Evaluation of Augmented Reality and Wearable Sensors to Assess Neurovestibular and Sensorimotor Performance in Astronauts for Extravehicular Activity Readiness

by

Hannah M. Weiss

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in the University of Michigan
2024

Doctoral Committee:

Associate Professor Leia Stirling, Chair
Professor Nadine Sarter
Professor Kathleen Sienko
Dr. Scott Wood, NASA

Hannah M. Weiss
hanweiss@umich.edu
ORCID iD: 0000-0002-2504-2763

DEDICATION

With love and utmost respect, I dedicate this dissertation to my parents, Tamara and Karl Weiss, whose unwavering belief in my potential has fueled my determination to reach for the stars. I am profoundly grateful for the values you instilled in me, the opportunities you provided, and the sacrifices you have made to support my success in all aspects of life. Ad astra per aspera.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to the following individuals and groups who have played instrumental roles in the completion of this dissertation:

Dr. Leia Stirling, your guidance, support, and steadfast belief in my abilities have been invaluable throughout this journey. Your mentorship has not only enriched my academic experience but also shaped me as a researcher, mentor, and future leader in my field, instilling in me the qualities of dedication, critical thinking, and the importance of passing knowledge to the next generation of scholars. I am truly grateful for your wisdom and the freedom you provided me to follow my passion for spaceflight research. I consider myself profoundly fortunate to have had the opportunity to work with such an outstanding researcher, professor, and mentor. Your guidance and personal example will always serve as a source of inspiration for me.

I extend my sincere thanks to the members of my dissertation committee, Dr. Nadine Sarter, Dr. Kathleen Sienko, and Dr. Scott Wood for their invaluable feedback, insightful suggestions, and dedication to ensuring the quality of my work. Your expertise has greatly contributed to the refinement of this dissertation.

To my parents, brothers, aunt, and sister-in-law your unwavering love, encouragement, and sacrifices have been the foundation upon which I have built my pursuits. Your unshakable belief in my abilities has been the driving force behind my ambitions in both athletics and academia, and I hold deep appreciation for each of you. This dissertation is a testament to the faith you had in me, and I hope this brings you as much pride and joy as it does to me.

To Jacob, thank you for your patience, understanding, and constant support from several states away. Your love and encouragement have been a steady source of inspiration and comfort throughout this demanding journey. There are no words that can express my gratitude for all you have done for me.

To my graduate colleagues and friends, who stood by me during both the highs and lows of this process, thank you for your support, countless late-night conversations, and laughter. Your friendship has been irreplaceable. Additionally, thank you to all who generously participated in my experimental research studies.

I would also like to acknowledge Olof Minto and Edwin Tang from IOE for their exceptional support in the construction of an adjustable wall adapter for the touch monitor used in the research and for IT troubleshooting to enable device connectivity. Your contributions were integral to the success of my research.

A special thank you goes to the undergraduate and graduate students who not only assisted in my research but also contributed significantly to advancing my mentoring abilities.

I extend my heartfelt appreciation to my mentors across my academic and professional career, Dr. Jill McNitt-Gray, and Dr. Han Kim, for your guidance and the opportunities you provided for my growth as a researcher. Your wisdom, expertise, and advice have been instrumental in directing my academic journey, spanning from my undergraduate studies to conducting research at NASA and extending into my pursuit of graduate studies.

Finally, I am grateful to the NASA Neuroscience Lab for their support as I navigated the dual responsibilities of completing my dissertation and taking on a new role in the lab, actively contributing to existing and novel sensorimotor research endeavors.

This dissertation would not have been possible without the collective support, encouragement, and belief in me from all of you. I am truly fortunate to have such wonderful individuals in my life. Thank you for being a part of this journey and for helping me reach this momentous milestone.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

TABLE

LIST OF APPENDICES

APPENDIX

LIST OF ACRONYMS

AR Augmented Reality

AURORA Augmented Reality Operations Readiness Assessment

EVA Extravehicular Activity

FSST Four Square Step Test

HIDH Human Integration Design Handbook

IET Ingress and Egress Task

ID Index of Difficulty

ISO International Organization for Standardization

IMU Inertial Measurement Unit

LEO Low Earth Orbit

OWT Obstacle Weave Task

SANS Spaceflight-associated neuro-ocular syndrome

TLX Task Load Index

TS Touch Screen

TW Tandem Walk

VOR Vestibulo-Ocular Reflex

XR Extended Reality

ABSTRACT

As next-generation space exploration missions require increased autonomy from astronaut crews, real-time diagnostics of astronaut health and performance are essential for mission operations, especially for determining Extravehicular Activity (EVA) readiness. Exposure to microgravity leads to changes in astronaut physiology such as loss of bone density and muscle mass, cardiovascular deconditioning, and the reorganization of sensorimotor integration that require distinct adaptation timelines. As astronauts transition from microgravity to gravity-rich environments (e.g., the Moon or Mars), sensorimotor impairment may affect their ability to perform mission-critical tasks such as landing procedures, capsule egress, and early EVA. To ensure mission success and astronaut safety, it is essential to identify sensorimotor impairment during missions. At present, no flight-ready tools are available for astronauts to assess their sensorimotor impairment throughout a mission, especially tools for which expert assistance is not needed. Current Earth-based assessments require extensive resources to administer and skilled clinicians to score and interpret the data. The development of lightweight, space-conscious assessments for astronauts transitioning between gravity environments is crucial to the success of future exploration missions. An Augmented Reality (AR) standalone system may be a viable solution by allowing holographic visual cueing to replace physical objects used in traditional assessments. AR headsets are easily deployable, enable the user to see their physical environment for safety in confined spaces, provide flexibility for new software integration, and are multifunctional for other mission tasks such as procedural guidance. The research focused on the development, usability testing, and evaluation of the Augmented Reality Operations Readiness Assessment (AURORA) to assess neurovestibular and sensorimotor performance with holographic visual cueing and integrated inertial measurement units.

The human-in-the-loop usability testing demonstrated positive perceptions of usability across two diverse age groups spanning current and future astronaut age demographics. Although usability concerns were identified with a subset of users, future design recommendations, such as proximity lighting for improved depth perception, are provided to promote self-administration for all users. While performance for the sensorimotor and vestibular assessments differed when compared to Earth-based standards and physical assessments, the research demonstrated the potential of AR as an assessment tool with the benefit of embedded sensors and onboard computing capabilities. The AR tool effectively quantified changes in hand-eye coordination performance as measured by accuracy, precision, error rate, throughput, and movement time with varying task difficulty. Users were able to successfully complete the tasks, yielding meaningful performance measures. When compared to the physical environment, AR led to altered user strategies, predominantly marked by increased task time, reduced head and torso movements, and increased head pitch amplitude aimed at enhancing hologram visibility due to the restricted field of view of the headset. User performance could have also been attributed to the novel environment for most users. AR-induced strategies resembled compensatory responses observed in astronauts after spaceflight and in vestibular patients, suggesting that AR would not exacerbate symptoms of motion sickness. These findings address gaps outlined in NASA's Human Research Roadmaps and offer guidance for AR design within NASA's Human Integration Design Handbook. The research advances AR's potential as a standalone assessment tool in evaluating astronaut performance throughout missions while considering spaceflight constraints. The implication of this research extends into applications for both the aerospace industry and the medical field as it relates to aging populations and individuals with balance and sensorimotor disorders.

CHAPTER 1

Introduction

1.1 Motivation

Vestibular and sensorimotor alterations associated with spaceflight are a potential risk to future long-duration missions beyond low Earth orbit as crew members will be required to operate more autonomously and sustain longer extravehicular activity (EVA) missions [7]. The risk is most significant during and immediately following the transitions between distinct gravity environments (i.e., microgravity and gravity-rich environments such as Earth, Lunar, or Martian surfaces) [8] (Figure 1.1). Earth's gravitation, which is usually sensed by the vestibular organs in the inner ear (i.e., the semicircular canals and otolith organs) [8, 9], is effectively absent in microgravity, disrupting normal vestibular function and eliciting sensory reinterpretation, which is maladaptive upon return to gravity environments [10]. Decrements in postural control [11], locomotion [12], and fine motor function can lead to motion sickness [13], spatial disorientation [14], and deficits in hand-eye coordination [15, 16] that may negatively impact critical mission tasks such as manual vehicle control, extravehicular activities (EVA), and capsule ingress and egress [17, 18]. Future human spaceflight exploration will present unique challenges to astronaut crews as the distance from Earth and mission duration increase. Currently, there are no fitness-for-duty assessments concerning neurovestibular and sensorimotor performance in mission operations on the International Space Station. New technologies are required to identify, diagnose, and reduce risks associated with space travel on astronauts' performance and determine when countermeasures may be necessary. Robust evaluation tools and devices are needed to perform EVA readiness assessments and to lay the foundation for countermeasure development and evaluation. The solution requires technologies and hardware systems that are validated, observe mission resource constraints, are flexible for diagnostic capabilities and risk mitigation, and ultimately systems that are informed by human factors principles for effective utilization.

Missions landing on the lunar or Martian surface will not receive the same levels of operational support available after landing on Earth, such as rescue teams and medical interventions [17]. It is crucial that new diagnostic systems provide relevant information to assess astronaut capability in these more autonomous operational settings. Current postflight Earth-based clinical measures of astronauts' neurovestibular balance performance are captured using the Sensory Organization Test battery with computerized dynamic posturography [17, 19] and functional mobility tests [20, 4], both of which require extensive resources and personnel to conduct. In clinical practice, various screening tests are implemented for vestibular disorders and elderly populations, including tandem walking balance [21], static balance with force platforms [22], the Star Excursion Balance Test [23], and the Berg Balance Scale [24].

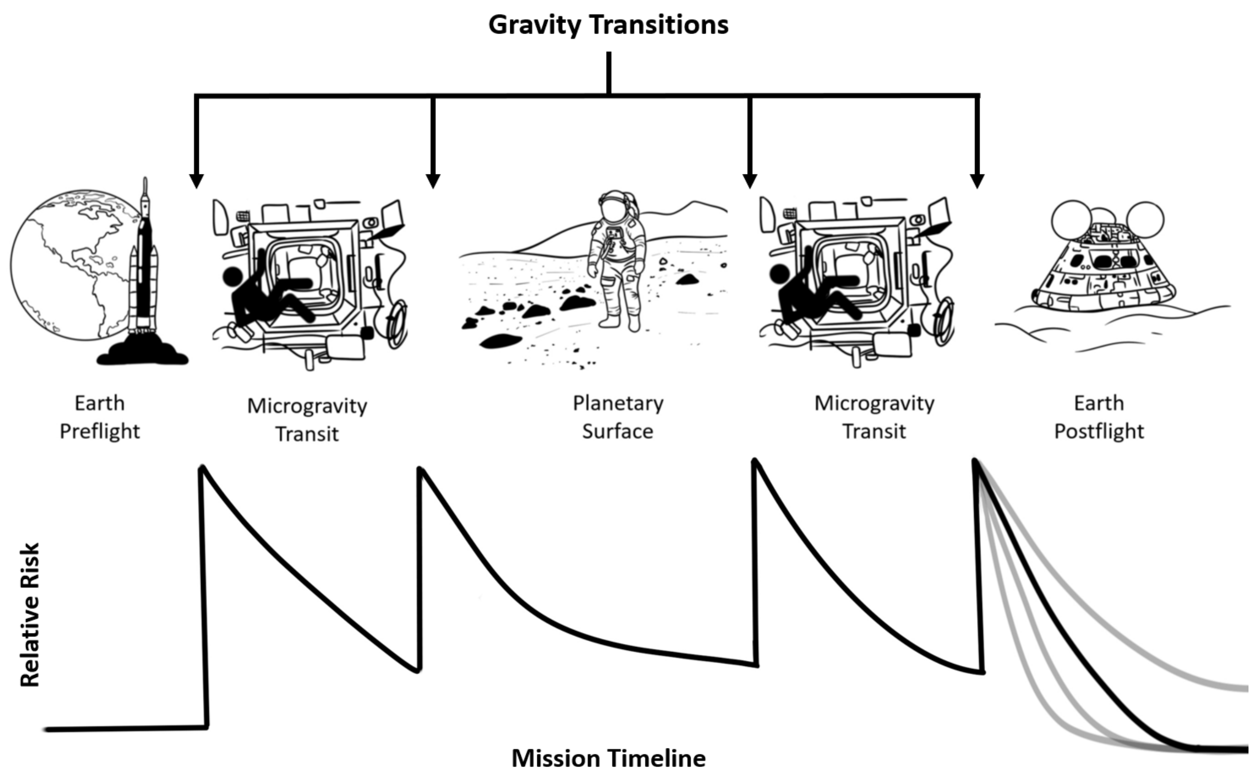

Figure 1.1: Relative sensorimotor risk progression for long duration exploration class missions. A comparison of different recovery curves based on mission duration length can be seen in the shaded regions to the right, where shorter missions exhibit faster recovery curves. This image was drawn with reference to [1] and adapted to include different recovery curves following conversations with Dr. Scott Wood.

Many of these clinical gold-standard balance measures require skilled clinicians to score and interpret the data. In spaceflight operations, volumetric and resource constraints must be considered when designing novel diagnostic tools and countermeasure devices. Volumetric capacity is constrained by the habitat or spacecraft module, thereby influencing the usage and stowage of assessment tools. Resource constraints include the demands required by the crew to set up, use, and stow the devices and the energy requirements for device operation. Additional constraints include the limitation of mass for launch considerations and the tool's ability to be used in microgravity and reduced-gravity environments to support evaluating physical performance throughout mission timelines. New assessment tools necessitate real-time feedback and data processing for the determination of the crew's physical fitness and capability to perform EVAs safely. As the validation of fitness for duty standards before landing on novel planetary surfaces may be limited, the tools will also have to be sensitive enough to track improvements from initial assessment conditions, such as soon after landing.

An Augmented Reality (AR) standalone system may be a viable solution that addresses many space-relevant requirements by allowing holographic visual cueing to replace physical objects used in traditional assessments. AR technology blends real-world elements with computer-generated content through multiple sensory modalities (e.g., auditory and visual). While balance tasks could be performed without AR using external cueing, AR systems enable evaluating performance with naturally synchronized task cueing with the benefit of embedded sensors to quantify human motion. The embedded inertial measurement unit consists of an accelerometer, gyroscope, and magnetometer, all of which can be leveraged to quantify postural control throughout balance assessments. The Microsoft HoloLens 2, the current state-of-the-art technology and headset architecture, features four visible light and two infrared cameras to support hand and eye tracking capabilities, respectively. The depth and environment sensing cameras embedded in augmented reality headsets can be leveraged to evaluate sensorimotor performance, such as hand-eye coordination, as well as facilitate user interactions. AR headsets are easily deployable, enable users to see their physical environment for safety in confined spaces, provide flexibility for new software integration, and are multifunctional for other mission tasks such as procedural guidance [25]. AR headsets feature built-in computational processing that can be used in conjunction with embedded sensors and additional wearable sensors to measure performance.

