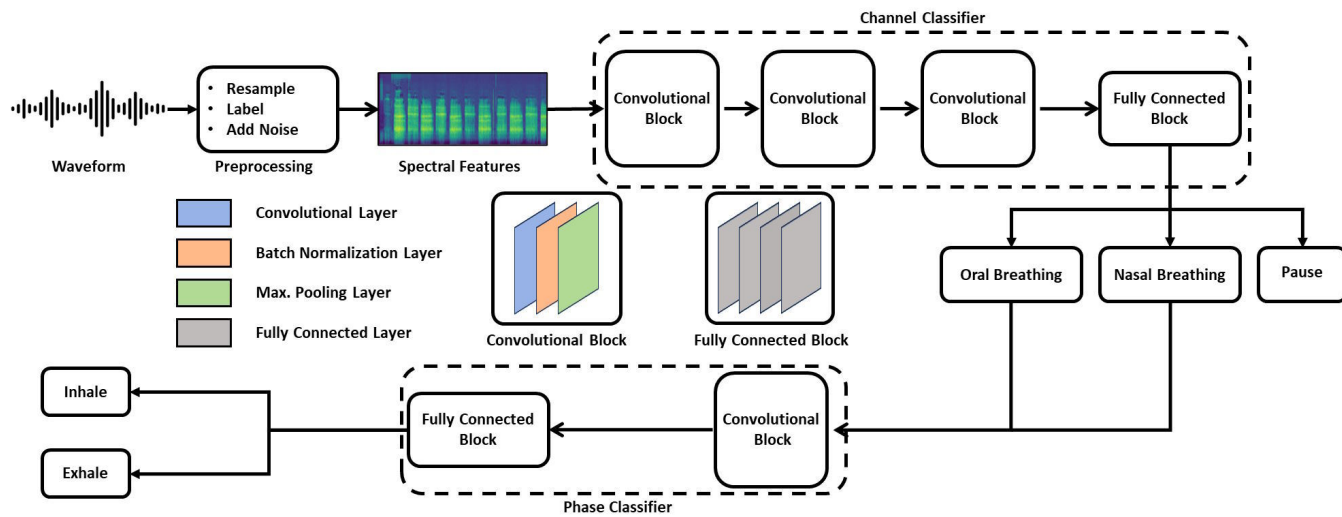
Figure 4.2: The system architecture for detecting pause, breathing channels, and breathing phases.

A binary cross-entropy loss function was used to train all models. In each instance, the model was trained until the F1 score of the test dataset stopped improving for 30 epochs, after which the highest F1 score was reported and the corresponding model checkpoint was saved.

4.3   Results and discussion

To evaluate the system performance, we conducted k-fold cross-validation on the dataset to provide a better estimate of the model performance on new data. Specifically, we used a leave-one-out cross-validation (LOOCV) method where data from seven subjects was used for training and the eighth subject's data was used for validation. This method prevents any data leakage between the training and testing datasets, and it also aids us in identifying any limitations of the model and detecting any shortcomings of the dataset.

With LOOCV of the channel classifier, we observed that using mel-spectrograms as an input feature produced the best results with an average F1 score of 93.98% (SD=5.02%). In comparison, using MFCCs as input features produced an average F1 score of 90.35% (SD=8.49%). When using mel-spectrograms, the highest F1 score of (97.99%) was obtained by excluding user #5 from the training dataset. Conversely, the lowest F1 score of (82.45%) was obtained by excluding user #6 from the training dataset. When using MFCCs, the results were similar to the mel-spectrograms case, with the highest F1 score of (97.50%) (by excluding user #5) and the lowest F1 score of (75.93%) (by excluding user #2).

The phase classifier was evaluated using a similar LOOCV approach and a comparison between mel-spectrograms and MFCCs as input features revealed that

mel-spectrograms were a better input feature for training the phase classifier. The average F1 score with mel-spectrograms was seen to be 76.20% (SD=8.76%) which was higher than the average F1 score of 75.56% (SD=8.55%) with MFCCs. The results also had a high standard deviation across the folds. Using mel-spectrograms, the highest F1 score of 89.46% was obtained by excluding user #4 from the training dataset, and the lowest score of 63.33% was obtained by excluding user #6 from the training dataset. When using MFCCs, the highest and lowest F1 scores obtained were 87.35% (by excluding user #4) and 63.71% (by excluding user #5), respectively.

The average F1 scores for phase classification were relatively low in comparison to the average F1 scores for channel classification. This suggests that either the CNN is not a very effective architecture to classify breathing phases, or there is insufficient variety in the dataset to fully capture the problem's complexity. Overall, the mel-spectrograms as input features performed better than MFCCs in training a neural network for breathing channel and phase classification. Table 4.1 shows the results of the experiments and provides a comparison between LOOCV using the two input features.

Table 4.1: Comparing mel-spectrograms and MFCCs as input features

Classifier	Features	F1 Score (%)		
		Avg. (SD)	Max.	Min.
Channel	Mel-spectrogram	93.98 (5.02)	97.99	82.45
	MFCC	90.35 (8.49)	97.50	75.93
Phase	Mel-spectrogram	76.20 (8.76)	89.46	63.33
	MFCC	75.56 (8.55)	87.35	63.71

4.4 Conclusion

This chapter presents a system for real-time detection of breathing channels and phases by capturing audio signals from a pair of wireless earphones when a user performs breathing exercises. Recognizing the challenges of monitoring therapy adherence outside the clinical setting – where specialized equipment and professional oversight are scarce – this research leverages consumer-grade hardware, specifically earphones and smartphones, as a viable solution for at-home therapy management. By employing two neural network classifiers trained on a custom dataset, the system is capable of real-time detection of breathing channels and phases, achieving significant accuracy. The efficacy of these classifiers, as evidenced by k-fold cross-validation, showcases the feasibility of using common wireless earphones for accurate, real-time breathing therapy monitoring. These findings serve as a promising starting point, presenting a practical and efficient approach to enhancing therapy compliance. Such compliance is vital for the successful outcome of therapy, especially in settings where traditional clinical resources are unavailable.

Chapter 5

Indoor localization of the elderly using ambient Wi-Fi

5.1   Introduction

According to the United Nations, the global population aged 65 and above is expected to double from 761 million in 2021 to 1.6 billion in 2050 [101]. By the end of the current decade, 1 in 6 people in the world will be age 60 or above. As the risk of developing comorbidities and chronic conditions grows with aging, a rise in the aging population exacerbates challenges faced by the healthcare system in general and caregivers in particular. A common solution proposed to address this problem is to facilitate aging in place with the help of AAL.

The concept of AAL refers to the use of intelligent technologies to help improve the quality of life for the elderly by enabling them to live independently and safely. Illustrative applications of AAL include monitoring the subject's vitals such as heart rate [102, 103] and blood glucose level [103], tracking location [104], recognizing

wandering behavior [105], detecting falls [106], recording health metrics such as sleep quality [107] and ROM [108, 109], and recognizing activities of daily living [110], among others. Before the advent of AAL, such information was gathered manually and the responsibility of data gathering and analysis fell on the caregivers and healthcare providers, respectively. As the Internet of Things (IoT) enabled devices gain popularity, some of the aforementioned monitoring tasks are being delegated to automated systems that require minimal user input, if any. Many IoT devices operate in the background and do not get in the way of the user as they perform their activities of daily living. The passive nature of these devices supports the development of systems that promote AAL and enable the elderly to live independently. This work focuses on telemonitoring of the elderly in an indoor environment by proposing a non-intrusive solution for performing indoor localization to facilitate AAL. Such an AAL telemonitoring system can allow caregivers to monitor and assess daily patterns of space use by an elderly and help locate them quickly and accurately in case of an emergency.

This chapter is divided into six sections. Section 5.2 examines previous literature on this topic and makes a case for the system proposed in this work. Section 5.3 elaborates on the working of fine timing measurement (FTM) protocol and discusses the approach used for the access point (AP) and mobile device localization. Section 5.4 discusses the experiments performed and presents the results. Section 5.5 discusses the practical considerations and limitations of the proposed solution. Finally, section 5.6 provides some concluding remarks.

5.2   Related work

Several indoor human monitoring and tracking systems have been proposed over the years. These systems can be classified into three categories based on their sensor type: (i) vision-based sensors, (ii) wearable sensors, and (iii) ambient sensors. Vision-based telemonitoring systems include camera-based systems [111, 112] and other optical sensors, such as infrared [113], to record and analyze the activities of the subject. These systems detect and track objects by analyzing the image data obtained from the image sensors. Despite gaining popularity in the past couple of decades, these systems raise privacy concerns due to the sensitive nature of the data collected by the sensors. This is a significant issue for vision sensors installed in a home setting, forcing the end-user to trade-off between privacy and autonomy [114]. Additionally, traditional optical sensors are limited by line-of-sight constraints that drive up the installation cost when ensuring effective coverage of the environment for performing indoor localization.

The sensors in the wearable category mostly include IMUs that are either used for motion tracking [109, 115] or as pedometers [116, 117]. These sensors are usually embedded in consumer devices such as smartphones and smartwatches [115, 118, 119] and are used in pedestrian dead-reckoning (PDR) systems. The use of wearable sensors is preferable from a privacy standpoint, but IMUs are prone to sensor drift and pedometer-based PDRs suffer from step misclassification, both of which lead to error accumulation over time. Cases where specialized hardware is used, such as foot-mounted pedometers, also entail proper installation and maintenance. Thus, such systems are unreliable when used by untrained individuals and require frequent error correction to function reliably.

The ambient sensor category includes a variety of wireless technologies such

as Bluetooth [120], radio frequency identification (RFID) [121], ultra-wideband (UWB) [122], Wi-Fi [120], etc. Based on the use case, several approaches can perform indoor localization using wireless devices with varying degrees of success. Some solutions utilize signal propagation path loss models and the received signal strength (RSS) of a wireless AP to determine its distance from a receiver device. This scheme is often supplemented with fingerprinting methods to locate the device indoors [104, 118]. Although this approach works in principle, the high variance in RSS data makes it impractical for use in small spaces, such as apartments and homes, which require a localization accuracy of less than a couple of meters. Moreover, fingerprinting techniques require the user to collect a large amount of data and use it to train the system offline before it can be used. Several studies have successfully used the channel state information (CSI) of wireless devices to perform localization [123, 124]. However, this information is not readily provided by commercially available mobile devices, making this approach infeasible for use in home environments using commodity hardware. Methods using the FTM protocol are a viable alternative for locating a mobile device indoors with a high degree of accuracy [125]. Several approaches have also combined FTM ranging with PDR systems that utilize IMUs [126, 127] to improve the accuracy.

FTM is a relatively new technology and the consumer wireless AP devices have only recently started to support it. Thus, some of the earlier studies used specialized hardware in their experiments to perform FTM-based indoor positioning [128]. A majority of the prior studies were conducted in a controlled environment and their primary objective was to show sub-meter accuracy when performing localization [129]. Finally, the prior studies either relied on *a priori* location information of the wireless APs when performing localization [129] or utilized fingerprinting methods

[129, 130], making them infeasible for use in unstructured environments. The primary aim of this study is to explore the potential of the FTM technology in the context of promoting AAL for the elderly by evaluating it in an unstructured environment, i.e., the location of multiple APs in the vicinity of a subject is unknown.

5.3   Wi-Fi fine timing measurement

Standardized by IEEE in 2016, FTM is a protocol that allows the measurement of round trip time entailed in the exchange of data packets between Wi-Fi-enabled mobile devices and Wi-Fi network APs [131]. The FTM protocol requires the Wi-Fi devices to keep track of the timestamps, with nanosecond accuracy, corresponding to each data packet's arrival and departure. The total travel time taken for a data packet to complete one cycle of FTM process is called the round trip time. Since electromagnetic waves traveling at the speed of light form the basis of Wi-Fi communication, the knowledge of round trip time can be exploited to determine the distance between the transmitting and receiving Wi-Fi devices.

A complete cycle of the FTM process consists of the following steps. First, an initiator (mobile device) requests an FTM session from a responder (AP). Then at timestamp t_1, a data packet is sent by the responder to the initiator. This packet arrives at the initiator at timestamp t_2, and a reply is sent back to the responder from the initiator at timestamp t_3. This reply is received by the responder at timestamp t_4, which completes one cycle of the FTM process. Since the clocks on the initiator and responder devices are not synchronized, using the timestamps to determine the distance between the two devices leads to errors in the calculation.

Thus, each device calculates a time difference for the data packets, which obviates the need for clock synchronization. Specifically, the initiator calculates the time difference $\Delta t_i = t_3 - t_2$ and the responder calculates the time difference $\Delta t_r = t_4 - t_1$. The initiator's time difference Δt_i is included as part of the data packet sent by it to the responder. Once a cycle of the FTM process is completed, the round trip time computed by the responder is given by the following equation

$$T_R = \Delta t_r - \Delta t_i,$$

where T_R is the total time taken for the exchange of a data packet during one cycle of the FTM process, excluding the processing time taken by the initiator and responding devices.

With the timestamp measurements of FTM process yielding nanosecond resolution and Wi-Fi signals operating at the speed of light, a small error in the round trip time determination can throw off the distance estimate. To mitigate this challenge, the FTM cycles are performed in bursts. Specifically, a burst is a short-duration transmission sequence where multiple cycles of the FTM process are performed in quick succession. At the end of an FTM burst, the average of all round trip times for the FTM cycles is calculated and considered as a more accurate measurement. This can be expressed in the following equation

$$T_m = \frac{1}{K} \sum_{k=1}^{K} T_{R_k},$$

where T_m is the average round trip time of the signal and K is the number of FTM cycles in an FTM burst. The average round trip time can be multiplied by the speed of light to get the round trip distance estimate. It is standard practice to add

a constant correction value ε, either to the round trip time in nanoseconds or the distance estimate in meters to account for the time it takes for the device hardware to receive the data packet and assign a timestamp to it. Now, the separation distance between the initiator and responding devices can be calculated as follows

$$d = \frac{c \times (T_m - \varepsilon)}{2},$$

where d is the estimated distance and c is the speed of light. Figure 5.1 provides a pictorial representation of FTM calculation.

In principle, FTM localization involves the following three steps.

- Estimate the location of the APs if they are unknown

- Perform multilateration [132] to locate the mobile device

- Post-process the data to improve accuracy and reliability
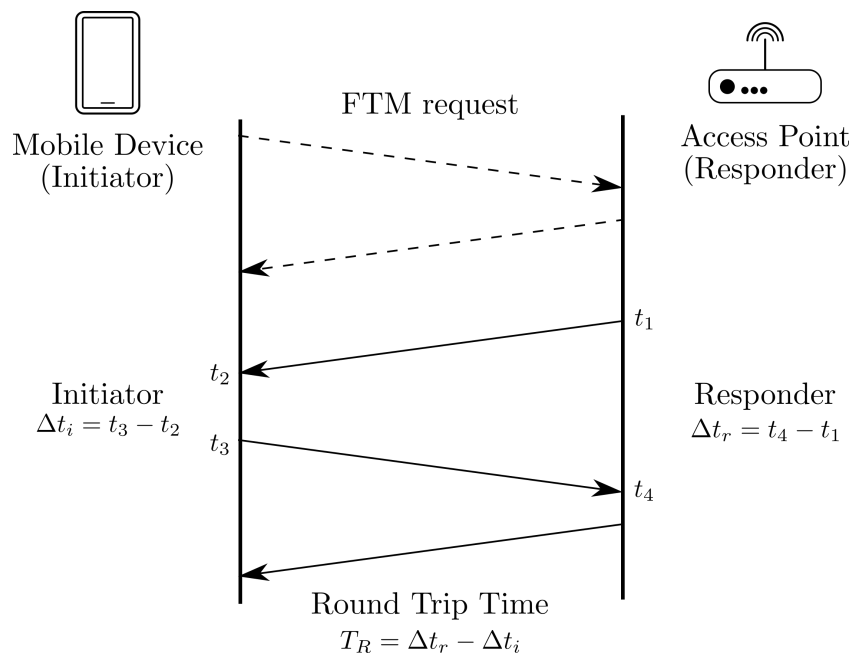
Figure 5.1: FTM calculation.

In prior works such as [129], the location of the APs was assumed to be known with a high degree of accuracy. Such *a priori* knowledge helps improve the location estimation of the mobile device because the AP locations act as ground truth against which the estimated position of the mobile device can be verified. However, in unstructured environments arising in the use case considered in this research, it is not always feasible to place multiple APs at known locations and it is often impossible to know the exact positions of one or more APs in the vicinity of the mobile device. In such cases, estimating the AP location becomes crucial for performing FTM-based localization. Thus, in the next section, the experimental setup and the position estimation of APs are discussed before delving into the FTM-based localization of the mobile device.

5.4 Experimental setup

The experiment was conducted inside and around the Mechatronics, Controls, and Robotics Laboratory at New York University. The test area, where all FTM measurements were performed, is fully furnished and covers 106 m^2. The laboratory is equipped with a significant amount of metallic test equipment and is actively used by people, making it a suitable setting for conducting the experiment. The total area, which includes the test area and the locations of all APs, spans 440 m^2.

To validate the indoor localization system, FTM data were collected in the test area using an Android mobile device at ten reference points (RPs) from six APs located in the vicinity. Only commercially available consumer-grade hardware was used for data collection. The Samsung Galaxy S20 Ultra 5G mobile device running Android was utilized as the FTM initiator, and six *Google Nest Wifi Pro*

Wi-Fi APs served as the FTM responders. Out of the six APs, only one (AP1) was situated inside the test area. Figure 5.2 shows the test area, the locations of the six APs, and ten RPs.

5.4.1 Analysis of indoor FTM measurements

Ideally, if environmental factors such as multipath propagation and signal fading were non-existent, the FTM method described in section 5.3 would yield perfect distance estimates. These estimates can then be translated into highly accurate positional information. However, realistic indoor environments introduce complexities that can significantly affect the accuracy of FTM measurements.

In an indoor setting, multipath interference occurs when signals reflect from various surfaces before reaching the receiver, causing interference and phase shifting of the signal, which can lead to signal fading. This can distort the perceived time of flight of the signal, leading to errors in the distance measurement. Another cause of errors in distance measurement is signal shadowing which occurs when the signal is blocked by large obstacles such as walls, large metallic equipment, or furniture, resulting in a weaker signal at the receiver and potentially inaccurate distance estimates.