This thesis is a novel investigation of augmented reality and wearable sensors to support neurovestibular and sensorimotor assessments in astronauts for future long-duration space missions. The work evaluates the following overarching research questions:

1. Can augmented reality be used as a standalone assessment tool for neurovestibular and sensorimotor evaluations in astronauts?

2. For self-assessment of physical abilities, what usability features for design and user interactions need to be considered?

3. What impact does the use of holographic objects in assessments traditionally performed with physical objects have on overall performance?

In the following sections, we will discuss space physiology and the effects of microgravity on the human body with respect to vestibular and sensorimotor performance (Section 1.2.1), thereby motivating the need for in-flight assessment tools to enable physiological monitoring. The literature review presents current Earth-based neurovestibular (Section 1.2.2) and sensorimotor assessments (Section 1.2.3) and current augmented reality applications for physiological evaluations (Section 1.2.4). In this literature review, we discuss various spaceflight capabilities and investigate how earth-based assessments may inform future in-flight assessment tools. Additionally, the literature review examines the capabilities of augmented reality systems for standalone assessment tools as well as specific usability heuristics for the design of new systems. Lastly, the final section summarizes the specific research questions of the thesis (Section 1.3), which relate to developing a standalone assessment tool that adheres to usability heuristics and the evaluation of the tool for assessments. Section 1.3 also presents the alignment of the thesis work with NASA Human Research Roadmap gaps to further contextualize the research goals and intended research contributions (Section 1.3.2).

1.2 Background

1.2.1 Space Physiology

In our everyday experiences, 1g (i.e., the acceleration due to gravity on Earth) is ubiquitous, making it difficult to acknowledge its significance as the critical stimulus contributing to physiological baselines and homeostasis, i.e., the maintenance and steady-state or equilibrium of the body's physiological processes. Exposure to microgravity (μg) leading to weightlessness presents a host of unique challenges and systemic effects on the human body. Several of the physiological impacts of microgravity exposure mirror those of aging and the complications that come with sedentary living. Prolonged exposure to microgravity can affect the human body systems, including the eyes [26, 27], brain [28, 29], vestibular [30, 31], cardiovascular [27, 32], musculoskeletal [27, 33], and immune systems [34, 35]. Motion control, vision, spatial orientation, motion sickness, and ambulation are all affected by changes in the body systems due to microgravity exposure [36]. Although many of the health concerns associated with time spent in low Earth orbit (LEO) persist during long-distance space travel, there are also unique challenges associated with the Lunar and Martian surfaces because of their distinct environmental characteristics, such as varying radiation levels [37].

The sensorimotor system governs the body's motor functions through a network of sensory organs, such as eyes, ears, skin, and the nervous systems [38]. Humans perceive and respond to the external environment and internal changes through the integration of these sensory organs and the central nervous system. A specialized sensorimotor capability, the neurovestibular system, maintains balance, posture, and eye gaze stabilization [39]. The following sections outline the effects of microgravity on the body systems that govern neurovestibular and sensorimotor functioning that support whole-body postural control and fine motor control, respectively.

1.2.1.1 Effects of Microgravity on Neurovestibular Performance

The neurovestibular system is a complex network of systems and integrative centers (e.g., vestibular and cerebellar) that serve to coordinate and consolidate multiple adaptive processes among various neural systems. To stabilize the body to control balance and smooth movements, the integrative centers of the brain integrate input from the eyes, inner ear, and proprioceptive sensors in the joints and muscles [39]. In addition to connecting to the brainstem, cerebellum, and somatic sensory cortex, the vestibular nuclei directly innervate motor neurons to control extraocular, cervical, and postural muscles [39]. As a whole system, the neurovestibular complex regulates postural control and reflexes, head orientation, and the stabilization of eye gaze during head movements through the vestibulo-ocular reflex (VOR) [40]. Alterations or damage to the vestibular system may adversely affect balance, sense of orientation within a given environment, and gaze stabilization, which may impact daily activities.

In microgravity, the input from the vestibular system and cerebellar regions are altered by the lack of chronic gravity-induced acceleration potentially leading to space motion sickness [41], especially during early exposure to changes in gravitational forces. Microgravity directly affects the vestibular system's peripheral or outermost organ, the inner ear. The structures of the inner ear function as inertial guidance systems and miniaturized accelerometers continually transmitting information to the brain stem. The vestibular labyrinth, which supports the perception of the body's orientation and motion relative to the gravitational vertical, consists of three semicircular canals and the otolith organs (the utricle and saccule) [40] (Figure 1.2). High-frequency motions are best sensed in the semicircular canals, while low-frequency motions are best sensed in the otolith organs [41].

The semicircular canals are oriented along three planes of movement that are orthogonal to one another (Figure 1.2). The semicircular canals support the sense of one's orientation and are sensitive to angular acceleration from head rotations, either self-induced or externally induced. These three planes of rotation are referred to as pitch (tilting of the head up and down; nodding ”yes”), roll (tilting of the head left or right), and yaw (lateral movement left and right; shaking the head ”no”). A fluid called endolymph fills the semicircular canals. The fluid moves with head rotations and flows into the ampulla, an enlarged portion of the canal with receptor cells for detecting movement [39] (Figure 1.2). During head movement, distortion of the receptor hair cells within the ampulla stimulates the vestibulocochlear nerve, which transmits information about the rotations of the head to the vestibular nuclei in the brain stem and cerebellum [39].

The otolith organs, i.e., the utricle and saccule, are sensitive to linear accelerations and the po-

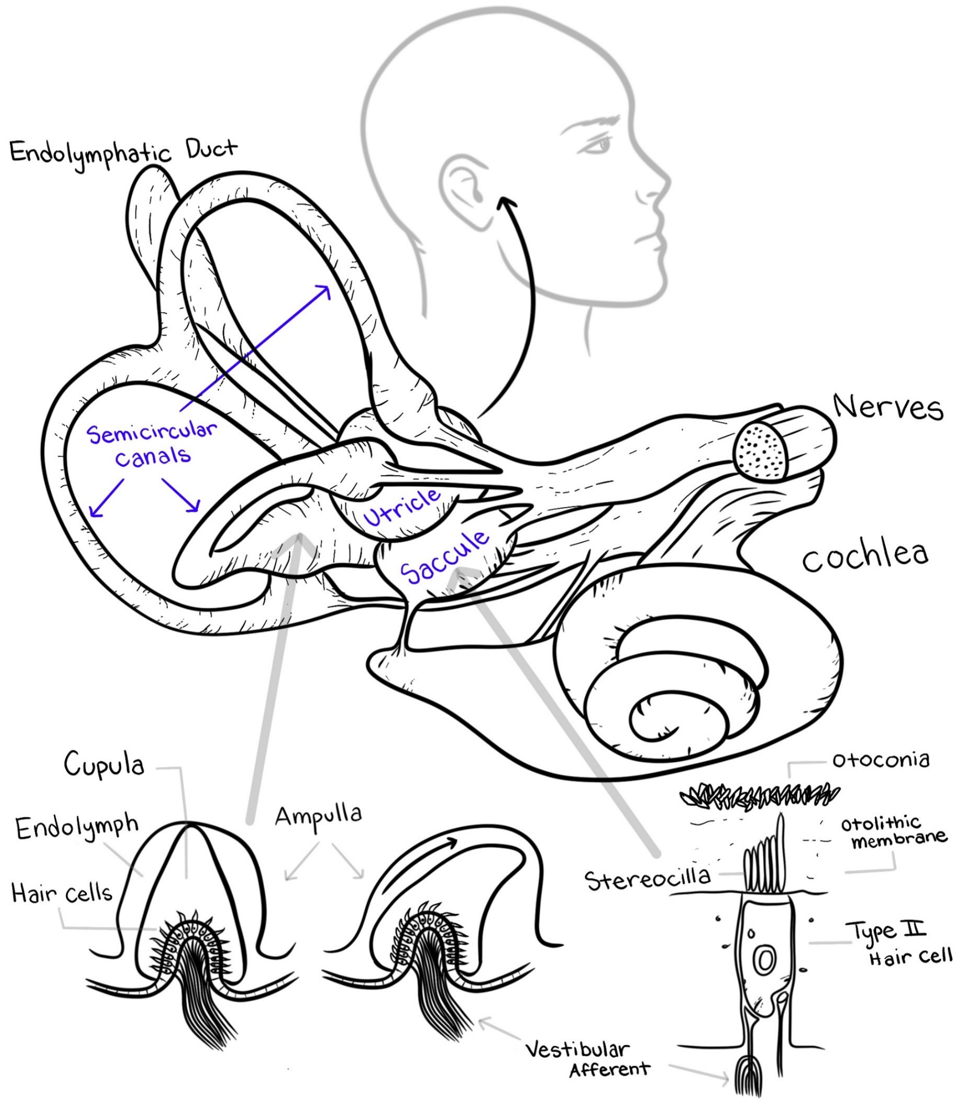

Figure 1.2: Located in the inner ear, the vestibular system comprises the semicircular canals and otolith organs (utricle and saccule). This figure was drawn by the author with reference to original artwork by Max Brödel [2]

sition or tilt of the head relative to the gravitational axis in static head positions [39]. While the utricle detects movement in the horizontal plane, the saccule is specialized to detect movement in the vertical plane. Both organs comprise sensory epithelium with hair cells and otolithic membranes with embedded calcium carbonate crystals called otoconia [39] (Figure 1.2). When the head tilts, gravity causes the otolithic membrane to shift relative to the epithelium resulting in a shearing motion that displaces the hair cells and generates receptor potentials in the hair cells to transmit information to the brain. Shearing motions of the otolithic membrane are also observed when the head undergoes linear accelerations [39]. Having a greater relative mass, the otolithic membrane tends to temporarily lag behind the epithelium during linear accelerations [39].

Throughout spaceflight, the vestibular labyrinth is primarily affected by the unloading of the otolith organs. On Earth, gravitational acceleration acts upon the epithelium in the saccule and utricle as the head tilts; however, in space, the otolith organs receive the same static stimulus irrespective of the body's orientation [42]. When the head is tilted in weightlessness, the otoliths are unaffected by a traditionally present steady pull. In this case, no signal is sent to the brain to indicate the head has changed orientation. Due to the inertia of the otolith organs, dynamic stimuli are still sensed by the otolith organs when one moves their head side to side, front to back, or up and down [42]. However, head movements, especially in the pitch and roll planes, are the predominant provocative stimuli for space motion sickness [43, 44]. These movements result in discordant cues being transmitted to the central nervous system for integration [31]. It is widely accepted that space motion sickness results from sensory conflict arising from mismatches between expected and actual sensory inputs from the vestibular system [13, 31, 45, 46]. It is especially difficult to resolve sensory conflict when there are incongruous inputs from other perceptual systems, such as the visual and proprioceptive senses. Interpreting sensory cues from otolith organs during and following g-transitions is complicated by the ambiguity between tilt and translation motions over a wide frequency range. When the sensory mismatch persists, motion sickness can result due to a loss of environmental calibration.

It is common for astronauts to experience motion sickness during their first two to three days in microgravity and upon returning to Earth as they adjust to new environmental constraints [13]. There is considerable plasticity within the balance control system, as evidenced by the recovery from space motion sickness later in mission timelines due to continuous weightlessness exposure [41]. Microgravity adaptation is believed to occur through the reinterpretation of sensory cues by the brain and the re-weighting of sensory input, particularly from the vestibular system's otolith organs. In this approach, sensory cues are reinterpreted in such a way as to reduce the influence of the otolith signals and provide greater emphasis on visual and proprioceptive signals [47, 42, 31]. However, this new re-weighting paradigm becomes maladaptive upon reentry to a gravity environment such as Earth, Lunar, or Martian surfaces. The neurovestibular system must readapt to the gravitational acceleration of the planetary body. In this readaptation period, there may be difficulties in physiological capabilities influenced by the vestibular system, such as gaze control, posture, and locomotion [36]. These decrements in performance may significantly impact early mission operations upon re-entry to a gravitational state [48]. Ultimately, the timelines for adaption and recovery are a function of time duration in microgravity exposure and have been shown to differ depending on the individual [36]. As such, vestibular assessments and evaluative tools are essential to incorporate into mission operations as long-duration crews are most susceptible to impairments leading to increased fall risk and injury.

Similar challenges to functional balance are experienced by astronauts and neurovestibular patients, resulting in disorientation, imbalance, disruption of gaze, and motion sickness symptoms [3]. Observations of vestibular patient populations suggest the risk of falls increases with the degree of vestibular hypofunction [49]. The analogous symptoms experienced by both populations pose an increased risk of falls and injuries, regardless of whether they are weightlessness-induced or pathology-induced vestibular disease, injury, or aging. In addition to vestibular impairments, astronauts experience skeletal decrements associated with spaceflight that may increase the risk of injury from vestibular-induced falls, thereby reducing long-term flight capabilities. For the skeletal system, there is currently a significant and unresolved health risk associated with microgravity-induced bone loss, as it increases the possibility of irreversible changes that weaken skeletal integrity and the onset of fracture injuries [50].

In order to ensure EVA readiness and reduce the risk of injuries from falls, assessment tools need to be integrated into mission operations and be available in flight to evaluate vestibular-induced postural instability during operationally-relevant tasks. Section 1.2.2 describes current earth-based vestibular assessments that can inform new assessment tools that are suitable for spaceflight constraints.

1.2.1.2 Effects of Microgravity on Sensorimotor Performance

The sensorimotor system regulates the coordination of motor movements, such as the execution of fine motor skills through integrating the visual system and proprioceptive sensors. Manual dexterity and hand-eye coordination are essential in performing mission tasks effectively, especially those that require human-system integration, such as capsule interfaces, telerobotic interfaces, and spacesuit systems. Due to changes in vision and muscular control with microgravity exposure, the ability to operate small controls accurately is affected, raising concerns for future long-duration crew members who will be required to perform critical tasks accurately [48, 16].

Spaceflight-associated neuro-ocular syndrome (SANS), previously termed visual impairment and intracranial pressure (VIIP) syndrome, represents the record of ophthalmologic and neurologic findings in astronaut populations after prolonged spaceflight [51]. Microgravity-induced ocular changes are believed to be caused by the redistribution of blood and cerebrospinal fluid toward the head during weightlessness [51]. Eye defects observed from spaceflight data include changes in visual acuity [52], refractive error [53], varying degrees of disk edema [54], globe flattening [54, 55], choroidal folds [53, 56], and hyperopic visual shifts [57]. An exposure-dependent response is observed in SANS symptoms, with up to twenty-three percent of astronauts experiencing changes in near-distance visual acuity after short-duration shuttle missions and up to forty-eight percent of astronauts experiencing changes after long-duration shuttle missions [57]. As mission duration increases, the impairment associated with visual acuity poses a significant risk to mission tasks where fine motor control and target acquisition are required.

The vestibulo-ocular reflex (VOR), a neurovestibular capability, promotes vision-stabilizing compensatory eye movements to stabilize the gaze on a target during head motion [58]. Semicircular canals and otolith organs sense the motion of the head, and the signals are synthesized to influence motor neurons that control eye muscle activity through the oculomotor nerve. Voluntary head oscillations between 0.25 and 1.0 Hz have been experimentally shown to elicit changes in vestibulo-ocular reflex gain in astronaut populations, particularly with a decrease in the gain of vertical and torsional VOR during and after spaceflight [59, 60, 61]. Disruption of the gaze compensation mechanism can result in blurred vision, oscillopsia (illusory movement of the visual world) [31], and/or reduced dynamic visual acuity [62], thereby impacting hand-eye coordination. Eye-head coordination has been quantified in thirty-one Space Shuttle astronauts during visual target acquisition [63]. The results indicate that there was an increase in gaze latency across all mission durations due to decreased velocity and amplitude in both eye saccades and head movements toward the target [63]. Eye movement control during stationary and dynamic head movements is vital for landing operations, particularly when trying to locate and read instrument displays under time pressure, identifying suitable landing locations free of obstructions, and tracking the motion of targets.

In addition to visual impairments, musculoskeletal decrements associated with spaceflight may impact the motor capabilities required to perform hand-eye coordination tasks, thereby reducing long-term flight capabilities. Upon first exposure to microgravity, astronauts have exhibited altered movement variability, reduced accuracy, and increased reaction time during controlled aimed arm movements [41, 64]; however, this is less attributed to muscle atrophy but rather an adaptation to the altered gravity environment. To control and predict motions, we use a forward dynamics model [65] that incorporates gravity. The forward model can be used to convert errors in outcomes into errors in motor commands, which facilitates distal supervised learning [66]. The forward model captures the relationship between motor commands and outcomes to incorporate adaptation from sensory error signals. Until the internal model is able to adapt to the absence of gravity, errors will occur. In addition, skeletal muscle atrophy occurs rapidly, i.e., diminished by up to 20% after only two weeks in space [67] and may rise to 50% without the use of countermeasures [68]. Muscle mass loss is pronounced in astronauts as a result of microgravity exposure with a decrease in muscle fiber mass, force, and power production [69, 70]. In order to mitigate the effects of spaceflight on the musculoskeletal system, current operations on the International Space Station employ countermeasures [71]. However, the exercise devices and programs that employ high resistance can only partially replicate the 1 g environment on Earth, which is conducive to maintaining muscle fiber mass. The relationship between long-duration upper body muscle atrophy as it relates to fine motor control in operationally relevant spaceflight settings has not been explored in depth.

Although eye-hand coordination for fine motor tasks has not been studied as extensively as oculomotor performance, some studies have been conducted during spaceflight missions and postflight. Flight studies for joystick or arm-reaching tasks have revealed increased task response times and decreased task accuracy [72, 73, 74]. postflight results have shown small but significant decreases in fine motor control as quantified through the Grooved Pegboard Test [75]. Within twenty-four hours of landing, astronauts' demonstrated decreased ability to operate simulated vehicles as compared to preflight performance [76]. A recent study evaluated astronaut performance preflight, in-flight, and postflight for a test battery of touch-screen-based fine motor tasks [16]. The tasks include pointing, dragging, tracing, and pinch-rotate tasks. Fine motor control decrements were found during gravitational transitions during the first week in orbit and the first month after landing, but none were found in-flight after adaptation following the first week [16]. Results indicate that although performance stabilizes in flight, the gravitational transitions from microgravity to a gravity-rich environment and vice versa disrupt fine motor performance, which may impact early operations post-landing.

Physiological changes that affect sensorimotor and vestibular functions pose important concerns to mission success and crew safety for future long-duration astronaut crews operating autonomously from Earth. It has been demonstrated that both vestibular [36] and sensorimotor impairments [16] are most severe during gravity transitions; therefore, assessment tools are needed to evaluate impairments prior to mission-critical tasks during and immediately following gravity transitions. The following Sections 1.2.2 and 1.2.3 describe current earth-based assessments used for astronauts and clinical care to inform future assessment tools for human exploration space missions. Section 1.2.4 provides background into current augmented reality applications used for physiological evaluations.

1.2.2 Current Earth-Based Neurovestibular Assessments

NASA implements several post-landing examinations for astronaut crews to monitor readaptation to Earth's gravity. The postflight assessments range from vestibular and motor function tests administered promptly after landing and subsequent assessments with additional equipment throughout adaptation timelines of 1-90 days postflight [9]. Assessments performed immediately after spaceflight missions near the landing sites include oculomotor tests, directed pointing, chair-to-stand test, postural stability, and tandem walking [9]. Although minimal equipment is required for these post-landing assessments, flight surgeons and medical personnel are required to perform the assessments, evaluate the performance of the astronaut crews, and provide post-landing support. Subsequent assessments conducted at NASA's Johnson Space Center include the Sensory

Organization Test battery with computerized dynamic posturography [77] and functional mobility tests [4] to observe balance performance and the time course of recovery to baseline measures. These functional tasks are administered twice (generally 180 and 90 days) preflight to allow for comparisons postflight.

While flight surgeons subjectively assess crew members' sensorimotor function immediately postflight using standard neurological assessments, the resources at Johnson Space Center provide direct measures and standardized methods to assess sensorimotor function. The computerized dynamic posturography system (Figure 1.3) features dual force plates supported by four force transducers mounted symmetrically to measure the distribution of vertical forces [19]. Modifications of the sensory conditions or unexpected perturbations are imposed using computer-controlled movements for the purpose of evaluating the relative importance of visual, vestibular, and somatosensory feedback in stabilizing posture. Sensory Organization Tests performed on the computerized dynamic posturography system consist of a set of increasingly challenging conditions to assess the subject's ability to maintain an upright posture through the effective use of visual, vestibular, and somatosensory information. When the eyes are closed and the support surface rotates in direct proportion to anterior-posterior (AP) body sway, the greatest decrements occur [19]. In clinical practice, computerized dynamic posturography can be used to assess, rehabilitate, and manage balance disorders [78]. Using computerized dynamic posturography, clinicians can identify patients at risk for recurrent falls and develop a safe rehabilitation program.

The Functional Mobility Test (FMT) (Figure 1.3), an obstacle avoidance task, has been used to test astronaut locomotion pre and postflight. Astronauts will need functional obstacle avoidance skills during emergency egress from a vehicle and while exploring planetary surfaces. The FMT task features a series of obstacles placed on 10 cm thick medium-density foam for safety and to increase the difficulty of the task. The obstacles include 31 cm styrofoam blocks, and foam pylons hung from the ceiling, to require stepping, ducking, and weaving movements [4]. It has been documented in the astronaut population that postflight performance is poorer than their pre-adaption scores on obstacle avoidance [20, 4]. Similarly, in clinical care, older adults perform worse than younger adults on obstacle avoidance tasks [79, 80]. Earlier studies have shown that obstacle avoidance tasks and computerized dynamic posturography are sensitive to vestibular dysfunction and that their combined results are highly sensitive [81]. Sensitivity was defined as the true positive diagnostic results divided by the sum of the true positive and false negative diagnostic results. Post-landing assessments such as dynamic posturography, Functional Mobility Task, and those immediately after landing are valid measures for quantifying vestibular dysfunction in astronaut populations, but they require personnel and extensive equipment to be performed in many cases. As such, current earth-based standard assessments implemented by NASA are not suitable for assessments beyond Earth. To support the long-term monitoring of astronaut crews throughout missions beyond low earth orbit, new space-conscious and efficient systems that can be conducted without external support will be required.

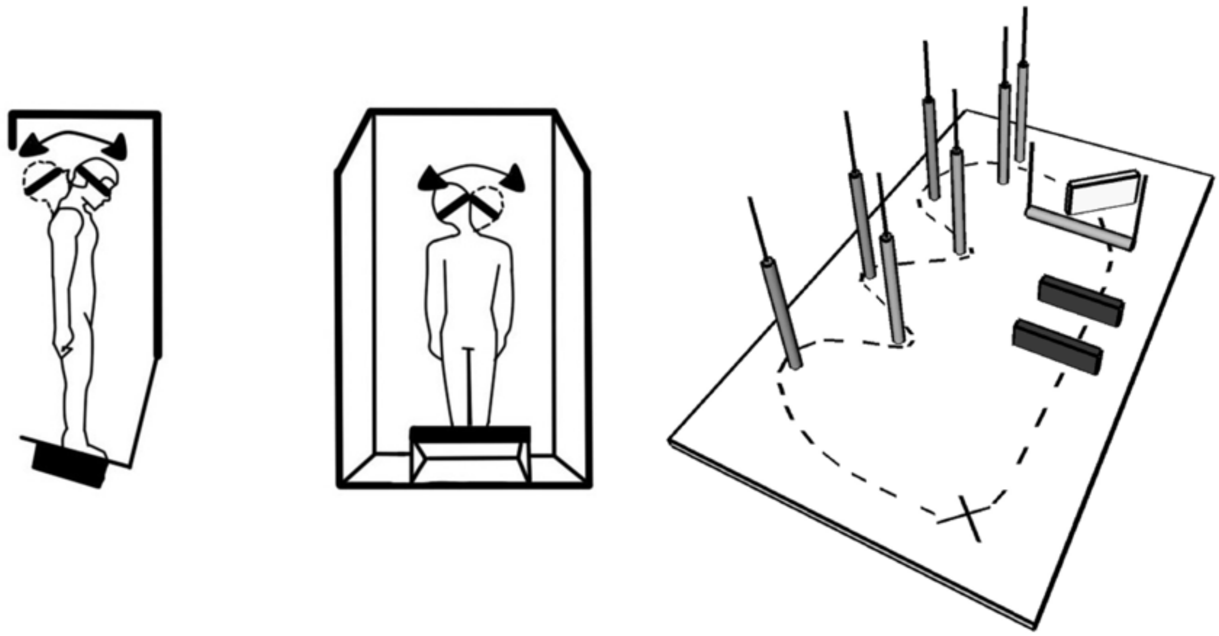

Figure 1.3: Left: Computerized Dynamic Posturography (CDP) [3]. Right: Functional Mobility Test (FMT) [4].

Assessments aimed at neurovestibular system performance are relevant both to astronauts and to those with vestibular impairments, such as the elderly or younger people with vestibular injury. In clinical practice, various additional screening tests are implemented for vestibular disorders and elderly populations, including tandem walking balance [21], static balance with force platforms [22], the Star Excursion Balance Test [23], and the Berg Balance Scale [24]. As with the previously mentioned astronaut assessments, however, these assessments may require equipment and certainly require clinician support in order to be conducted and interpreted for patient care. Although assessments like tandem walking or the star excursion can be implemented with minimal equipment (e.g., tape), the ability to capture relevant data metrics can be augmented with wearable sensors to support clinicians. Sensor-integrated systems such as augmented reality, allow for a standalone device with embedded sensors that are suitable for clinical applications and spaceflight constraints to support decision-making by flight surgeons and astronauts. Chapter 4 of this thesis aims to validate the augmented reality administered balance assessments by comparing them to these Earth-based standards.

1.2.3 Current Earth-Based Sensorimotor Assessments

As a standard measure collected for astronaut crews to monitor their adaptation postflight compared to preflight baselines, eye-hand coordination for fine motor tasks has not been extensively researched compared to vestibular and locomotor evaluations. Hand-eye coordination assessments have been previously implemented in spaceflight research as relatively independent examinations in-flight and pre- or postflight [75, 16]. In postflight operations and clinical examinations, finger-to-nose pointing has been used as an easy, equipment-free method to assess dysmetria [9]. Other sensorimotor tasks are either exclusively administered with a touchscreen or tablet-based equipment [16] or using devices such as the Grooved Pegboard Test [75, 82]. While touchscreen devices are expected to be included in future long-duration exploration missions, an augmented reality-based system can provide additional information beyond touchscreen capabilities.

Augmented reality headsets, such as the Microsoft HoloLens 2, feature four visible light and two infrared cameras to support hand and eye tracking capabilities, respectively. While touchscreens can capture reaction time, movement time, and accuracy data, the manner by which the user acquires targets is largely unmeasurable. For AR, hand tracking provides the ability to evaluate the entire task, including performance measures but also user strategy and movement patterns that may change with varying levels of sensorimotor disruption. In addition to the benefits of embedded sensors, augmented reality systems offer a hands-free replacement of tablet-based procedural guidance, which is the predominant use of touchscreens in flight today [25]. Ultimately, AR systems are multifunctional for spaceflight tasks and provide flexibility for new software integration. Chapter 3 aims to validate the hand-eye coordination measures acquired in AR as compared to a traditional gold-standard touchscreen device to verify the benefit of AR over a touchscreen.

1.2.4 Current Augmented Reality Applications

To date, a handful of Earth-based augmented reality and virtual reality systems have been developed for balance rehabilitation and exercise that focus on aging populations, including tasks such as standing balance [83], [84] and gait during walking [85, 86]. However, these tools feature a limited number of assessments and are constrained by the extent of software and equipment required (i.e., motion capture, force plates, wearable sensors, cloud storage, and processing infrastructures), thereby diminishing their usability for other applications and populations, such as astronaut crews in-flight.

In addition to balance, sensorimotor performance is important for astronaut assessment. Sensorimotor performance, which depends on the visual system and the brain's visuomotor networks to initiate movement towards a target, is nicely encapsulated in Fitts' Law for reaching and pointing tasks that reveal a tradeoff of speed and accuracy [87]. Fitts' Law relates movement time, distance, and accuracy for rapid aimed movements, where the time required to reach a target increases with distance and decreases with size. This law applies to pointing and dragging tasks using a mouse, trackball, stylus, joystick, and touchscreen [6]. A key component of Fitts' pioneering work consists of a conceptual model of human performance which utilizes units that are derived from information notation and are analogous to the processing of information. Information theory is the study of how information can be quantified, stored, and communicated. The Index of difficulty (ID), which represents the difficulty of a movement task as a function of the movement distance and target width, is presented in bits, i.e., the basic unit of information for computing and digital communication [87]. Throughput, a combined measure of movement time and accuracy (Equation 4.3), is presented in bits per second to represent the rate at which information is transmitted by a series of movements [87]. Different input devices are characterized by their throughput, which is the ubiquitous measure used in the literature to compare human performance, i.e. information processing through movement tasks.