To assess the performance of FTM ranging in these realistic scenarios, an analysis of indoor FTM measurements was performed under both line-of-sight (LOS) and non-line-of-sight (NLOS) conditions. This process involved collecting a dataset of FTM measurements where AP1 (primary AP) was in direct view (LOS) and the rest of the APs were obstructed (NLOS). The collected data was subjected to statistical analysis to ascertain the distribution of the FTM ranging measurements. For each AP, a histogram and an empirical cumulative distribution function (ECDF)
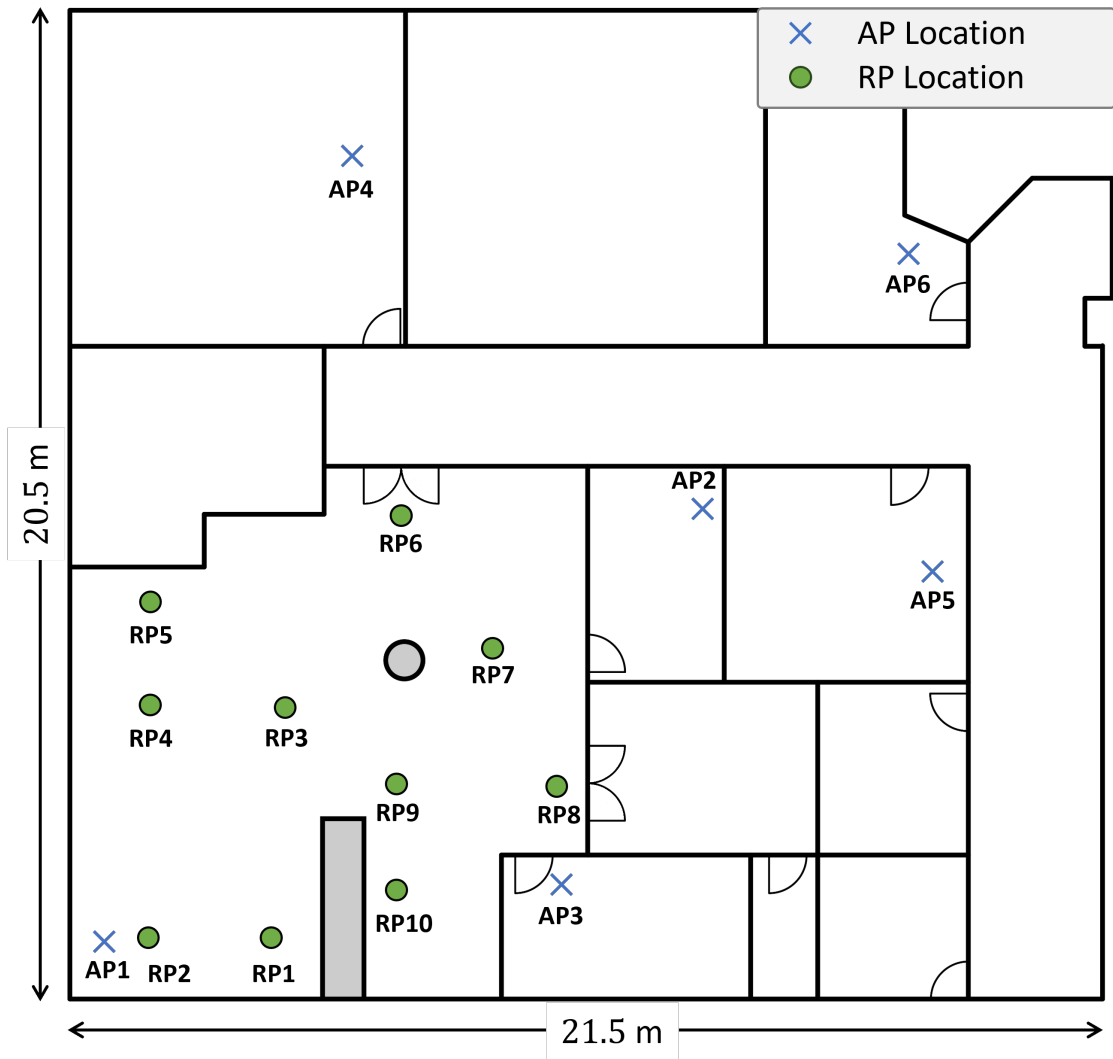
Figure 5.2: Test area marked with AP and RP locations.

were computed to visualize the measurements to determine whether the system is overestimating or underestimating the distances. Figure 5.3 shows the histograms and ECDFs for the six APs at an RP 5.2 m from AP1 inside the test area.

As shown in Figure 5.3, inconsistencies were observed in the system when estimating the distance of an AP from an RP. Notably, for AP2 and AP4, the system significantly underestimated the distance, while for AP5, it slightly overestimated it. This behavior is unusual, as distances are typically overestimated due to multipath effects, which add additional time to the signal's round trip time. This discrepancy suggests that the Android device preprocesses the FTM ranging data before providing it to the user, necessitating a different strategy to accurately determine the positions of APs in the vicinity.

Another investigation involved determining whether the data follows a Gaussian distribution. Although the mean error measurements occasionally deviated significantly from zero, the error values themselves showed limited variation, suggesting a potentially Gaussian distribution. However, the Shapiro-Wilk test [133] revealed that the ranging measurements did not follow a Gaussian distribution. Consequently, the median was used as a measure of central tendency to estimate the distance of the AP from an RP.
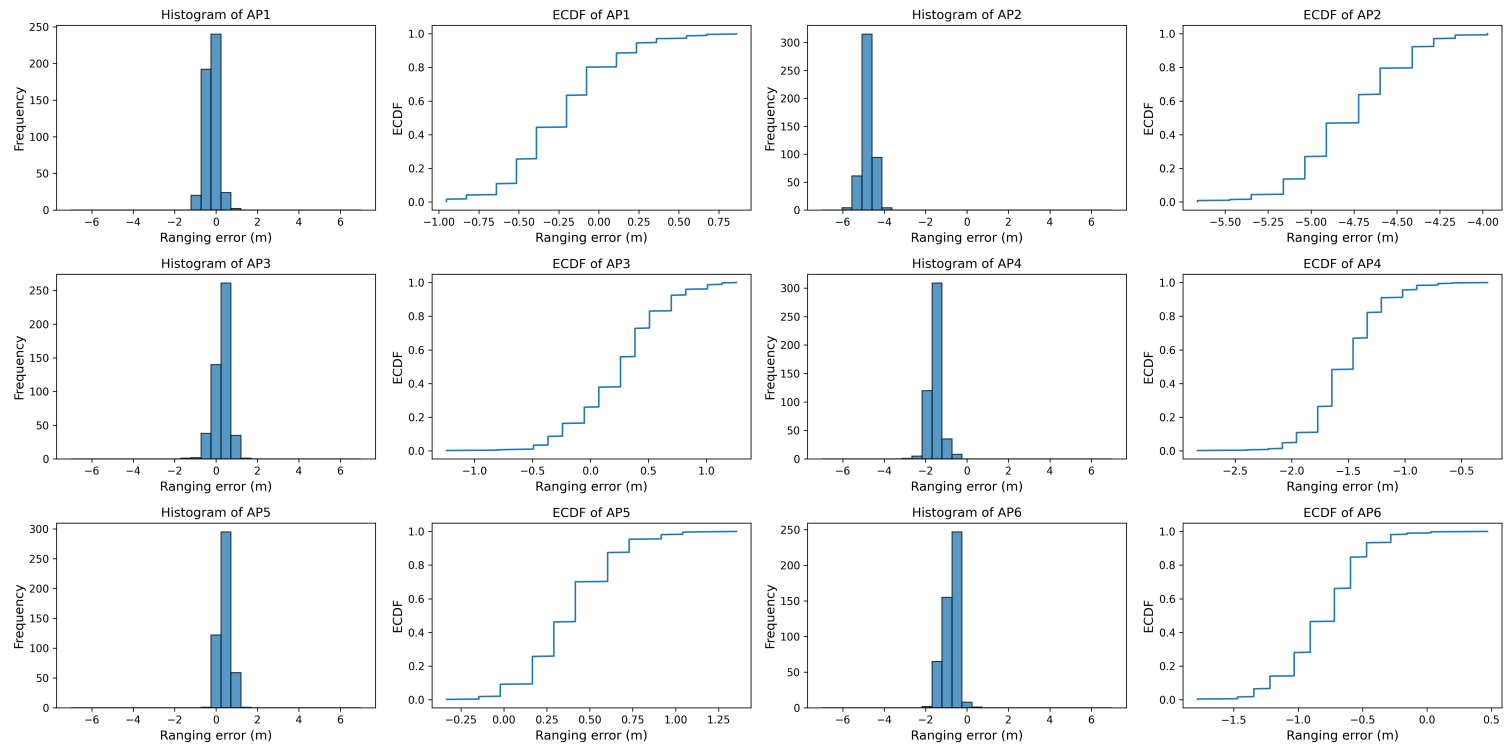
Figure 5.3: Histograms and ECDFs for the six APs at an RP. Only AP1 maintains LOS to the RP.

5.4.2   Access point localization

It is assumed that most residential establishments, such as apartments and homes, have at least one AP located at a known position, this AP can be designated as the primary AP. Next, the position of the primary AP is used to establish a common coordinate system such that the positions of other neighboring and detectable APs, whose exact locations are generally inaccessible to the user, can be estimated with respect to this primary AP.

To perform the initial position estimation, within their apartment or home, the user moves to several different RPs whose coordinates relative to the primary AP are *known*. Next, the user collects FTM-based distance measurements to all available APs at each RP for a chosen time duration. The coordinates of each identified AP are estimated relative to the previously established common coordinate system at the primary AP. As delineated below, to estimate the position of each of these APs, the measured distances are used to perform multilateration using weighted least squares approximation. Note that the weights used in the least squares algorithm are the inverse of the estimated distance measurement for each AP at the corresponding RP. This approach is based on the observation that, generally, the likelihood of encountering significantly large outliers increases with greater distances between the initiator and the responder [134]. Hence, assigning weights inversely proportional to the estimated distances helps mitigate the impact of such outliers, leading to more reliable localization results. Informed by the analysis in subsection 5.4.1, the median of the distance measurements has been adopted as the distance estimator for each AP from an RP, offering a robust measure against outliers.

To elaborate further, the ith AP position $p_i = (p_{i_x}, p_{i_y})$ is estimated using

$$p_i = \underset{p_{i_x}, p_{i_y}}{\arg\min} \sum_{n=1}^{N} w_{i_n} \left(e_{i_n} \right)^2$$

where w_{i_n} is the weight associated with the ith AP at the nth RP, e_{i_n} is the distance estimation error from the nth RP to the ith AP, $p_i = (p_{i_x}, p_{i_y})$ is the estimated location of the ith AP, and N is the total number of RPs, where $N \geq 3$ non-collinear points are considered for the 2D case considered here. For the 3D case, $N \geq 4$ non-coplanar points are required. The estimation error e_{i_n} is defined as follows

$$e_{i_n} = d_{i_n} - \sqrt{(p_{i_x} - r_{n_x})^2 + (p_{i_y} - r_{n_y})^2}$$

where d_{i_n} is the FTM-based measured distance at the nth RP from the ith AP, and $r_n = (r_{n_x}, r_{n_y})$ is the known location of the nth RP. Figure 5.4 shows an example of AP localization with three RPs.