Recent efforts have been made to extend Fitt's law into newer interface technologies, such as augmented reality. Unlike touchscreen devices, augmented reality devices allow for different interaction methods beyond pointing selection with the fingertip. Raycasting is an AR method to enable users to interact with holograms at a distance. A ray is cast from the user's hand and acts as an extension of the fingertip by permitting selection at the end of the ray. Some commercial-off-the-shelf AR headsets also utilize handheld controllers to enable human-system interactions. A recent study demonstrated that speed-accuracy tradeoffs exist with different interaction methods (touchpad, pointing gesture, and raycast), but no performance differences resulted from the transparency of holographic content as measured by throughput, movement time, error rate, and incorrect click count [88]. The study identified that the pointing gesture, in which a cursor was placed at the end of the user's dominant pointer finger, yielded throughput values of 1.52 bits/s and an error rate of 3.69 percent. As defined in a recent ISO standard [89] characterizing the evaluation of pointing devices, throughput is the predominant performance indicator that combines accuracy and speed. A review of nine studies that used the 2D target acquisition task procedure described in the ISO 9241-9 standard reported throughput values for isometric joystick (1.6-2.55 bits/s), touchpad (0.99-2.9 bits/s), and mouse input (3.7-4.9 bits/s) [6]. Investigations on newer touch-based devices, such as mobile touchscreens, have demonstrated throughput of 6.95 bits/s, which is 42-88% greater than mouse input performance [90]. The limited literature suggests that the augmented reality pointing gesture yielded throughput values similar to that of a touchpad device and, in some cases, outperforms the touchpad, however, under-performs compared to touchscreen devices. The comparison of AR to 2D planar interaction devices is valid with the assumption that the targets are presented in 2D; however, AR may introduce more complex 3D movements that should be taken into account. Extension of the literature needs to address the three-dimensional interaction space afforded by AR environments to inform the comparison to 2D interface screens.

The findings from the previous literature motivate the need for future examinations of accuracy and precision for the pointing gestures in AR interactions, including investigations into error rates. Understanding how holographic targets affect user accuracy, speed, and precision can support future augmented reality designers for improved usability across all users and applications as well as contextualize the interpretation of performance in AR for assessments. In order to interpret AR-based hand-eye coordination performance and inform crew readiness decisions, an understanding of nominal user performance during AR-based tasks is important. Astronauts' precision and accuracy are essential for tasks such as manual landings, where selecting interface buttons is crucial to mission success. Chapter 3 of this thesis examines the performance of healthy adults on a multidirectional tapping task within an augmented reality (AR) environment and a traditional touchscreen device to address the gaps identified.

1.2.5 Usability Design and Testing of Augmented Reality Applications

The field of human-computer interaction has produced many important frameworks of human factors principles and practices that are integral to the design, development, and testing of systems. Some of the most prominent and widely used frameworks included Shneiderman's 8 golden rules of interface design [91], Nielsen's usability heuristics [92], and Wicken's principles of display design [93].

Sheiderman's 8 golden rules of interface design include consistency, universal usability, information feedback, dialogues to yield closure, error prevention, easy reversal of actions, keeping the users in control, and reducing short-term memory load [91]. Nielsen's usability heuristics promote the visibility of system status, a match between the system and the real world, aesthetic and minimalist design, and accessibility to additional documentation for help [92]. Furthermore, Wickens emphasizes the legibility and audibility of displays, the avoidance of absolute judgment limits, redundancy gain, pictorial realism, minimizing information access costs, proximity compatibility, and replacing memory with visual information [93].

These frameworks share common themes and display standards to promote ease of use by a wide range of users. As well as being integrated into the design of interfaces, these standards can also be used to categorize the usability of an interface through human-in-the-loop testing with first-time users. All new systems that require human-computer interaction should be assessed for usability to ensure the features accommodate users' needs and limitations to perform tasks safely, effectively, and efficiently. In the case of augmented reality technology, standard usability protocols can be used; however, additional AR-specific usability heuristics need to be considered to address new interaction modalities available such as hand gesture-based interactions.

Iterative usability testing under goal-oriented conditions is the basic principle of user-centered design, and it is an essential component of ensuring that new systems are used efficiently and effectively. The primary objective of usability testing is to ensure a system's functions, features, and overall purpose are in line with what users require by observing how real people, preferably the target audience, use the system. The usability of a system is determined by a combination of features such as functionality, visual appeal, and usefulness. The system must be tailored to the context in which it will be used, as well as the characteristics of the users. To date, there have been many attempts to standardize usability procedures through questionnaires such as the System Usability Scale [94] and the Questionnaire for User Interaction Satisfaction [95]. However, these surveys are administered retrospectively and focus on attitudinal measures where an individual's design preferences may actually hinder the user from achieving their goals [96]. By requiring users to evaluate an experience retrospectively, they are more likely to remember and make judgments on the most recent or intense aspects of the experience [97]. As a result, questions administered immediately after a usability task or test scenario, along with research staff actively watching participants interact with the system, can provide a better measure of overall usability and can also provide a diagnosis of specific usability issues [98].

Usability testing helps to verify newly developed systems for efficient and effective use by a wide range of users with various proficiencies. It is also important to consider the physiological capabilities of the target audience, as well as the environment in which the system will be used when identifying usability concerns. For astronaut populations, the ability to self-administer assessments improves the crew's autonomy to evaluate performance throughout long-duration mission timelines. The ability to self-administer assessments supports future long-duration missions where communication with ground support may not be feasible.

Understanding the human challenges present in space exploration, such as physiological and perceptual needs, will support the design and performance of integrated human-machine systems. The long-term effects of weightless, prolonged exposure to the space environment (radiation, microgravity, gravity fields), as well as the mental toll of protracted isolation, may also impact system usability. The astronaut's ability to properly operate a device is limited by constraints of their physiology, capacity to process input and stimuli, physical limitations as to viewpoints and observability, as well physical preparedness. Therefore, the development of AR systems for crew members in altered physiological states needs to be taken into consideration during the design of the interface and interaction elements. An understanding of baseline usage with healthy, non-impaired users is necessary before we can examine altered state usage.

In an effort to generalize AR assessment tools, a wide range of ages should be incorporated into the usability testing to support all users. A wide range of ages are represented in the astronaut population, therefore usability testing with astronaut-like demographics is desired. In addition, as mentioned previously, an AR-based assessment tool may be beneficial for clinical evaluations of aging populations and those with sensorimotor impairment. Although digital literacy among the aging population is improving, future designs and the development of user-centered technologies need to accommodate aging users. In order to identify usability concerns from meaningful comparisons between distinct populations, human-in-the-loop usability testing should be conducted with younger and older adults for future iterations of the software to ensure the tool is universally operable and effective for physiological evaluations.

1.3 Research Objectives

The aim of this work is to develop and evaluate an augmented reality application with embedded wearable sensors as a potential tool to support neurovestibular and sensorimotor assessments in astronauts for future long-duration space missions. At present, there are no fitness-for-duty assessments with respect to neurovestibular and sensorimotor performance in mission operations on the International Space Station. Robust evaluation tools and devices are needed to perform EVA readiness assessments and to lay the foundation for countermeasure development and evaluation. The solution requires novel technologies and hardware systems that are validated, observe mission resource constraints, are flexible for diagnostic capabilities and risk mitigation, and ultimately systems that are informed by human factors principles for effective utilization.

1.3.1 Research Questions

The work evaluates the following primary and secondary research questions:

1. Can augmented reality be used as a standalone assessment tool for neurovestibular and sensorimotor evaluations in astronauts?
    - 1.1 What dependent measures can be derived from the sensor capabilities of AR systems in evaluating balance and postural control?
    - 1.2 What dependent measures can be derived from the sensor capabilities of AR systems in evaluating hand-eye coordination?

During the development of AURORA (Chapter 2), a strong emphasis was placed on addressing questions 1.1 and 1.2 in order to aid in the evaluation of question 1. In addition to traditional physical assessments being integrated into a holographic configuration, the available sensor data streams were examined for metrics that could be extracted to support the assessments of task performance. Relevant metrics were defined by reviewing the literature and creating additional surrogate measures that align with the assessment goals. Chapters 4, 5,and 6 implemented the derived metrics to compare the holographic assessments with the physical assessments.

2. For self-assessment of physical abilities, what usability features for design and user interactions need to be considered?

   2.1 What usability concerns are discovered with younger healthy adults?

   2.2 What usability concerns are discovered with older healthy adults?

By identifying usability concerns from both younger and older users, human-in-the-loop usability testing provided an appropriate testbed to evaluate questions 2.1, and 2.2. Based on the results of the usability tests, updates were made to the AR application to improve user self-assessment capabilities.

3. What impact does the use of holographic objects in assessments traditionally performed with physical objects have on overall performance?

   3.1 When compared to traditional balance assessments, how does AR impact postural control and task completion measurements?

   3.2 When compared to traditional hand-eye coordination assessments, how does AR impact hand-eye coordination and task completion measurements?

By comparing performance within the AR modality to physical implementations of the assessment with real objects, a repeated measures design was used to evaluate questions 3.1, and 3.2.

1.3.2 Alignment of Research with NASA Human Research Program Gaps

This dissertation research focuses on addressing the gaps in current human spaceflight knowledge and capabilities regarding sensorimotor performance, as defined by the National Aeronautics and Space Administration (NASA) Human Research Program. The research also addresses the limitations of current earth-based measures by performing preliminary evaluative research on the use of augmented reality as a viable alternative. The following outlines these aforementioned gaps in more detail.

The NASA Human Research Program is comprised of five elements, namely the Research Operations and Integration, Space Radiation, Human Health Countermeasures, Exploration Medical Capability, and Human Factors and Behavioral Performance. A key goal of the Human Research Program is to discover methods and technologies to support safe and productive human spaceflight through risk identification and organized research plans. To this end, the Human Research

Roadmap [7] was developed to describe risk evidence reports that support the existence of human system risk, gaps in knowledge about characterizing or mitigating the risk, and the research tasks to be carried out in order to close the gaps and reduce the risk.

Evidence reports are compiled by the Human Research Program Elements and subject matter expert scientists. Human system risks encompass the physiological and performance impacts of space travel hazards. The unique environment features altered gravity, radiation, isolation and containment, distance from Earth, and unique challenges related to medical support, human factors, and behavioral health. Gaps represent the critical unknowns that need to be addressed to mitigate risk and therefore serve to focus the areas of research work to address risk reduction milestones. By better characterizing a risk or developing mitigation capabilities for reducing a distinct risk to an acceptable level, a research task may partially or completely close a gap. Ultimately, each outlined research task or series of tasks will lead to a product or deliverable to expand NASA's human spaceflight knowledge and capabilities.

The subject matter experts within the Human Health Countermeasures (HHC) Element of the NASA Human Research Program have identified the risk of altered sensorimotor and vestibular function, thereby impacting critical mission tasks. It has been widely documented with returning astronaut crews that microgravity exposure causes changes in sensorimotor and vestibular function, which leads to motion sickness, spatial disorientation, impairment in postural control and locomotion, and deficits in manual and fine motor control [20]. In the wake of these gravity transitions, crew members may experience performance decrements in manual vehicular control, extravehicular activities, and capsule egress [48]. As a result of prolonged microgravity exposure and the distance from Earth, the altered sensorimotor and vestibular function may be of particular concern for future long-duration missions to the Lunar and Martian surfaces for risk of injury and compromised performance.

NASA's Human Research Roadmap identifies the following knowledge gaps regarding extended exposure to microgravity, including the ability to control a vehicle during G-transitions (SM-102) [99], motion sickness incidence and severity (SM-103) [100], changes in brain functions (sleep, cognition, and attention) (SM-104) [101], and the ability to perform EVAs right after a G-transition (SM-101) [102]. Secondary gaps identify the need for the development of countermeasures to aid in manual control tasks (SM-202) [103], to mitigate space motion sickness (SM-203) [104], and provide the ability to self-administer vestibular and sensorimotor assessments and rehabilitation tools without assistance from ground support (SM-204) [105]. Sensorimotor training, self-administered rehabilitation, sensory augmentation, and in-flight balance training are among the countermeasures solicited and those currently under study. In order to characterize the long-term health effects of this sensorimotor mission risk, long-term monitoring will be necessary.

This thesis directly aims to address the need for a countermeasure that provides astronaut crews the ability to self-administer vestibular and sensorimotor assessments throughout mission timelines to monitor physiological capabilities and inform extravehicular activity readiness (SM-204) [105]. The augmented reality countermeasure proposed may also serve to characterize the effects of long-duration weightlessness identified in the primary sensorimotor gaps by adding to the depth of knowledge obtained from long-term health monitoring. In addition, the research seeks to address specific technological issues outlined in NASA's 2020 Technology Roadmaps [106], TA 6.3.1 Medical Diagnosis/Prognosis (6.3.1.1) and TA 6.3.2 Long-Duration Health (6.3.2.4), regarding novel diagnostic and countermeasures hardware that are small, compact, rapidly deployable, and stowable for crew health monitoring and long-term health. The research is informed by TA 6.3.4 Human Factors (6.3.4.1), and aims to provide design guidance for AR mission display devices in NASA's Human Integration Design Handbook (HIDH) [107], which currently lacks guidance for Extended Reality (XR) technologies.

1.3.3 Research Outline

Chapter 2 of the thesis outlines the development of the AR application, the Augmented Reality Operations Readiness Assessment (AURORA). In this section, we examine the interface affordances, neurovestibular and sensorimotor tasks, physiological performance measures, and user interaction methods. The work addresses aspects of the second research question.

Chapter 3 reports the methods and results of two distinct human-in-the-loop usability tests with younger and older healthy adult populations. The Chapter includes the methods by which usability was evaluated as well as the usability concerns identified that were addressed to improve the application before implementation into evaluative testing. Comparisons between the two distinct cohorts are evaluated. This section addresses questions 2, 2.1, and 2.2 with respect to usability and efficient autonomous use.