Once the initial estimation of AP positions is completed, the APs are considered to be anchor nodes, i.e., their location is assumed to be fixed, and the user is considered to be the mobile node, i.e., their location is dynamic. The accuracy of the AP position estimates directly affects the subsequent position estimation of the user. Thus, as time progresses and the user explores new locations in the apartment or home, the estimation process can be repeated periodically to refine and update the AP location estimates.

After conducting AP localization, we found that the most precise results were obtained by employing multilateration with only the nearest four RPs to each AP. This approach has been consistently applied across the entirety of this study. Table
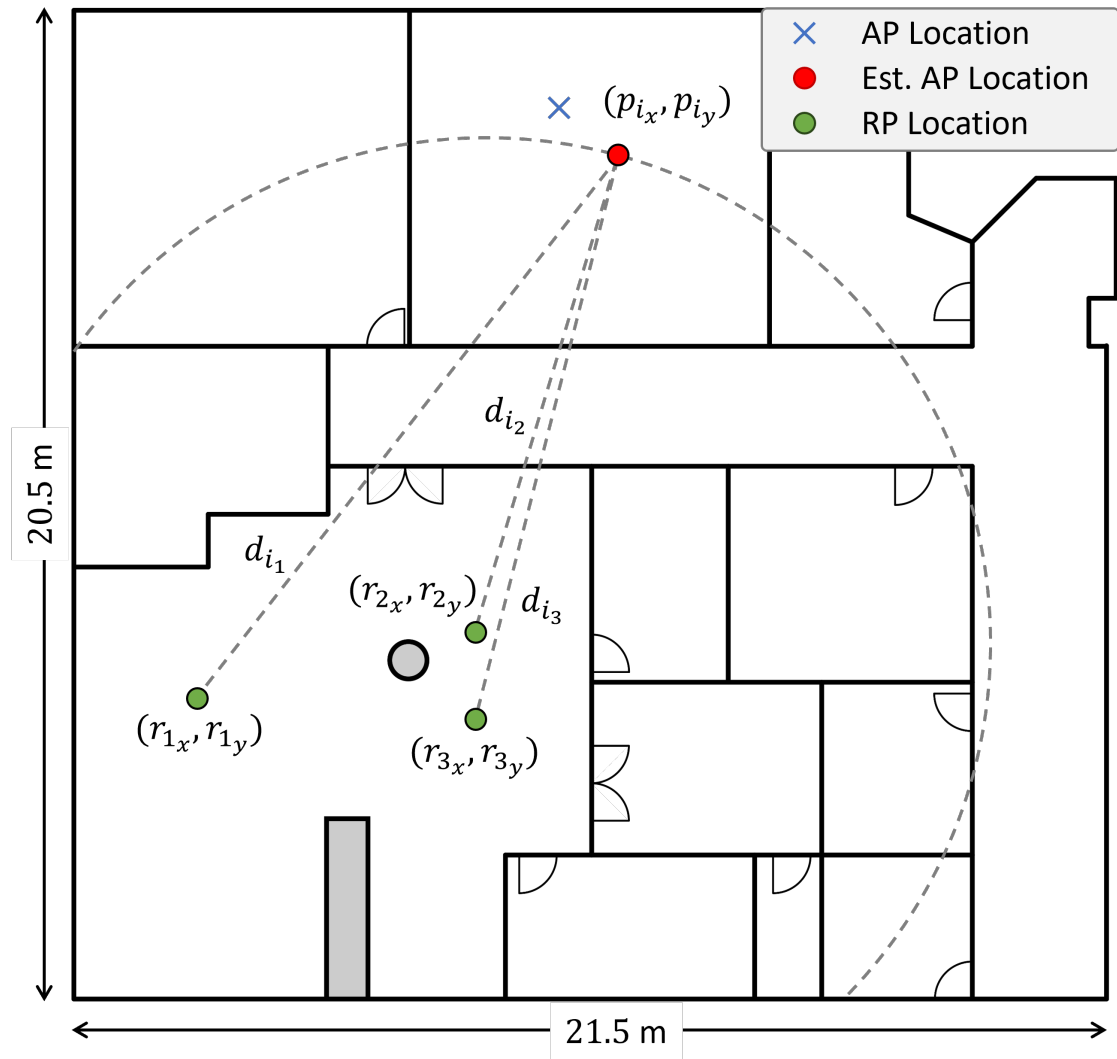
Figure 5.4: AP localization in 2 dimensions with three RPs.

Table 5.1: Absolute positioning error for AP localization for both datasets.

Dataset	Absolute positioning error (m)					
	AP1	AP2	AP3	AP4	AP5	AP6
1	0.00	0.99	0.99	0.78	3.14	29.02
2	0.00	1.04	0.97	1.68	2.71	28.94

5.1 presents the absolute positioning error when AP localization was performed over data derived from two datasets collected within the test area. For each dataset, data were collected at every RP over a duration of two minutes. Figure 5.5 provides a visual depiction of these positioning errors.

It can be observed that the closest four APs yield the most accurate results whereas AP5 and AP6 have high positioning errors.

5.4.3   Mobile device localization

After estimating the positions of the APs, the next step is to determine the location of the mobile node, i.e., the user. This process is similar to the one described in the preceding section, with a critical distinction: the APs now serve as *fixed* RPs, and the FTM ranging data between the APs and the mobile device are leveraged to determine the mobile device's location. At each moment, the mobile device's position is calculated using the weighted least squares method, capitalizing on the fact that FTM bursts provide new ranging data every second. However, the system's accuracy diminishes if the user is in motion, owing to the increasing variance in measurement, rendering it less viable for real-time indoor localization unless the user is within LOS of multiple APs. Consequently, our analysis concentrates on scenarios where the user remains stationary, akin to situations where an individual might require urgent medical aid due to a fall. This specific scenario was simulated by collecting data at three distinct locations, with the findings presented in Table
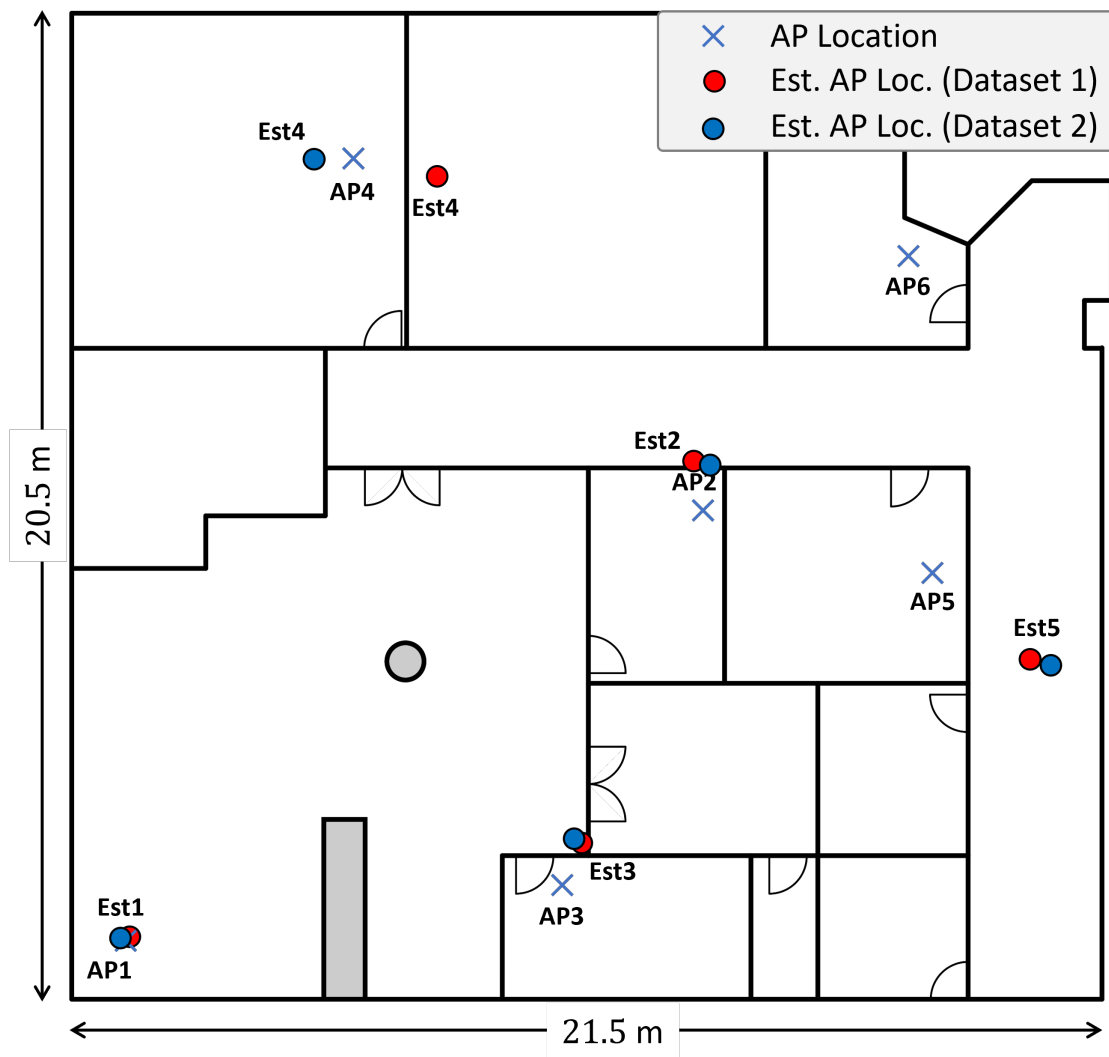
Figure 5.5: Estimated positions of APs for each dataset.

Table 5.2: Absolute positioning error for three test points using estimated AP locations from two datasets.

Dataset	Absolute positioning error (m)		
	Test point 1	Test point 2	Test point 3
1	0.53	0.79	0.95
2	0.49	0.47	0.89

5.2. At each of the three test points, the positioning error was less than 1 m which is sufficient for indoor localization of a person. Additionally, it is important to note that at each of the test points, data collection was limited to a duration of only 30 seconds.