Chapter 4 features the evaluation of AURORA's two-dimensional hand-eye coordination task with healthy adults. The results are interpreted with respect to user performance on a touchscreen device. The work addresses questions 1, 1.2, 3, and 3.2 with respect to the use of AR as an effective assessment tool for hand-eye coordination.

Chapter 5 examines the evaluation of AURORA's dynamic balance assessments as compared to the physical implementation of the tasks. The work investigates user performance for the tandem walk and four square step test. This section addresses aspects of questions 1, 1.1, 3, and 3.1 regarding the use of AR as an effective assessment tool for balance and postural control during dynamic tasks.

Chapter 6 reports findings from the comparison study of AURORA's operational balance assessments within AR as compared to physical objects of the same scale and dimensions. This

Chapter investigates user strategy and task performance for the ingress and egress task and the obstacle weave task. The work addresses questions 1, 1.1, 3, and 3.1.

Chapter 7 discusses the results in the context of the research motivation and research gaps. The section touches on recommendations and future applications of the augmented reality application developed through this work as well as future research opportunities to explore.

CHAPTER 2

Augmented Reality Application Development

The following chapter discusses the development of the AR application while informing the second research question regarding what usability features for design and user interaction need to be considered in AR applications. The system was developed with consideration for prominent human factors design principles to promote usability and self-administration of assessments. The embedded assessments were informed by Earth-based standards currently implemented in post-flight assessments and clinical care for aging populations and vestibular patients. The following sections describe the application development, including design aims and intended use cases (Section 2.1.1), system capabilities (Section 2.1.2), user interactions (Section 2.1.3), neurovestibular tasks (Section 2.2.1), and sensorimotor tasks (Section 2.2.2).

2.1 Augmented Reality Operations Readiness Assessment

The Augmented Reality Operations Readiness Assessment (AURORA) is a standalone application that assesses neurovestibular and sensorimotor performance with holographic visual cueing and integrated inertial measurement units (IMU). AURORA is currently developed for the Microsoft HoloLens 2 using Unity Technologies real-time 3D development platform in conjunction with Microsoft's Mixed Reality Toolkit 2. The application consists of two main modules, the neurovestibular balance assessments with three sub-modules (static, dynamic, and operational balance) and the sensorimotor hand-eye assessments with three sub-modules (2D, 3D, and random target acquisition) (Figure 2.1). The assessments featured in the application were chosen from clinical gold standard evaluations [108] and tasks currently implemented by NASA for observing crew adaptation timelines pre/post flight on Earth [9, 1].

2.1.1 Design Aims and Intended Use

AURORA was designed for implementation in ground-based applications, cis-lunar microgravity, and altered gravity conditions on a planetary surface. The system is intended for crew readiness

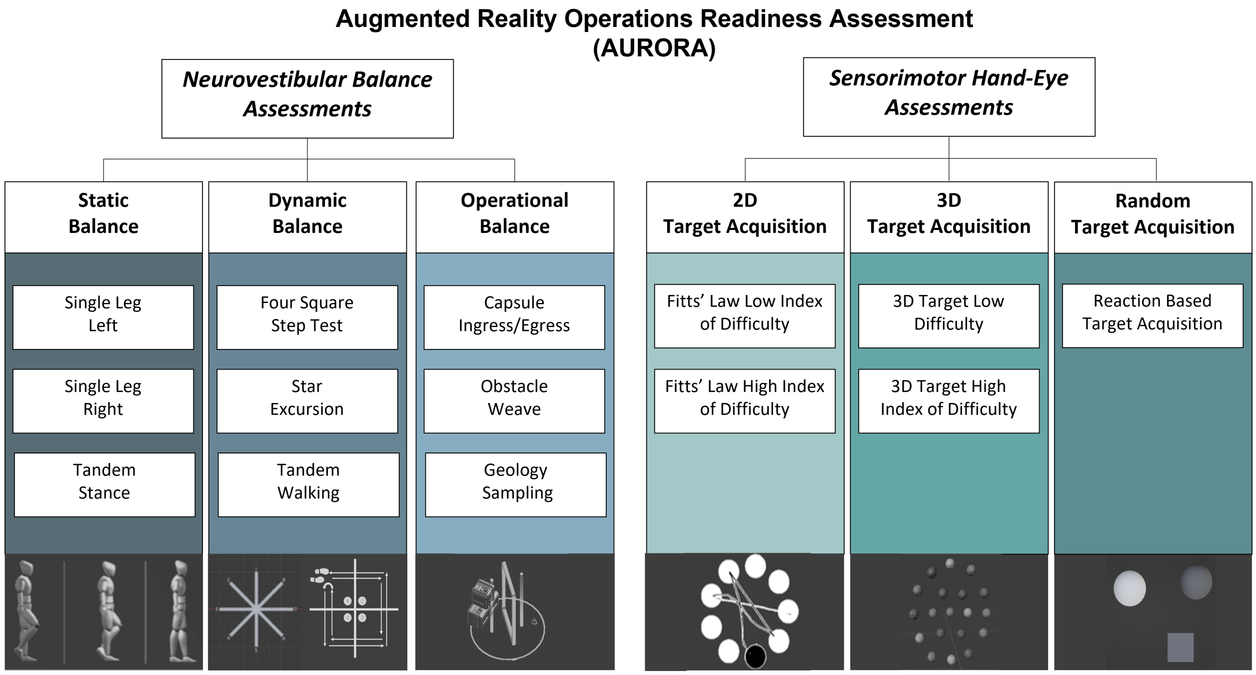

Figure 2.1: Augmented Reality Operations Readiness Assessment (AURORA) Application Modules

assessments and does not directly evaluate motion sickness, although the metrics obtained may characterize disorientation during task performance. The neurovestibular assessments are gravity-dependent by requiring ambulation and ground surfaces to perform the evaluations; however, the sensorimotor evaluations may be conducted in both planetary surface settings and microgravity (where hand-eye coordination is the primary measure of interest). AURORA serves as a tool for acquiring baseline performance metrics preflight on Earth, monitoring performance adaptation in-flight and during gravity transitions, and evaluating postflight astronaut performance.

The application features a modular design for flexibility in operations. Volume 1 of the NASA Space Flight Human-System Standard for Crew Health (NASA-STD-3001) [109] declares that the nature of mission-associated critical operations shall guide in-mission Fitness-for-Duty requirements (including but not limited to vehicle control, robotic operations, and EVAs). To align with the flexibility specified, an example use case would be to utilize the tool before initiating an operational task in-flight, where crew members may choose the appropriate performance assessment to evaluate aspects of the upcoming task. For example, if the operational task focuses mainly on ambulating rough surfaces, the user may select the dynamic assessments and forego the static balance assessments. Moreover, if the astronaut's operational task requires handling interface controls accurately, such as manually landing a spacecraft, the hand-eye coordination assessment may be selected. All the tasks within AURORA are designed to be executed and completed within a minute or less to support operational time constraints. Most assessments require as little as twenty seconds to complete the required task, with additional minimal setup time prior to execution. The development focused primarily on the use case scenarios for spaceflight habitats and modules, with all assessments conducted within the confines of a 2x2 meter work area, consistent with representative hardware volumes taken from ISS exercise equipment [110]. Flexibility in performance assessments and easily deployable systems are critical for efficient operations to inform EVA readiness.

AURORA necessitates minimal resource requirements for implementation and storage. The application is completely housed within the Microsoft HoloLens 2 and therefore requires only the headset and charger to operate. The charger included with HoloLens 2 can fully charge the battery in less than 65 minutes when the device is in standby mode. The 5660 mAh battery gives the device a battery life of up to three hours of active use [111]. In standby mode, the battery of the device lasts up to two weeks. A single session can support EVA readiness evaluations for multiple astronauts with the current device without requiring a charge.

The Microsoft HoloLens 2 is an untethered holographic headset with embedded computing, cameras, and sensors [5]. An inertial measurement unit is positioned within the headset's visor near the middle of the wearer's forehead. The placement of the IMU is consistent with the literature for wearable sensors to assess standing balance [112]. The device is the current state-of-the-art technology that supports hand and eye tracking using four visible light and two infrared cameras (Figure 2.2). The embedded visible light cameras have a diagonal field of view of 96.1 degrees, with two cameras facing forward and two at an oblique angle to support hand tracking and scene understanding. At present, the HoloLens 2 device provides the largest field of view capabilities for the user with 52 degrees diagonal, 43 degrees horizontal, and 29 degrees vertical fields of view [113]. While the headset's field of view covers a large part of human vision, the visibility is still far from encompassing the entirety of human vision, i.e., approximately 180 degrees forward-facing field of vision [114].

The Microsoft HoloLens 2 supports eye calibration to ensure the system accurately tracks the user's eyes and to improve the viewing and interaction experience. The calibration process is accompanied by dynamic eye tracking to follow where the user looks in real time to improve comfort, hologram alignment, and hand tracking. The eye-tracking approach also ensures that the headset's visuals are displayed correctly as the headset shifts slightly on the head throughout use. Up to fifty recently used calibration profiles are stored on the device. When the user dons a device again that has previously been calibrated to their eyes, the display automatically recognizes the user from the stored calibration profiles and adjusts for quality and comfort [115]. The calibration procedures require less than 5 minutes and can be conducted preflight on the device to be flown or in-flight anytime.

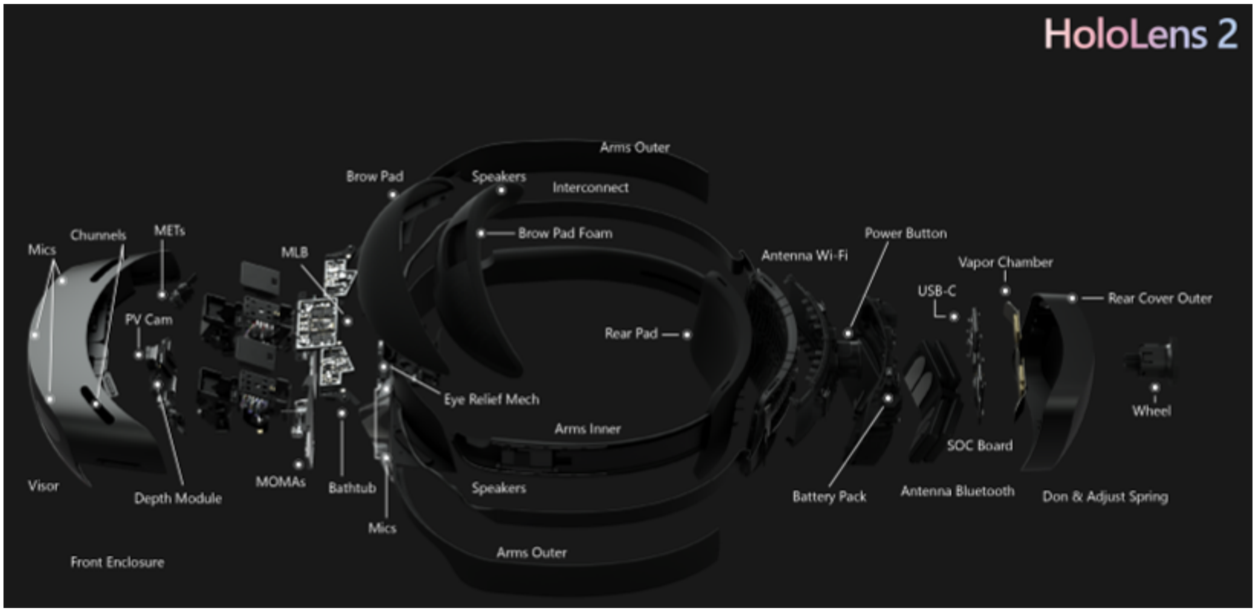

Figure 2.2: Microsoft HoloLens 2 Device Specifications [5]

The HoloLens 2 device provides the capability to define task metrics and save data for further processing. The data obtained from the application, i.e., task metrics and the raw hand and eye tracking data, are stored directly on the device. The data accumulation is only limited by the system's storage capacity, which is currently 64 GB of Universal Flash Storage. Data can be easily offloaded from the device to a computer for further processing using the charging cable, which serves as a multipurpose tool for connecting to external devices. Data can also be offloaded via a Wireless connection using the Microsoft Device Portal. Data offloading procedures can be included in mission operations when more in-depth analyses are needed or for performance documentation.

The headset's storage allows for user data to be saved and accessed again for comparisons. In this case, a crew member can perform an assessment in-flight and immediately compare their performance to preflight measures to inform extravehicular activity readiness or to monitor physiological capabilities. Although the Microsoft HoloLens 2's limited storage (64-GB) [5] can be mitigated by implementing consistent offloading procedures, a device with increased storage capacity would be preferred for future long-duration missions. During dynamic operations, numerous offloading procedures would require crew time, and mission operations may not be flexible.

AURORA is intended to be used by crew members in-flight without assistance from Earth-based ground support. It is expected that crew members will be trained on how to use AURORA prior to flight during their preflight testing operations. For systems to be used in flight during operations, astronauts typically undergo extensive training. However, multiple training sessions with a system like AURORA are not guaranteed. AURORA was designed to minimize the training required for first-time users and to support training retention. As astronauts may rely on the device and application for long-duration missions, it is imperative that skills are retained months after training.

2.1.2 Design Affordances

To support effective user utilization without external assistance, the development process focused on implementing prominent human factors principles, including consistency across displays, intuitive assessment navigation, and animations accompanied by verbal and written task descriptions. Specific frameworks that were applied included Shneiderman's 8 golden rules of interface design [91], Nielsen's usability heuristics [92], and Wicken's principles of display design [93].

Sheiderman's 8 golden rules of interface design include consistency, universal usability, information feedback, dialogues to yield closure, error prevention, easy reversal of actions, keeping the users in control, and reducing short-term memory load [91]. Furthermore, Nielsen's usability heuristics promote the visibility of system status, a match between the system and the real world, aesthetic and minimalist design, and accessibility to additional documentation for help [92]. Furthermore, Wickens emphasizes the legibility and audibility of displays, the avoidance of absolute judgment limits, redundancy gain, pictorial realism, minimizing information access costs, proximity compatibility, and replacing memory with visual information [93].

Consistency within AURORA was maintained not only across sequences of actions for similar tasks but also within prompts, menus, interface colors, fonts, layouts, and symbols. The text was maintained at a constant size relative to the viewer. Each assessment features a hand-attached menu, a task information screen, audio and written instructions, as well as an animation or pictorial representation of the task. These design aspects are described in detail in Section 2.1.3. All routinely used buttons, such as start and stop assessment, information dictation, assessment navigation, hologram placement, and reloading or resting the assessment scene, were maintained consistent throughout the application and between assessment scenes. In accordance with Wicken's proximity compatibility principle, buttons with similar functions, such as start and stop assessment buttons and navigation buttons for assessments, were grouped together.