5.5   Limitations of the proposed system

FTM-based localization solutions represent a significant advancement over RSSI-based solutions, offering more accuracy and reduced susceptibility to multipath effects. Despite these advantages, there are still limitations to consider, especially when designing solutions for the elderly. The initial phase of data collection for AP localization is critical, as the quality of FTM ranging data collected at each RP greatly influences the precision of AP location estimates. A particular challenge arises from the fact that the locations of most APs utilized by the system are unknown, exacerbating the issue. Should any APs be relocated, a recalibration becomes necessary. Moreover, the system's functionality is compromised in the absence of a sufficient number of APs within the vicinity. While this is less of a concern in densely populated urban environments, it becomes more problematic in suburban areas where residences are more dispersed. A possible solution can be the installation of additional APs within a household. Another hurdle is the current landscape of Wi-Fi infrastructure, where many homes have yet to upgrade to

FTM-compatible routers. Although this issue may diminish over time as consumer technology progresses, it remains relevant for contemporary application of the technology. On its own, the system can provide a coarse estimation of a user's location in near real-time, which may suffice for caregivers seeking a general sense of the elderly individual's well-being. However, for more precise, real-time location estimation, enhancements are needed either through increased sampling rates or by incorporating additional sensors to improve the system's accuracy. These considerations are essential for tailoring the system to effectively meet the needs of its intended users, ensuring both its current utility and future potential.

5.6   Conclusion

This chapter explores the utility of ambient Wi-Fi data for the indoor localization of the elderly. Our experiments demonstrate that in static conditions, FTM ranging proves to be a reliable method for achieving high accuracy in localization within a short timeframe. Although the hardware both over and underestimated the FTM measurements at several RPs, the use of the median as a measure for central tendency helped achieve relatively low positioning error. Three test points in the test area were chosen to test the algorithm. The positioning accuracy achieved was below 1 m which is more than enough for localizing an individual indoors. This approach holds promise for monitoring stationary positions, which can be critical in various health and safety contexts.

Chapter 6

Conclusions and future work

6.1 Summary and contributions

Chapter 1 outlines the transformative role of digital health in redefining health-care, highlighting its benefits for both patients and providers. It emphasizes the impact of technological advancements, such as artificial intelligence, machine learning, and wearable technologies, in enhancing healthcare delivery and patient outcomes. The chapter addresses the challenges and opportunities presented by an aging global population, emphasizing the necessity for innovative healthcare solutions like Health-IoT and AAL to support aging in place and mitigate care-giver shortages. Additionally, it explores the accelerated adoption of digital health technologies in response to recent global health crises, such as the COVID-19 pandemic, and the potential of human-robot interaction to improve healthcare services. Through this discussion, it introduces the main focus of the dissertation which is the advancement of digital health and telerehabilitation which sets the stage for further exploration of digital health's capacity to create a more accessible,

personalized, and efficient healthcare system.

Chapter 2 follows by introducing a remote control system for dialysis machines using mobile HRI, aimed at enhancing the COVID-19 emergency response. The proposed system utilizes smartphones or tablets and allows healthcare professionals to remotely operate a robot to interact with the dialysis machine's touch control panel, ensuring social distancing and reducing PPE usage. A study conducted involving participants using a tablet to control the robot through a live video feed demonstrated successful completion of tasks related to the remote manipulation of the machine's interface. The overall feedback for the device was positive and it led to improvements in the system's interface for a more intuitive user experience. This system avoids the need for custom UI development, enabling rapid deployment and efficient use of resources in emergencies.

Chapter 3 explores the utility of DVAs equipped with cameras beyond their intended use, employing them for ROM assessment. This approach uses real-time video analysis to monitor joint movements, with a focus on the shoulder and elbow, offering a promising alternative to traditional methods. The system, validated against both synthetic data and Kinect sensor measurements, proves to be a reliable tool for ensuring therapy adherence, highlighting its potential to enhance patient care through at-home monitoring. It also demonstrates that commercially available devices that are not originally designed for healthcare applications, can be effectively repurposed to advance digital health and telerehabilitation.

Chapter 4 explores the use of accessible technology, such as earphones and smartphones, as a viable solution for at-home therapy management. Using two neural network classifiers trained on a custom dataset containing audio recordings of breathing from healthy individuals, this system can accurately identify breathing

channels and phases in real-time. The reliability of these classifiers was validated through k-fold cross-validation, demonstrating the potential of off-the-shelf wireless earphones for accurate monitoring of breathing therapies. Reflecting the findings of Chapter 3, this chapter further highlights the utility of consumer-grade hardware in facilitating tasks like therapy adherence monitoring.

Chapter 5 discusses the use of ambient Wi-Fi data for indoor localization of elderly individuals, emphasizing the potential of FTM technology in unstructured environments. Experimental analysis shows that FTM ranging is highly accurate in static conditions, making it a viable solution for accurately determining a person's location within their home. This advancement provides caregivers with an additional tool to enhance elderly care by facilitating the remote monitoring of indoor movements without compromising their privacy.

To summarize, this dissertation makes several key contributions to the field of digital health, leveraging emerging technologies to enhance healthcare delivery and patient care. These contributions are a step forward in addressing contemporary healthcare challenges, particularly those brought forth by global health crises such as pandemics and the ongoing issue of an aging population.

Throughout this dissertation, the use of a diverse array of technologies–ranging from using live video feeds from commercially available webcams as UIs to using the cameras of DVAs for ROM assessments, from the use of standard earbuds for detecting breathing phases and channels to leveraging ambient Wi-Fi signals for indoor localization–demonstrates the versatility and potential of cross-disciplinary technology applications in healthcare. By repurposing devices and data not initially intended for medical use, such as DVAs and ambient Wi-Fi, this research highlights the immense possibilities for crafting innovative, effective, and efficient healthcare

solutions. This not only broadens the scope of digital health technologies but also underlines the potential for innovative integration of everyday technologies into healthcare practices and provide healthcare solutions that are both effective and efficient.

6.2   Future research

Digital health encompasses a broad spectrum of research opportunities. Among these, significant challenges specific to the geriatric population present urgent and important problems warranting further exploration. Issues like incontinence, dementia, depression, and loneliness are some of the most pressing issues affecting the elderly. Addressing these problems through research not only represents a noble endeavor but also a vital necessity. This dissertation has primarily concentrated on AAL and the innovative use of existing technologies to enhance the accessibility and efficiency of healthcare, pointing toward a future where digital health solutions can significantly improve the quality of life for the elderly.

For the work presented in Chapter 2, the integration of force feedback control on the robot's end effector can be a promising next step. This will improve the precision and reliability of the system by closing the control loop and making the system more robust. Additionally, placing multiple fiducial markers on the ICPT could substantially increase the robot's accuracy. As technological advancements continue to drive down the costs of hardware, the webcam used for visual input can be replaced with a dedicated depth camera to eliminate the need for placing fiducial markers on the ICPT. This can be aided by deep-learning-based methods for the automatic detection and localization of the ICPT. The possibility of a mobile

robot platform can also be explored, which would enable remote interaction with various medical equipment within a given setting, vastly expanding the system's applicability and flexibility. Finally, additional intuitive modes of interaction involving wearable technologies and AR could revolutionize the user experience, enhancing efficiency and engagement while simultaneously reducing the cognitive load associated with task execution.

The work presented in Chapter 3 can be further improved by exploring the possibility of training deep learning models with DVA-based measurements to perform ROM assessment and determine if such a model can detect and track changes in the ROM of an individual over time. Such a system will be useful for the elderly and it will help them prevent or, at the very least, slow down the aging-related loss of ROM by ensuring adherence to therapy. Healthcare providers can also use the data provided by the system to assign therapies tailored to specific patients, which can lead to a more effective therapy outcome. Chapter 4 opens future research avenues for the design and development of at-home therapy compliance applications involving breathing exercises. The system can be further enhanced by utilizing advanced deep learning methods such as joint embeddings for coupling audio and video data. Future work can explore the combination of vision-based models with acoustic models to simultaneously detect and track physical and breathing exercises for a holistic compliance monitoring solution. Such work can lead to the development of a personalized therapy compliance monitoring system that is tailored to specific users, furthering the scope of this work in remote monitoring systems. Finally, the methodologies outlined in Chapter 5 could be enhanced by implementing on-device machine learning, since machine learning models have the potential to better understand the complexities of the

environment. With further engineering efforts, the frequency at which FTM data is collected could be increased, leading to more rapid and accurate indoor position estimations. Additionally, augmenting FTM data with other sensing modalities, such as IMUs, could further improve positioning accuracy, making the system more effective for localizing both stationary and moving individuals. There is substantial potential in harnessing this localization data to analyze behavioral patterns of the elderly. Such analyses could lead to early detection of behavioral anomalies or shifts indicative of health issues, including depression. This advancement could significantly improve the quality of care and monitoring for the elderly, offering a more nuanced understanding of their daily activities and overall well-being. Our work lays the groundwork for future research in this domain, highlighting the necessity for more sophisticated models and technologies to address the challenges of dynamic indoor localization. Further exploration and development of these systems will be crucial in realizing the full potential of indoor localization technology in elder care and beyond.

Bibliography

[1] P. Tsinganos and A. Skodras, "A smartphone-based fall detection system for the elderly," in *International Symposium on Image and Signal Processing and Analysis.* IEEE, 2017, pp. 53–58.

[2] A. V. Román, D. P. Martínez, Álvaro Lozano Murciego, D. M. Jiménez-Bravo, and J. F. de Paz, "Voice assistant application for avoiding sedentarism in elderly people based on IoT technologies," *Electronics*, vol. 10, no. 8, p. 980, 2021.