Usability across all users with visual perception ability was in part supported by comprehensive task explanations (i.e., written, verbal, and visual) for novice users and shortcuts for expert users. More experienced users, i.e., those who have used AURORA on several occasions, can bypass the assessment instructions and proceed with the assessment evaluations. New users of AURORA may read the instructions and also view how the assessment tasks are performed as the instructions are dictated in order to better understand the task goals and how to execute the task. The task animations provided visual information to support task understanding and limit the mental effort required to perform the task.

Information feedback was provided with audio and visual cues. As the index finger approached the user interface buttons for selection, proximity lighting changed the surface hue and lighting of the button to indicate distance. Following the successful acquisition of a button by the index finger, a clicking noise further indicated success. For timed tasks such as the static balance assessments, mentioned further in detail in Section 2.2.1, a sound effect is played during the completion of the task to notify the user that the task is complete. Additionally, sound effects were used to indicate the start and end of each assessment; upon selecting the ”start assessment” or ”stop assessment” buttons, dictation would indicate ”begin” or ”assessment complete,” respectively. The dictation feedback was the chosen dialogue to yield closure as encouraged in Sheiderman’s 8 golden rules [91]. During sensorimotor assessments, where time to completion is critical, a five-second visual countdown is used and projected in front of the user to prepare them for the start of the task. Information feedback displays, such as heads-up display panels, are presented to inform completion and instruct for following tasks, an example described in Section 2.2.1 for the Geology Sampling task.

The prevention of errors and easy reversal of actions were managed with reset buttons and the option to reload the assessment scene entirely. Dialogues were presented to coach users through stepwise procedures, such as the placement and fine-tuning of the location of floor-based holograms. Redundancy in the design ensured most user actions, even those not in line with the instructed procedure, produced the intended result. For example, the users are instructed to start data collection by selecting the start assessment button and subsequently end the data collection by selecting the stop assessment button. If the user selected reset assessment instead of stop assessment, the system was designed to stop data collection, save, and store the data before resetting the scene. Human-in-the-loop usability (3) testing provided real user feedback to identify unanticipated actions despite taking into account many user actions and errors during the design process. Although prominent human factors principles for interface design were consulted and implemented in the design process, usability testing is a crucial aspect of the design process to identify any usability concerns by target users before the application is released for use.

2.1.3 User Interactions

Users interact with AURORA using standard augmented reality hand gestures such as touch selection with their index finger, raycasting for distant manipulation of holograms, and hand-attached menus for action selections (Figures 2.4, 2.5). The touch selection approach is the most intuitive way to interact with buttons because it is similar to touchscreens; however, the gesture requires adequate evaluation of the depth between the fingertip and button location. Touch selection also requires the visibility of the index finger by the headset's cameras to support tracking. Index fingers held forty-five degrees above the horizontal were qualitatively shown to support the headset's tracking abilities as well as the isolation of the index finger by curling the remaining fingers in a fist (Figures 2.3, 2.4, 2.5). The HoloLens supports touch selection by providing a white floating circle pointer (similar to a mouse pointer) near the tip of the index finger to help target elements. When the white float circle pointer does not appear on the index finger, this indicates the headset is having difficulty tracking the fingertip, and therefore interactions may be affected. The lack of the white circle pointer should prompt the user to alter their finger or gesture position to support hand tracking. Raycasting implements a laser pointer (hand ray) from the index finger to the target and allows the user to interact with holograms at a distance (Figures 2.4, 2.5). After an item is targeted with the hand ray, users can act on the target in different ways, such as pinch grabbing to translate and rotate or using air tap (quick pinch motion of the index finger and thumb) to select holograms at a distance.

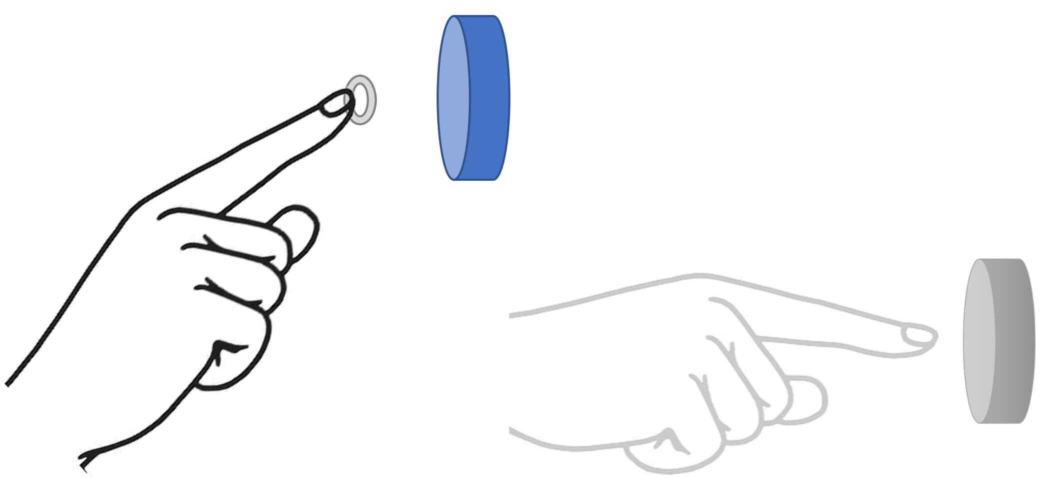

Figure 2.3: The left image depicts the hand posture that best supports the headset's finger tracking. The right figure depicts the hand posture that inhibits finger tracking as the hand occludes the fingertip.

Hand menus provide a robust interaction that allows the user to quickly bring up hand-attached user interfaces for frequently used functions and prevents false activation when the palm is no longer visible by the headset's cameras. The hand menu is positioned on the ulnar side of the palm and allows the user to instantly access hand menu options such as assessment navigation and starting/stopping assessments (Figure 2.5). Hand menus and world-anchored menus incorporate

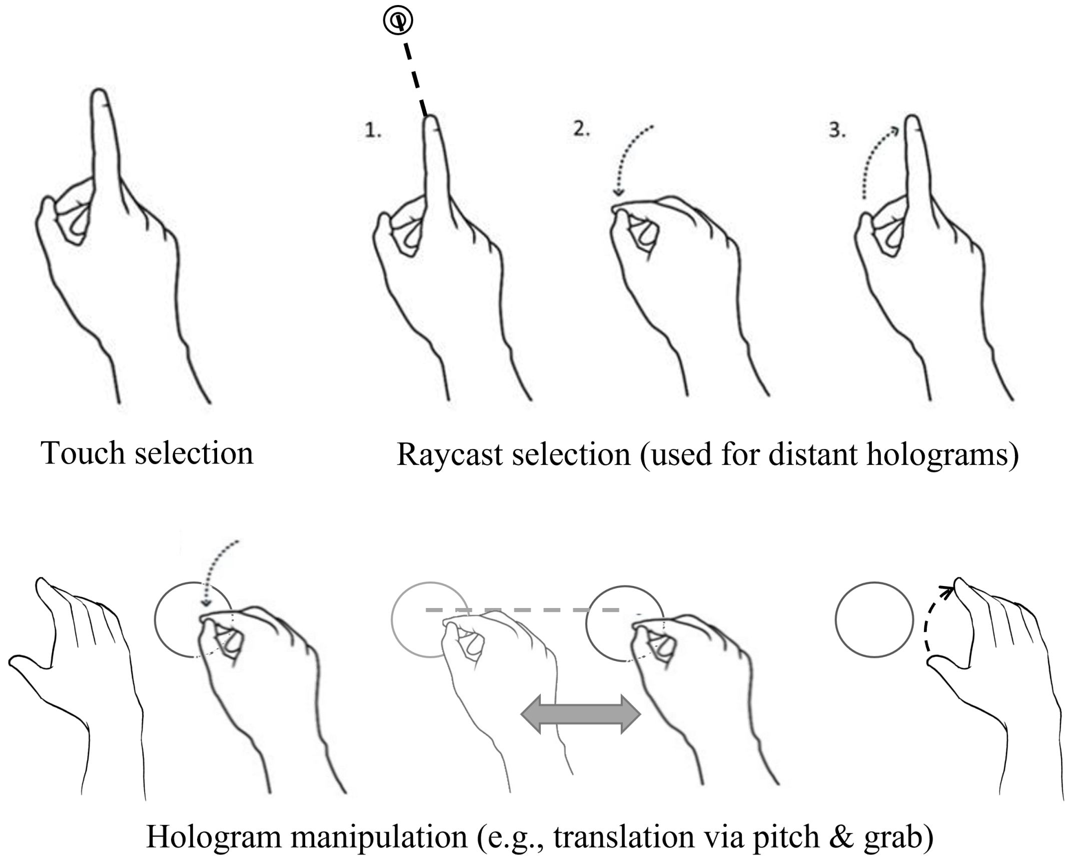

Figure 2.4: Left: Touch Selection. Right: Raycasting for distant manipulation. Bottom: Hologram manipulation to move within 3D space.

proximity lighting for buttons, where the surface shading of the button changes hue as the user's finger approaches to assist with depth perception.

The assessment instruction panels include information about the task and provide buttons to activate and deactivate the audio dictation that is accompanied by animation examples (Figure 2.5). The neurovestibular balance assessments that require floor-based holograms also provide a button to initiate the placement of holographic content on the floor with sliders that correspond to axes in three-dimensional space.

Users are able to fine-tune the placement of holographic content on the floor for all floor-based assessments using the vertical and horizontal sliders that allow for up and down, side to side, and forward and back motions (Figure 2.6). Current AR algorithms for scene understanding proved to be imperfect estimations of floor location, impacting the automatic placement of holograms. Slider capabilities were implemented to address these concerns and align with considerations for volumetric constraints of habitat designs. The holograms are projected automatically below the user's head position during placement; however, fine-tuning may be required to reposition the

Figure 2.5: Left: Example assessment instructions panel. Right: Hand menu with scene navigation and start/stop capability. Demonstration of raycasting interaction with the right hand.

hologram between physical capsule structures that may impede the execution of the assessment.

2.2 Assessment Tasks

The following section describes the assessment tasks provided in the AURORA system. The tasks were chosen from clinical gold standard evaluations [108] and tasks currently implemented by NASA for observing crew adaptation timelines pre/post flight on Earth [9, 1]. The detailed descriptions of the Neurovestibular tasks (Section 2.2.1) and Sensorimotor tasks (Section 2.2.1) serve to outline the task-specific goals and user instructions. To contextualize the selection of these assessments, Section 2.2.3 discusses their relevance and importance to spaceflight operations.

2.2.1 Neurovestibular Tasks

The Neurovestibular module features three sub-modules (static, dynamic, and operational balance) to evaluate astronaut readiness under varying degrees of balance difficulty and task objectives (Figure 2.7). The static balance sub-module assesses three static balance poses with and without neurovestibular disturbances, both visual (eyes open, eyes closed) and motor (anterior/posterior and lateral head tilting). Head movements, especially in the pitch and roll planes, are the predominant provocative stimuli for space motion sickness [43]. The neurovestibular disturbances

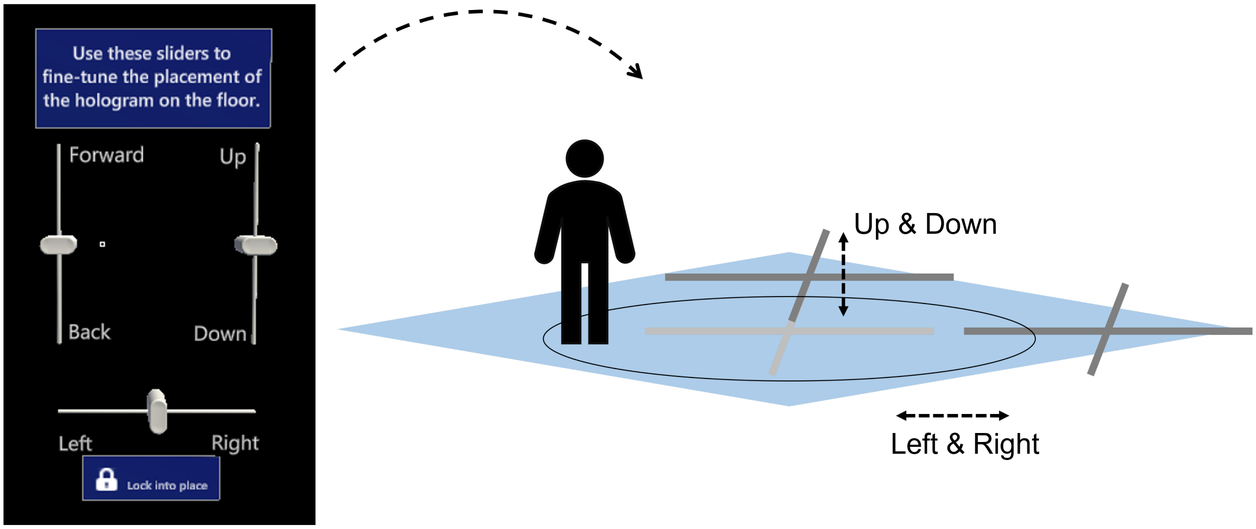

Figure 2.6: Left: Slider interface to fine-tune positioning of floor-based objects. Right: Demonstration of Four Square object placement. Circle depicts 2x2 meter work volume.

present varying difficulty levels, the most challenging being the eyes closed configuration paired with either head tilting or rolling.

The selected static balance postures, i.e., single leg stance [116] and tandem stance [117], challenge the participants' unilateral stability and stability with a reduced medial/lateral base of support. The balance postures are evaluated for a maximum of 20 seconds within each posture and visual or motor condition as is traditional in the Balance Error Scoring System [118]. The AR system provides holographic visual cueing (i.e., static and dynamic eye gaze focus objects) to support the execution of assessment tasks (Figure 2.7). For example, a dynamic eye gaze object is projected in front of the user to model the desired head-pitching movement patterns. The eye gaze object oscillates (+/- 20 degrees) within the user's field of view and is accompanied by audio tones at 0.33 Hz. Voluntary head oscillations between 0.25 Hz and 1.0 Hz have been experimentally shown to elicit changes in vestibulo-ocular reflex gain in astronaut populations [119]. A circular eye gaze object projected 0.5 meters from the user travels in an arc to prompt head rolling (Figure 2.7).