[3] United Nations Department of Economic and Social Affairs, Population Division, "World population prospects 2022: Summary of results," 2022, Accessed: Mar. 7, 2024. [Online]. Available: https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/wpp2022_summary_of_results.pdf

[4] J. Vespa, "The US joins other countries with large aging populations," US Census, 2018, Accessed: Mar. 7, 2024. [Online]. Available: https://www.census.gov/library/stories/2018/03/graying-america.html

[5] Y. A. Qadri, A. Nauman, Y. B. Zikria, A. V. Vasilakos, and S. W. Kim, "The future of healthcare internet of things: A survey of emerging technologies,"

IEEE Communications Surveys & Tutorials, vol. 22, no. 2, pp. 1121–1167, 2020.

[6] J. K. Eckert, L. A. Morgan, and N. Swamy, "Preferences for receipt of care among community-dwelling adults," *Journal of Aging & Social Policy*, vol. 16, no. 2, pp. 49–65, 2004.

[7] J. Binette and K. Vasold, "Home and community preferences survey: A national survey of adults age 18-plus," AARP Research, 2018, Accessed: Mar. 7, 2024. [Online]. Available: https://www.aarp.org/content/dam/aarp/research/surveys_statistics/liv-com/2018/home-community-preferences-survey.doi.10.26419-2Fres.00231.001.pdf

[8] M. J. Giummarra, B. Haralambous, K. Moore, and J. Nankervis, "The concept of health in older age: Views of older people and health professionals," *Australian Health Review*, vol. 31, no. 4, pp. 642–650, 2007.

[9] National Academies of Sciences, Engineering, and Medicine, *Social isolation and loneliness in older adults: Opportunities for the health care system.* The National Academies Press, 2020.

[10] N. Super, "Who will be there to care? the growing gap between caregiver supply and demand," *National Health Policy Forum*, no. 89, 2022.

[11] K. C. Fleming, J. M. Evans, and D. S. Chutka, "Caregiver and clinician shortages in an aging nation," in *Mayo Clinic Proceedings*, vol. 78, no. 8, 2003, pp. 1026–1040.

[12] J. Graham, "A shortage of caregivers," *The New York Times*, February 2014,

(accessed: Aug. 15, 2023). [Online]. Available: https://archive.nytimes.com/newoldage.blogs.nytimes.com/2014/02/26/a-shortage-of-caregivers/

[13] E. Flaherty and S. J. Bartels, "Addressing the community-based geriatric healthcare workforce shortage by leveraging the potential of interprofessional teams," *Journal of the American Geriatrics Society*, vol. 67, no. S2, pp. S400–S408, 2019.

[14] A. Oldenburg, "Nationwide caregiver shortage felt by older adults," AARP Public Policy Institute, 2022, Accessed: Mar. 7, 2024. [Online]. Available: https://www.aarp.org/caregiving/basics/info-2022/in-home-caregiver-shortage.html

[15] U.S. Department of Labor, Bureau of Labor Statistics, "Employment projections – 2021-2031," 2022, Accessed: Mar. 7, 2024. [Online]. Available: https://www.bls.gov/news.release/archives/ecopro_09082022.pdf

[16] J. L. Wolff, B. C. Spillman, V. A. Freedman, and J. D. Kasper, "A national profile of family and unpaid caregivers who assist older adults with health care activities," *JAMA Internal Medicine*, vol. 176, no. 3, pp. 372–379, 2016.

[17] S. C. Reinhard, H. M. Young, C. Levine, K. Kelly, R. B. Choula, and J. Accius, "Home alone revisited: Family caregivers providing complex care," AARP Public Policy Institute, 2019, Accessed: Mar. 7, 2024. [Online]. Available: https://www.aarp.org/content/dam/aarp/ppi/2019/04/home-alone-revisited-family-caregivers-providing-complex-care.pdf

[18] Y.-R. Hong, J. Lawrence, D. Williams Jr, and A. Mainous III, "Population-level interest and telehealth capacity of US hospitals in hesponse to COVID-19:

Cross-sectional analysis of Google search and national hospital survey data," *JMIR Public Health and Surveillance*, vol. 6, no. 2, p. e18961, 2020.

[19] D. M. Mann, J. Chen, R. Chunara, P. A. Testa, and O. Nov, "COVID-19 transforms health care through telemedicine: evidence from the field," *Journal of the American Medical Informatics Association*, vol. 27, no. 7, pp. 1132–1135, 2020.

[20] S. P. McGrath, K. M. McGovern, I. M. Perreard, V. Huang, L. B. Moss, and G. T. Blike, "Inpatient respiratory arrest associated with sedative and analgesic medications: impact of continuous monitoring on patient mortality and severe morbidity," *Journal of Patient Safety*, vol. 17, no. 8, pp. 557–561, 2021.

[21] J. P. Trevelyan, S.-C. Kang, and W. R. Hamel, "Robotics in hazardous applications," in *Springer Handbook of Robotics*, B. Siciliano and O. Khatib, Eds. Berlin, Germany: Springer, 2008, pp. 1101–1126.

[22] M. Hägele, K. Nilsson, J. N. Pires, and R. Bischoff, "Industrial robotics," in *Springer Handbook of Robotics*, B. Siciliano and O. Khatib, Eds. Berlin, Germany: Springer, 2008, pp. 1385–1422.

[23] K. Nagatani *et al.*, "Emergency response to the nuclear accident at the Fukushima Daiichi nuclear power plants using mobile rescue robots," *Journal of Field Robotics*, vol. 30, no. 1, pp. 44–63, 2013.

[24] E. Krotkov *et al.*, "The DARPA robotics challenge finals: Results and perspectives," *Journal of Field Robotics*, vol. 34, no. 2, pp. 229–240, 2017.

[25] R. R. Murphy *et al.*, "Search and rescue robotics," in *Springer Handbook of Robotics*, B. Siciliano and O. Khatib, Eds.   Berlin, Germany: Springer, 2008, pp. 1151–1173.

[26] K. Yoshida and B. Wilcox, "Space robots and systems," in *Springer Handbook of Robotics*, B. Siciliano and O. Khatib, Eds.   Berlin, Germany: Springer, 2008, pp. 1031–1063.

[27] A. Wang, J. Ramos, J. Mayo, W. Ubellacker, J. Cheung, and S. Kim, "The HERMES humanoid system: A platform for full-body teleoperation with balance feedback," in *IEEE-RAS International Conference on Humanoid Robots*, 2015, pp. 730–737.

[28] T. B. Sheridan, "Human-robot interaction: Status and challenges," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 58, no. 4, pp. 525–532, 2016.

[29] ——, *Telerobotics, automation, and human supervisory control.*   Cambridge, MA: MIT Press, 1992.

[30] H. Yamada, N. Tao, and Z. DingXuan, "Sketch and run: A stroke-based interface for home robots," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2009, pp. 36–40.

[31] D. Sakamoto, K. Honda, M. Inami, and T. Igarashi, "Construction tele-robot system with virtual reality," in *IEEE Conference on Robotics, Automation and Mechatronics*, 2008, pp. 197–200.

[32] S. Abidi, M. Williams, and B. Johnston, "Human pointing as a robot directive,"

in *ACM/IEEE International Conference on Human-Robot Interaction*, 2013, pp. 67–68.

[33] M. T. Wolf, C. Assad, M. T. Vernacchia, J. Fromm, and H. L. Jethani, "Gesture-based robot control with variable autonomy from the JPL BioSleeve," in *IEEE International Conference on Robotics and Automation*, 2013, pp. 1160–1165.

[34] S. Kasahara, R. Niiyama, V. Heun, and H. Ishi, "Spatially-aware embodied manipulation of actuated objects mediated by augmented reality," in *International Conference on Tangible, Embedded and Embodied Interaction*, 2013, pp. 223–228.

[35] S. Hashimoto, A. Ishida, M. Inami, and T. Igarashi, "Touchme: An augmented reality based remote robot manipulation," in *International Conference on Artificial Reality and Telexistence*, 2011, pp. 61–66.

[36] S. M. Chacko and V. Kapila, "Augmented reality as a medium for human-robot collaborative tasks," in *IEEE International Conference on Robot and Human Interactive Communication*, 2019, pp. 1–8.

[37] J. A. Frank, S. P. Krishnamoorthy, and V. Kapila, "Toward mobile mixed-reality interaction with multi-robot systems," *IEEE Robotics and Automation Letters*, vol. 2, no. 4, pp. 1901–1908, 2017.

[38] J. A. Frank, M. Moorhead, and V. Kapila, "Mobile mixed-reality interfaces that enhance human-robot interaction in shared spaces," *Frontiers in Robotics and AI*, vol. 4, 2017.

[39] Google, ARCore, developers.google.com, https://developers.google.com/ar (accessed: Mar. 7, 2024).

[40] M. Lanham, *Learn ARCore - Fundamentals of Google ARCore: Learn to build augmented reality apps for Android, Unity, and the web with Google ARCore 1.0.* Packt Publishing Ltd, 2018.

[41] S. Chacko, A. Granado, and V. Kapila, "An augmented reality framework for robotic tool-path teaching," in *Proceedings of CIRP Conference on Manufacturing Systems*, 2020, pp. 1218–1223.

[42] M. L. Ranney, V. Griffeth, and A. K. Jha, "Critical supply shortages – the need for ventilators and personal protective equipment during the covid-19 pandemic," *New England Journal of Medicine*, vol. 382, no. 13, pp. e41(1)–e41(3), 2020.

[43] K. Mizumoto, K. Kagaya, A. Zarebski, and G. Chowell, "Estimating the asymptomatic proportion of coronavirus disease 2019 (covid-19) cases on board the Diamond Princess cruise ship, yokohama, japan, 2020," *European Communicable Disease Bulletin*, vol. 25, no. 10, pp. 2 000 180–2 000 184, 2020.

[44] M. Klompas, C. A. Morris, J. Sinclair, M. Pearson, and E. S. Shenoy, "Universal masking in hospitals in the covid-19 era," *New England Journal of Medicine*, vol. 382, no. 21, pp. e63(1)–e63(3), 2020.

[45] J. Wang, M. Zhou, and F. Liu, "Reasons for healthcare workers becoming infected with novel Coronavirus disease 2019 (COVID-19) in China," *Journal of Hospital Infection*, vol. 105, no. 1, pp. 100–101, 2020.

[46] S. Naicker, C.-W. Yang, S.-J. Hwang, B.-C. Liu, J.-H. Chen, and V. Jha, "The novel Coronavirus 2019 epidemic and kidneys," *Kidney International*, vol. 97, no. 5, pp. 824–828, 2020.

[47] USRDS, "The United States renal data system 2022 annual data report," U.S. Department of Health and Human Services, 2022, Accessed: Mar. 7, 2024. [Online]. Available: https://usrds-adr.niddk.nih.gov/2022/introduction

[48] R. Abelson, "Dialysis patients face close-up risk from Coronavirus," *The New York Times*, 2020, last accessed 27 September 2020. [Online]. Available: https://www.nytimes.com/2020/04/11/health/dialysis-risk-coronavirus.html

[49] FDA, "Enforcement policy for non-invasive remote monitoring devices used to support patient monitoring during the Coronavirus Disease 2019 (COVID-19) public health emergency (revised)," Food and Drug Administration, 2020, Accessed: Mar. 7, 2024. [Online]. Available: https://www.fda.gov/media/136290/download

[50] C. Hale, "Companies roll out remote COVID-19 monitoring tools to free up hospital space," *The New York Times*, 2020, last accessed 27 September 2020. [Online]. Available: https://www.fiercebiotech.com/medtech/companies-roll-out-remote-covid-19-monitoring-tools-to-free-up-hospital-space

[51] H. Liao, T. Inomata, I. Sakuma, and T. Dohi, "3-D augmented reality for MRI-guided surgery using integral videography autostereoscopic image overlay," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 6, pp. 1476–1486, 2010.

[52] J. Wang *et al.*, "Augmented reality navigation with automatic marker-free

image registration using 3-D image overlay for dental surgery," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 4, pp. 1295–1304, 2014.

[53] K. Wada, T. Shibata, T. Saito, K. Sakamoto, and K. Tanie, "Psychological and social effects of one year robot assisted activity on elderly people at a health service facility for the aged," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2005, pp. 2785–2790.

[54] D. E. Logan *et al.*, "Social robots for hospitalized children," *Pediatrics*, vol. 144, no. 1, 2019.

[55] R. Murai *et al.*, "A novel visible light communication system for enhanced control of autonomous delivery robots in a hospital," in *IEEE/SICE International Symposium on System Integration*, 2012, pp. 510–516.

[56] Z. Li, P. Moran, Q. Dong, R. J. Shaw, and K. Hauser, "Development of a tele-nursing mobile manipulator for remote care-giving in quarantine areas," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 3581–3586.

[57] J. Craig, *Introduction to Robotics: Mechanics and Control.* New York, NY, USA: Pearson, 2018.

[58] Robotis, OpenManipulator-X, emanual.robotis.com, https://emanual.robotis.com/docs/en/platform/openmanipulator_x/overview/ (accessed: Mar. 7, 2024).

[59] M. W. Spong, S. Hutchinson, and M. Vidyasagar, *Robot modeling and control.* John Wiley and Sons, Inc., 2006.

[60] Y. Jianjun, S. Xiaojie, Z. Shiqi, Y. Ming, and Z. Xiaodong, "Monte Carlo method for searching functional workspace of an underwater manipulator," in *Chinese Control And Decision Conference*, 2018, pp. 6431–6435.

[61] Y. Guan and K. Yokoi, "Reachable space generation of a humanoid robot using the Monte Carlo method," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006, pp. 1984–1989.

[62] M. Mihelj *et al.*, *Robotics*. Cham, Switzerland: Springer, 2019.

[63] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[64] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, 2nd ed. Cambridge University Press, 2004.

[65] S. Garrido-Jurado, R. Munoz-Salinas, F. J. Madrid-Cuevas, and R. Medina-Carnicer, "Generation of fiducial marker dictionaries using mixed integer linear programming," *Pattern Recognition*, vol. 51, pp. 481–491, 2016.

[66] F. J. Romero-Ramirez, R. Muñoz-Salinas, and R. Medina-Carnicer, "Speeded up detection of squared fiducial markers," *Image and Vision Computing*, vol. 76, pp. 38–47, 2018.

[67] P. Corke, *Robotics, vision and control: Fundamental algorithms in MATLAB*, 1st ed. Springer, 2013.

[68] S. G. Hart, "NASA-task load index (NASA-TLX); 20 years later," in *Proceed-*

ings of the Human Factors and Ergonomics Society Annual Meeting, vol. 50, no. 9, 2006, pp. 904–908.

[69] J. Brooke, "SUS: A quick and dirty usability scale," in *Usability Evaluation In Industry.* London: Taylor and Francis, 1996, pp. 189–194.

[70] L. M. Koonin *et al.*, "Trends in the use of telehealth during the emergence of the COVID-19 pandemic–United States, January–March 2020," *Morbidity and Mortality Weekly Report*, vol. 69, no. 43, pp. 1595–1599, 2020.

[71] K. Olmstead, "Nearly half of americans use digital voice assistants, mostly on their smartphones," Pew Research Center, 2017, Accessed: Mar. 4, 2024. [Online]. Available: http://pewrsr.ch/2kquZ8H

[72] M. Duque, S. Pink, Y. Strengers, R. Martin, and L. Nicholls, "Automation, wellbeing, and digital voice assistants: Older people and Google devices," *Convergence: The International Journal of Research into New Media Technologies*, vol. 27, no. 5, pp. 1189–1206, 2021.

[73] A. L. Neves, C. Lygidakis, K. Hoedebecke, L. de Pinho-Costa, and A. Pilotto, "Digital health in an ageing world," in *The Role of Family Physicians in Older People Care*, J. Demurtas and N. Veronese, Eds. Cham, Switzerland: Springer, 2022, pp. 107–118.

[74] E. Sezgin, Y. Huang, U. Ramtekkar, , and S. Lin, "Readiness for voice assistants to support healthcare delivery during a health crisis and pandemic," *npj Digital Medicine*, vol. 3, no. 1, pp. 1–4, 2020.

[75] K. O'Brien, A. Liggett, V. Ramirez-Zohfeld, P. Sunkara, and L. A. Lindquist,

"Voice-controlled intelligent personal assistants to support aging in place," *Journal of the American Geriatrics Society*, vol. 68, no. 1, pp. 176–179, 2020.

[76] N. Bott *et al.*, "A protocol-driven, bedside digital conversational agent to support nurse teams and mitigate risks of hospitalization in older adults: Case control pre-post study," *Journal of Medical Internet Research*, vol. 21, no. 10, p. e13440, 2019.

[77] A. Rajkumar, F. Vulpi, S. R. Bethi, H. K. Wazir, P. Raghavan, and V. Kapila, "Wearable inertial sensors for range of motion assessment," *IEEE Sensors Journal*, vol. 20, no. 7, pp. 3777–3787, 2019.

[78] D. Xiang, H. Joo, , and Y. Sheikh, "Monocular total capture: Posing face, body, and hands in the wild," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 957–10 966.

[79] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "Blazepose: On-device real-time body pose tracking," *arXiv preprint arXiv:2006.10204*, 2020.

[80] D. H. Gates, L. S. Walters, J. Cowley, J. M. Wilken, and L. Resnik, "Range of motion requirements for upper-limb activities of daily living," *American Journal of Occupational Therapy*, vol. 70, no. 1, pp. 7 001 350 010p1–7 001 350 010p10, 2016.

[81] G. Pavlakos *et al.*, "Expressive body capture: 3D hands, face, and body from a single image," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 967–10 977.

[82] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *The Lancet*, vol. 327, no. 8476, pp. 307–310, 1986.

[83] I. Herawati, A. F. M. Ludin, M. Mutalazimah, I. Ishak, and N. M. F. Farah, "Breathing exercise for hypertensive patients: A scoping review," *Frontiers in Physiology*, vol. 14, pp. 70–85, 2023.

[84] V. C. Goessl, J. E. Curtiss, and S. G. Hofmann, "The effect of heart rate variability biofeedback training on stress and anxiety: A meta-analysis," *Psychological Medicine*, vol. 47, no. 15, pp. 2578–2586, 2017.

[85] M. R. Fu *et al.*, "Proactive approach to lymphedema risk reduction: A prospective study," *Annals of Surgical Oncology*, vol. 21, pp. 3481–3489, 2014.

[86] M. Thomas *et al.*, "Breathing exercises for asthma: A randomised controlled trial," *Thorax*, vol. 64, no. 1, pp. 55–61, 2009.

[87] A. Bruton *et al.*, "Physiotherapy breathing retraining for asthma: A randomised controlled trial," *The Lancet Respiratory Medicine*, vol. 6, no. 1, pp. 19–28, 2018.

[88] Y. Lu *et al.*, "Effects of home-based breathing exercises in subjects with COPD," *Respiratory Care*, vol. 65, no. 3, pp. 377–387, 2020.

[89] H. Liu, J. Allen, D. Zheng, and F. Chen, "Recent development of respiratory rate measurement technologies," *Physiological Measurement*, vol. 40, no. 7, p. 07TR01, 2019.

[90] C.-H. Shih, N. Tomita, Y. X. Lukic, Á. H. Reguera, E. Fleisch, and T. Kowatsch, "Breeze: Smartphone-based acoustic real-time detection of breathing phases for a gamified biofeedback breathing training," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 4, pp. 1–30, 2019.

[91] E. P. Doheny *et al.*, "Estimation of respiratory rate and exhale duration using audio signals recorded by smartphone microphones," *Biomedical Signal Processing and Control*, vol. 80, no. 1, pp. 1–30, 2023.

[92] C. Romano *et al.*, "Respiratory rate estimation during walking and running using breathing sounds recorded with a microphone," *Biosensors*, vol. 13, no. 6, pp. 637–652, 2023.

[93] A. Kumar, V. Mitra, C. Oliver, A. Ullal, M. Biddulph, and I. Mance, "Estimating respiratory rate from breath audio obtained through wearable microphones," in *International Conference of the IEEE Engineering in Medicine & Biology Society*, 2021, pp. 7310–7315.