The dynamic balance sub-module comprises the Four Square Step Test [120], the Star Excursion/Y Balance test [23], and tandem walking [21] (Figure 2.8). As a clinical test, the Four Square Step Test is reliable, valid, easy to score, quick to administer, requires little space, and needs no special equipment [121]. This task is a clinical standard test of dynamic balance and stability in which participants step forward, backward, left, and right over a low object, always facing the same direction following a square foot pattern. The participant performs the pattern clockwise followed by counterclockwise. The cross object on the floor is traditionally made with tape or small rods.

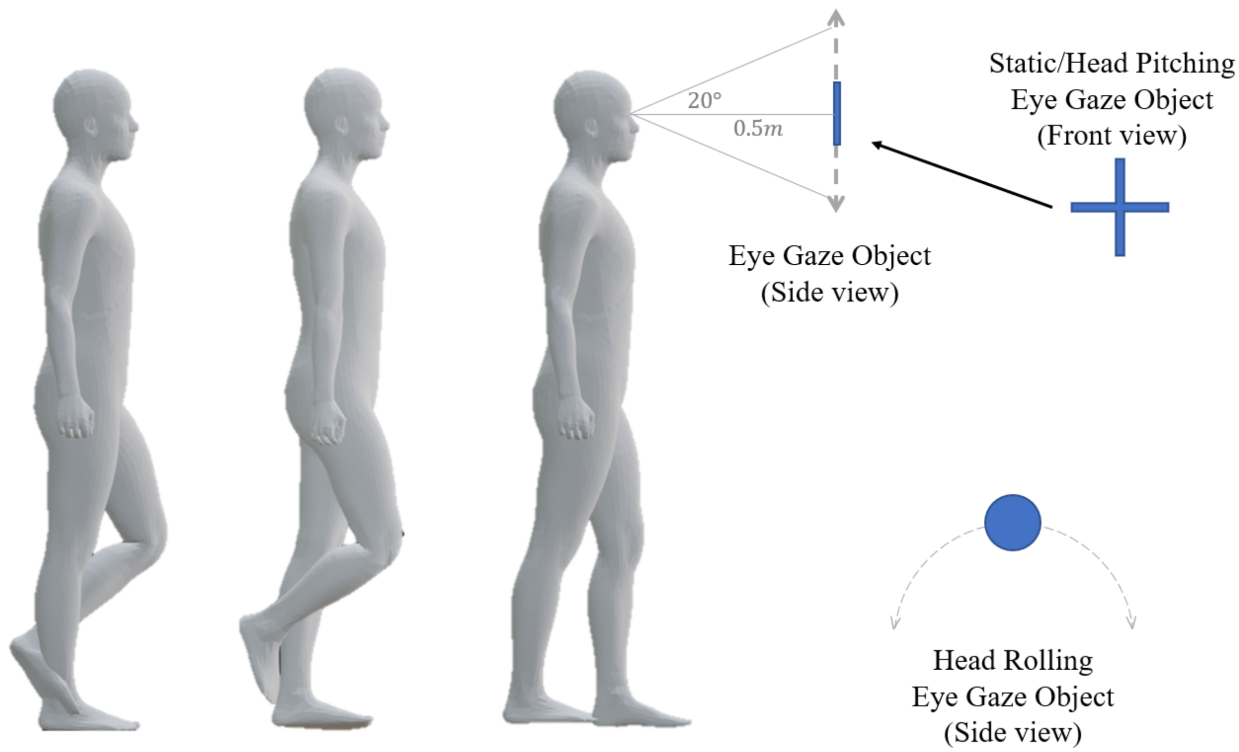

Figure 2.7: Static balance assessments with holographic visual cueing objects.

In AURORA, the cross object is placed on the floor as a hologram object.

The Star Excursion traditionally features the placement of 6 strips of tape on the ground at an angle of 45 degrees [122]; in AURORA, the star object is holographically projected. The star excursion is a dynamic test that requires strength, flexibility, and proprioception. The Y balance test is an abbreviated assessment with only three axes (i.e., axes 1, 4, and 6) in Figure 2.8. To perform these assessments, the participant stands at the center of the object on one foot. With their hands on their hips, the participant reaches out with their offloaded foot as far as possible along the axes of the object and lightly touches the floor before returning to the origin. The task is to perform in a clockwise fashion while balancing on the left foot and anti-clockwise while balancing on the right foot. For both assessments, the participant begins with the anterior or forward axis of the object. During the star excursion, once the participant passes the midline, they will continue to move the offloaded leg posterior to the grounded leg until they reach the most lateral axes. To tap the final axis located at a 45-degree angle to the front left or right, the participant will bring their offloaded leg anterior to their grounded leg.

The star excursion assessment requires the user to stand at the center of the star structure and reach their offloaded foot as far as possible along the axes of the star. In a traditional setting, the distance reached in each direction is measured to the nearest 0.5cm and averaged after three

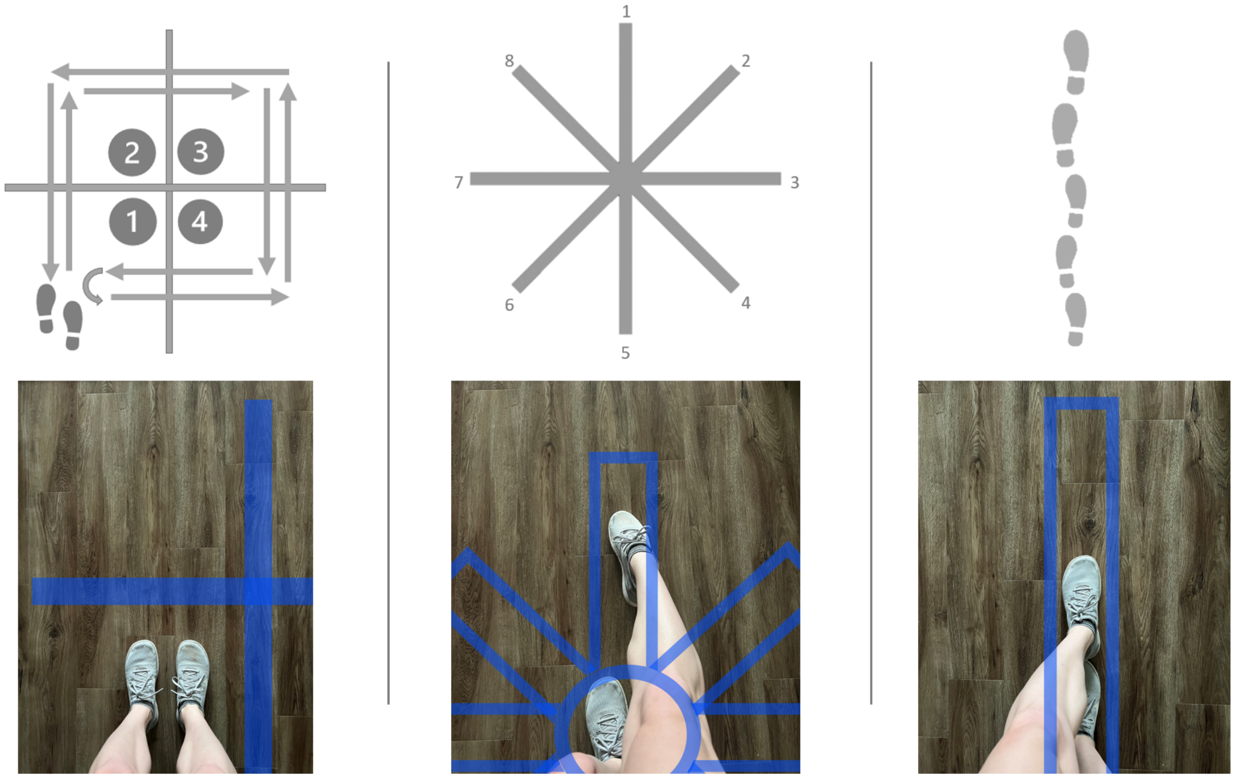

Figure 2.8: Dynamic Balance Assessments. Left: Four Sqauare Step Test. Middle: Star Excursion. Right: Tandem Walking

reach attempts [123]. The AR headset is unable to provide such precision; therefore, the relative locations the user reaches are considered based on a three-score color-coded scale (red, yellow, green). Video data can be captured from the headset's built-in camera to provide a first-person perspective. The video data can be used for review with automated post-processing visual imaging of reach distance and can be evaluated by flight surgeons and EVA flight controllers.

Tandem walking, a heel-to-toe dynamic walking task, features a holographic beam (1.82 meters long) projected on the floor. The tandem walking assessment evaluates heel-to-toe walking within the specified volumetric constraints, 2x2 meters. The Microsoft HoloLens 2 does not feature foot tracking, and therefore, the feet can occlude the holographic content placed on the floor. To mitigate occlusions, the tandem walking beam and other floor-based holograms such as the star excursion object, are hollowed out to present a skeleton-like structure where the feet can be placed with minimal hologram obstruction (Figure 2.8).

The operational balance sub-module incorporates three EVA-based tasks (Figure 2.1, 2.9), namely capsule ingress and egress, obstacle weave, and a postural task similar to that required for geology sampling. The operational balance assessments require dynamic balance during operationally relevant tasks that may be performed on future lunar missions.

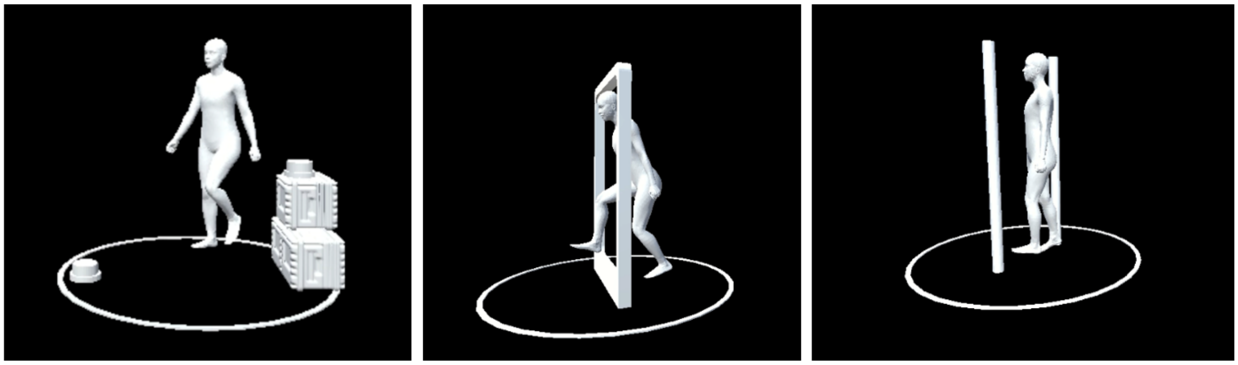

Figure 2.9: Operational Balance Assessments. Left: Geology Sampling. Middle: Capsule Ingress and Egress. Right: Obstacle Weave

The capsule ingress and egress task requires the participant to ambulate through an airlock portal while minding their posture to allow clearance for their head and the top of the airlock. The participants are instructed to step through the airlock with one leg, turn 180 degrees, and step through the airlock again with the opposite leg to evaluate the relative stability of both legs as support legs during the dynamic motion. The airlock task simulates an astronaut entering or exiting their spacecraft upon landing on a planetary body, a difficult task immediately following a gravity transition from extended exposure in microgravity. Although the participant will not feel physical contact with the holographic airlock, audio sounds can be used to indicate if the participant's head comes in contact with the top of the airlock.

The obstacle weave task requires participants to weave around two objects placed on the floor. The task simulates astronauts ambulating on a planetary surface which will require dynamic directional changes in movement to avoid large geology samples or craters. The participant is instructed to weave in a figure eight pattern to allow for both rotations about their right side and left side.

The geology sampling task comprises four sequences of dynamic movements: 1) walking to a geology sample location, 2) bending down and/or kneeling to press a button object to acquire a sample, 3) standing up and turning around 180 degrees, and 4) returning to the origin or sample return location. Geology sampling will be a primary mission objective for future lunar missions. Understanding performance during the four distinct sequences may inform which aspect of the task is most affected by sensorimotor impairment. It is likely that astronauts will be required to kneel throughout mission operations, even with extended grip tools. These operational balance assessments seek to evaluate dynamic balance during operationally relevant tasks that will be performed on future Lunar EVAs. Using AR in conjunction with physical object manipulation could make these tasks more similar to those encountered in operations.

2.2.2 Sensorimotor Tasks

The sensorimotor sub-modules assess two-dimensional, three-dimension, and random target acquisition. The 2D target acquisition assessment evaluates Fitts speed-accuracy trade-off law within an AR environment. The 2D hand-eye multidirectional tapping task, adapted from the ISO9241-9 standard [89] with 16 targets rather 25, is an extension of the Fitts paradigm with the primary benefit of controlling for the effect of target direction [6]. The hand-eye task features 16 targets arranged equidistantly in a circular array with the principal aim of tapping the targets as quickly and accurately as possible. The sequence in which the participant acquires the targets follows a predefined tapping pattern that alternates the active target in a clockwise procedure across the full diameter of the array with the starting and ending positions at the apex of the array. Two indexes of difficulty (ID), defined by the movement distance (diameter of the array) and the target widths, are implemented. The nominal index of difficulty incorporates a diameter of 0.152 meters and a target width of 0.025 meters, yielding an index of difficulty of 2.824 bits (Figure 2.10). The more challenging index of difficulty comprises a diameter of 0.229 meters and a target width of 0.019 meters, yielding an index of difficulty of 3.700 bits. The selection of IDs is in line with the 2-8 bits range of IDs employed within the literature [124, 125, 6]. For interface push buttons with fingertip activation, NASA specifies a minimum diameter of 0.010m when barehanded and 0.019m when gloved [107]. Our selected indexes of difficulty align with the NASA standards where the more difficult target size (0.019) matches that of the gloved hand minimum requirement and the nominal target size (0.025m) is slightly larger than the barehanded minimum requirement.

The array of 16 targets is holographically projected into the user's physical space, where world-anchoring ensures the array's position and size remain constant irrespective of the user's physical movements (Figure 2.10). In this manner, the participant is allowed to position themselves at an appropriate personalized arm-length distance to interact with the target space. Only pointing interactions with the index finger of the dominant hand is permitted. Raycasting, the distant manipulation AR gesture interaction, is disabled within the augmented scene to enforce the use of the pointing gesture. Consistent use of the pointing gesture supports comparisons between hand-eye coordination tasks across touchscreen-based devices in Fitts' Law research.

The 3D target acquisition and random target acquisition tasks are alternative full-body hand-eye coordination assessments focusing on reaction-based target acquisition. A series of circles and squares appear randomly within the user's view. Participants are instructed to use their dominant hand to touch the circles as quickly and accurately as possible, but not the squares. The 3D target acquisition task extends the interaction space beyond a 2D planar array and requires full arm movement to acquire targets at different depths relative to the user.