[94] T. Röddiger, D. Wolffram, D. Laubenstein, M. Budde, and M. Beigl, "Towards respiration rate monitoring using an in-ear headphone inertial measurement unit," in *Proceedings of the International Workshop on Earable Computing*, 2019, pp. 48–53.

[95] M. M. Rahman *et al.*, "Detecting physiological responses using multimodal earbud sensors," in *International Conference of the IEEE Engineering in Medicine & Biology Society*, 2022, pp. 01–05.

[96] B. Islam *et al.*, "Breathtrack: Detecting regular breathing phases from

unannotated acoustic data captured by a smartphone," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 3, pp. 1–22, 2021.

[97] H. K. Wazir, S. R. Bethi, A. R. Kumar, F. Caruso, and V. Kapila, "A wearable pendant sensor to monitor compliance with range of motion lymphatic health exercise," in *International Conference of the IEEE Engineering in Medicine & Biology Society*, 2020, pp. 4588–4591.

[98] D. Bhattacharya *et al.*, "Coswara: A respiratory sounds and symptoms dataset for remote screening of SARS-CoV-2 infection," *Scientific Data*, vol. 10, no. 1, pp. 397–407, 2023.

[99] B. M. Rocha *et al.*, "A respiratory sound database for the development of automated classification," in *International Conference on Biomedical and Health Informatics*, 2017, pp. 33–37.

[100] Y.-Y. Yang *et al.*, "TorchAudio: Building blocks for audio and speech processing," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2021, pp. 6982–6986.

[101] J. R. Wilmoth, D. Bas, S. Mukherjee, and N. Hanif, "World social report 2023: Leaving no one behind in an ageing world," UNDESA, 2023, Accessed: Mar. 7, 2024. [Online]. Available: https://social.desa.un.org/sites/default/files/publications/2023-02/WorldSocialReport2023.pdf

[102] F. Adib, H. Mao, Z. Kabelac, D. Katabi, and R. C. Miller, "Smart homes that monitor breathing and heart rate," in *Annual ACM conference on human factors in computing systems*, 2015, pp. 837–846.

[103] M. Esposito, A. Minutolo, R. Megna, M. Forastiere, M. Magliulo, and G. De Pietro, "A smart mobile, self-configuring, context-aware architecture for personal health monitoring," *Engineering Applications of Artificial Intelligence*, vol. 67, pp. 136–156, 2018.

[104] C. Yang, W. Wang, F. Li, and D. Yang, "A sustainable, interactive elderly healthcare system for nursing homes: An interdisciplinary design," *Sustainability*, vol. 14, no. 7, pp. 4204–4224, 2022.

[105] J. C.-W. Cheung, E. W.-C. Tam, A. H.-Y. Mak, T. T.-C. Chan, and Y.-P. Zheng, "A night-time monitoring system (enightlog) to prevent elderly wandering in hostels: A three-month field study," *International Journal of Environmental Research and Public Health*, vol. 19, no. 4, pp. 2103–2118, 2022.

[106] M. Saleh and R. L. B. Jeannès, "Elderly fall detection using wearable sensors: A low cost highly accurate algorithm," *IEEE Sensors Journal*, vol. 19, no. 8, pp. 3156–3164, 2019.

[107] J.-Y. Kim, C.-H. Chu, and M.-S. Kang, "IoT-based unobtrusive sensing for sleep quality monitoring and assessment," *IEEE Sensors Journal*, vol. 21, no. 3, pp. 3799–3809, 2020.

[108] H. K. Wazir, K. Gaikwad, and V. Kapila, "Range of motion assessment using a digital voice assistant," in *IEEE International Conference on Engineering in Medicine & Biology Society*, 2022, pp. 2577–2580.

[109] A. RajKumar, F. Vulpi, S. R. Bethi, H. K. Wazir, P. Raghavan, and V. Kapila,

"Wearable inertial sensors for range of motion assessment," *IEEE Sensors Journal*, vol. 20, no. 7, pp. 3777–3787, 2020.

[110] G. Sebestyen, I. Stoica, and A. Hangan, "Human activity recognition and monitoring for elderly people," in *IEEE International Conference on Intelligent Computer Communication and Processing*, 2016, pp. 341–347.

[111] A. H. Nasution and S. Emmanuel, "Intelligent video surveillance for monitoring elderly in home environments," in *IEEE 9th Workshop on Multimedia Signal Processing*. IEEE, 2007, pp. 203–206.

[112] A. T.-Y. Chen, M. Biglari-Abhari, I. Kevin, and K. Wang, "Investigating fast re-identification for multi-camera indoor person tracking," *Computers & Electrical Engineering*, vol. 77, pp. 273–288, 2019.

[113] R. Naccarelli, S. Casaccia, and G. M. Revel, "The problem of monitoring activities of older people in multi-resident scenarios: An innovative and non-invasive measurement system based on wearables and PIR sensors," *Sensors*, vol. 22, no. 9, pp. 3472–3494, 2022.

[114] H. Kelly, "For seniors using tech to age in place, surveillance can be the price of independence," *The Washington Post*, November 2021, (accessed: Aug. 15, 2023). [Online]. Available: https://www.washingtonpost.com/technology/2021/11/19/seniors-smart-home-privacy/

[115] W. Kang and Y. Han, "SmartPDR: Smartphone-based pedestrian dead reckoning for indoor localization," *IEEE Sensors Journal*, vol. 15, no. 5, pp. 2906–2916, 2015.

[116] B. Khalili, R. Ali Abbaspour, A. Chehreghan, and N. Vesali, "A context-aware smartphone-based 3D indoor positioning using pedestrian dead reckoning," *Sensors*, vol. 22, no. 24, pp. 9968–9996, 2020.

[117] Q. Tian, Z. Salcic, K. I.-K. Wang, and Y. Pan, "A multi-mode dead reckoning system for pedestrian tracking using smartphones," *IEEE Sensors Journal*, vol. 16, no. 7, pp. 2079–2093, 2016.

[118] L. Árvai, "Application of smartwatches in elderly care with indoor localization functionality," *International Journal of Interactive Mobile Technologies*, vol. 15, no. 5, pp. 174–186, 2021.

[119] A. Poulose, O. S. Eyobu, and D. S. Han, "An indoor position-estimation algorithm using smartphone IMU sensor data," *IEEE Access*, vol. 7, pp. 11 165–11 177, 2019.

[120] R. Montoliu, E. Sansano, A. Gascó, O. Belmonte, and A. Caballer, "Indoor positioning for monitoring older adults at home: Wi-Fi and BLE technologies in real scenarios," *Electronics*, vol. 9, no. 5, pp. 728–747, 2020.

[121] S.-C. Kim, Y.-S. Jeong, and S.-O. Park, "RFID-based indoor location tracking to ensure the safety of the elderly in smart home environments," *Personal and Ubiquitous Computing*, vol. 17, pp. 1699–1707, 2013.

[122] Y. Zhang and L. Duan, "Toward elderly care: A phase-difference-of-arrival assisted ultra-wideband positioning method in smart home," *IEEE Access*, vol. 8, pp. 139 387–139 395, 2020.

[123] T. F. Sanam and H. Godrich, "A multi-view discriminant learning approach

for indoor localization using amplitude and phase features of CSI," *IEEE Access*, vol. 8, pp. 59 947–59 959, 2020.

[124] K. Wu, J. Xiao, Y. Yi, D. Chen, X. Luo, and M. N. Lionel, "CSI-based indoor localization," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 7, pp. 1300–1309, 2013.

[125] J.-H. Seong, S.-H. Lee, W.-Y. Kim, and D.-H. Seo, "High-precision RTT-based indoor positioning system using RCDN and RPN," *Sensors*, vol. 21, no. 11, pp. 3701–3716, 2021.

[126] G. Guo *et al.*, "A robust integration platform of Wi-Fi RTT, RSS signal, and MEMS-IMU for locating commercial smartphone indoors," *IEEE Internet of Things Journal*, vol. 9, no. 17, pp. 16 322–16 331, 2022.

[127] K. Han, S. M. Yu, S.-L. Kim, and S.-W. Ko, "Exploiting user mobility for WiFi RTT positioning: A geometric approach," *IEEE Internet of Things Journal*, vol. 8, no. 19, pp. 14 589–14 606, 2021.

[128] G. Guo, R. Chen, F. Ye, X. Peng, Z. Liu, and Y. Pan, "Indoor smartphone localization: A hybrid WiFi RTT-RSS ranging approach," *IEEE Access*, vol. 7, pp. 176 767–176 781, 2019.

[129] M. Sun, Y. Wang, S. Xu, H. Qi, and X. Hu, "Indoor positioning tightly coupled Wi-Fi FTM ranging and PDR based on the extended Kalman filter for smartphones," *IEEE Access*, vol. 8, pp. 49 671–49 684, 2020.

[130] X. Wang, X. Wang, S. Mao, J. Zhang, S. C. Periaswamy, and J. Patton, "Indoor radio map construction and localization with deep Gaussian processes," *IEEE Internet of Things Journal*, vol. 7, no. 11, pp. 11 238–11 249, 2020.

[131] "IEEE standard for information technology–Telecommunications and information exchange between systems local and metropolitan area networks–Specific requirements - Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications," IEEE Std 802.11-2016, 2016.

[132] A. Norrdine, "An algebraic solution to the multilateration problem," in *International Conference on Indoor Positioning and Indoor Navigation*, 2012.

[133] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3-4, pp. 591–611, 1965.

[134] M. Bullmann, T. Fetzer, M. Ebner, S. Kastner, F. Deinzer, and M. Grzegorzek, "Data driven sensor model for wi-fi fine timing measurement," in *IEEE International Conference on Indoor Positioning and Indoor Navigation*, 2022, pp. 1–8.

[135] B. Siciliano and O. Khatib, *Springer Handbook of Robotics*, B. Siciliano and O. Khatib, Eds. Berlin, Germany: Springer, 2008.