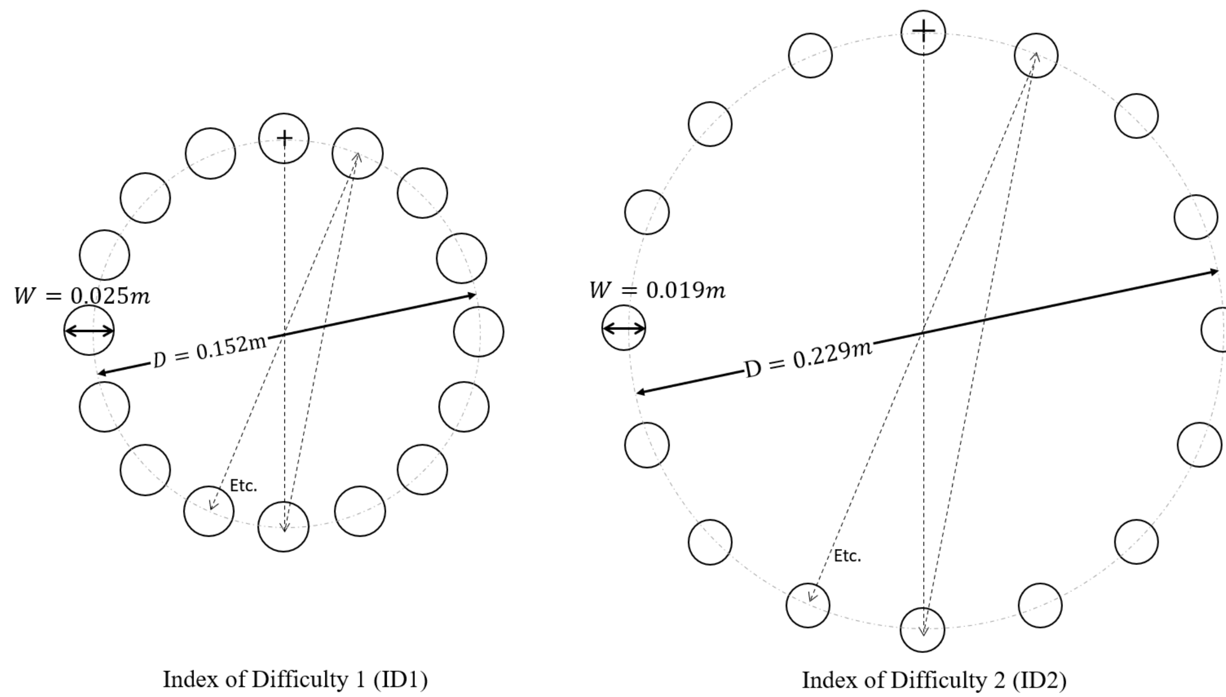

Figure 2.10: 2D Target Acquisition Task, i.e., a multidirectional tapping task with 16 targets. Left: Nominal Index of Difficulty. Right: The more challenging Index of Difficulty.

2.2.3 Mapping of Assessment Tasks to Spaceflight Operations

The assessments integrated into the AURORA system were informed by clinical assessments and functional attributes of tasks crew members may perform during long duration missions. The sensorimotor hand-eye coordination assessments, such as the multidirectional tapping task directly map to intravehicular crew activities such as manual landing procedures and telerobotic operation of robotic hardware such as the Space Station Remote Manipulator System, also referred to as the Canadarm2. Astronauts are required to monitor numerous displays and interact with various interfaces during these operational tasks, potentially under time constraints. Effective hand-eye coordination is imperative for successful completion of manual landing tasks. A representative multidirectional tapping task can provide an opportunity for assessing crew performance inflight prior to execution of the operational tasks to inform readiness. Recent spacecraft systems have integrated touch screen displays, as opposed to physical buttons used within previous spacecraft. The use of AR provides the evaluation and tracking of touch selection gestures, however, lacks the tactile feedback component of a touch screen display.

The dynamic and operational balance assessments mimic functional attributes of extravehicular activity on a planetary surface without the operational constraints such as a spacesuit. If a crew member exhibits difficulties in the simplified augmented reality assessment tasks, further time for adaptation or rehabilitation may need to be implemented in mission operations to support the safe completion of extravehicular activities with the added difficulty of a spacesuit. The dynamic balance assessments, such as the four square step test and the tandem walk tasks, are clinically-based assessments for evaluating fall risk in various populations. Increased risk of falls in an astronaut population inflight could result in serious health concerns such as a higher risk of musculoskeletal injury due to muscle atrophy and reduced bone density. The four square step test evaluates postural and locomotor stability during translations in all directions, a functional capability important for crew members while ambulating on planetary terrain. The tandem walk task seeks to evaluate locomotor performance with a reduced base of support and reduced stability without the support of arm movements which may correlate with suited ambulation. The operational assessments tasks are directly related to operational performance such a geology sampling, obstacle avoidance, and crew ingress and egress from the spacecraft. For these assessments, various body postures such as kneeling, torso and head forward pitch, and rotational motion are required, all of which are important for extravehicular activity tasks.

2.3 Discussion

The development of AURORA focused on supporting first-time users and repeat users. The application utilizes standard augmented reality gestures used for augmented reality devices, and the assessments featured in the application were chosen from clinical gold standard evaluations [108] and tasks currently implemented by NASA for observing crew adaptation timelines pre/post flight on Earth [9, 1]. Assessment instructions were provided in written, audio, and visual forms to support the understanding of task goals. Design decisions made during the development of AURORA were focused on allowing the user to assess their own performance without external assistance and prior experience using AR systems.

Although prominent human factors principles for interface design were consulted and implemented in the design process, usability testing is a crucial aspect of the design process to identify any usability concerns by real users. The following chapter (3) describes the procedures and results of the human-in-the-loop usability testing conducted on the AURORA application prior to the evaluation studies (Chapters 4, 5, and 6).

CHAPTER 3

Human-in-the-Loop Usability Testing

The following chapter presents the human-in-the-loop usability testing of AURORA, a non-intrusive system that employs augmented reality and inertial measurement units (IMU) to evaluate sensorimotor and neurovestibular performance for astronaut populations. Historically, astronaut candidates have been selected between the ages of 26 and 46, with the average age being 34 years old at the time of selection [126]. Upon completion of the selection process, the astronaut candidates undergo a training and evaluation period lasting approximately two years before being inducted into the astronaut corps for their first mission selection. As the astronaut corps exhibits a wide range of ages, the human-in-the-loop usability study aimed to recruit users of similar ages and related educational backgrounds to evaluate the application. In addition to astronauts, other users such as aging populations who experience sensorimotor and vestibular deficits similar to astronauts may benefit from AURORA's assessments. As such, the human-in-the-loop usability testing was evaluated by two distinct user populations, a younger cohort and an older cohort to gather feedback from a wide range of users that align with the target populations listed above. In the two distinct usability studies, the same protocol was followed. The studies were conducted to determine similarities and differences in usability across the age groups and to identify usability issues before implementing AURORA into evaluation studies.

The usability studies focus primarily on nominal usability and do not simulate crew members' physical capabilities and potential decrements while using the system. The work implemented immediate usability feedback, external observations of user interactions, and a post-test survey (A) to distinguish measures of user preference and overall usability of the augmented reality system. The usability study focused on three primary outcomes: (1) how supportive is the AURORA application in self-administering assessments, (2) what is the ease of use and learning perceptions for first-time users, and (3) what usability concerns need to be addressed for future iterations of the application before the integration into evaluation studies. The results from the study's usability testing serve to inform iterative improvements to the application and training protocols before evaluation testing of the system's capabilities for physiological evaluation (Chapters 4, 5, and 6). This work seeks to inform the second research question relating to usability concerns with healthy younger and older adults to improve usability through AR design for promoting the self-assessment of physical abilities.

A peer-reviewed conference proceeding has been produced from the work related to this chapter in the International IEEE Aerospace 2023 conference [127]. In March 2023, the results from the usability study with the younger cohort were presented (Section 3.2). The following chapter synthesizes the results from the younger and older cohorts (Section 3.3) and the emergent usability findings found by comparing the two groups (Sections 3.3.4 and 3.3.5).

3.1 Methods

3.1.1 Participants

Younger cohort. Fifteen University of Michigan undergraduate and graduate students (7 males, 8 females; mean age 23.80 years, SD = 4.59) participated in the usability testing protocol. The research was approved by the University of Michigan Institutional Review Board for Health Sciences and Behavioral Sciences. The participants were informed of the purpose and procedures before the experimental session and provided written informed consent. The participants reported normal to corrected-to-normal vision, no musculoskeletal, auditory, or vesitbular disorders. Eight of the fifteen participants reported English as their first language with the remaining participants reporting Korean (n=2), Mandarin (n=4), and Hindi (n=1). Twelve participants indicated previous experience with augmented and virtual reality platforms such as the HoloLens and Oculus devices.

Older Cohort. Six female adults (mean age 66.3 years, SD = 3.9, min 59, max 70) participated in the usability testing protocol. The research was approved by the University of Michigan Institutional Review Board for Health Sciences and Behavioral Sciences. The participants were informed of the purpose and procedures before the experimental session and provided written informed consent. The participants reported normal to corrected-to-normal vision, no auditory or vestibular impairments, and no history of dizziness, vertigo, or nausea during past use of an immersive system. Participants also confirmed they had not sustained a balance-related fall within six months prior to the study. One participant self-reported a tremor affecting their right arm and hand stability. All six participants reported English as their first language and daily use of a computer. Five participants reported daily use of a tablet or smartphone device, with the remaining participant recording use only a few times per month. Two participants indicated previous limited experience (i.e., single use) with virtual reality platforms.

3.1.2 Protocol

Each participant performed an hour protocol comprised of training and familiarization with the HoloLens device and associated hand gestures, followed by evaluative usability testing. The participants followed the Microsoft-developed HoloLens Tips application to learn the gestures required to interact with holograms, including touch selection, raycasting selection, object pinch manipulation for rotation, scaling, and translation; menu selection; and closing applications. Training comprised a self-paced instructional application for hand gesture interactions followed by verbal instruction from the research staff for application navigation. The headset was individually calibrated (reference the calibration procedure in section 2.1.2) to each user's pupillary distance to ensure a stable and clear presentation of holograms before commencing the evaluation of the AURORA application.

System design and interface questions were asked verbally throughout the usability testing to obtain immediate user feedback and perceptions of device capabilities and hologram interactions. Observational notes of the user's interactions from the research staff's perspective were additionally taken with support from the Microsoft Device Portal by allowing the first-person view of the user to be streamed on a PC device. Qualitatively evaluating the user's ability to use the correct gestures to interact with the system properly helped appropriately evaluate the subjective user feedback obtained throughout the study. A post-test survey was administered at the end of the testing session to obtain overall feedback and perceptions of the system. A custom usability survey (Appendix A) to evaluate AURORA was created based on conventional usability methods such as the System Usability Scale (SUS) [97], the Technology Acceptance Model (TAM2) [128], and the Questionnaire for User Interface Satisfaction (QUIS) [129]. Usability methods explicitly related to augmented reality include the AR acceptance model [130] and the Handheld Augmented Reality Usability Scale (HARUS) [131]. The primary survey sections include overall reactions to the software, the screen, and user interface visibility and interactability, the system's terminology and information, perceived ease of use, system capabilities, intention to use in the future, perceived usefulness/benefit, and comfort physically and mentally. The custom post-test survey questions were administered by a 7-point Likert scale. Previous research suggests the accuracy of data obtained from a Likert scale decreases significantly with point scales below five and above seven [132].

It is important to note that consecutive iterations were made throughout the usability testing; therefore, newer improved interactions were evaluated by later participants. The newer iterations addressed areas of interaction where the application did not perform as intended. For example, navigating from one assessment to another assessment would occasionally cause the application to crash due to a memory leak issue while the IMU data was being logged. This was discovered early on during the first few younger cohort sessions, and a workaround (i.e., temporarily disabling the IMU logging capability) was implemented to mitigate this issue. The newer iterations also addressed emergent usability concerns and feedback discussed in the following sections. Ultimately, the primary usability concerns were discovered early in testing and were addressed for future usability participants.

All neurovestibular assessments that are available in the AURORA architecture were evaluated by the participants beginning with the single-leg left static assessments, followed by the four square step test, star excursion, tandem walking, and operation balance assessments (i.e., capsule ingress/egress, geology sampling, and obstacle weave). The required system interactions included reading/listening to assessment instructions, placing holographic content on the floor for floor-based assessments, starting/stopping assessments, and resetting or navigating to different assessment scenes. One sensorimotor assessment, namely the 2D target acquisition task, was evaluated by the participants in this usability study. The remaining sensorimotor assessments, 3D and random target acquisition, were not included. The 3D and random target acquisition tasks require the same user interactions as the 2D target acquisition task tested; therefore, results may be generalized to the remaining sensorimotor assessments. Verbal questions administered in real-time by the study team focused on the user's ability to understand the assessment with the information provided by the instruction panels and audio instructions.

3.2 Results: Younger Cohort

3.2.1 Training, Learning, and Implementing Gestures

The time required to train the younger participants on the primary hand gestures associated with the headset varied greatly between participants, where some users finished the Microsoft-developed HoloLens Tips gesture training within 5 minutes and others upwards of 15 minutes. Although the users' capabilities improved with extended use of the system, three participants continued to have difficulty successfully interacting with holograms inside and outside of the AURORA application. One participant notably struggled with the hand gestures and spent half the testing session within the training phase. This participant was unable to evaluate the entire AURORA system within the allotted testing session. From observational notes by the research staff and verbal comments acquired from the participants, the primary issue that led to poor interaction with holograms emerged from the user's inability to perceive depth within the headset. It was observed that in some situations, users would attempt to touch select holographic buttons where in reality, their fingers were too far in front to contact the hologram's bounding box, and therefore the click was not triggered by the system. The exact separation between the user's finger and the holograms were not obtained; however, visual streaming from the headset revealed large discernible distance offsets (roughly 0.03 - 0.25 meters determined qualitatively by the research staff) in many cases.

Raycasting, the alternative to touch selection, also posed a significant challenge to six participants. Successfully performing a pinch requires the user to be very deliberate with their finger placement and isolate the index and thumb fingers to support the system's hand-tracking capabilities. In many cases, the participants who struggled with raycasting would use their fists or all fingers to grab holograms, impeding the headset's ability to detect the user's intentions. The inability to act in accordance with the established hand gestures led to the participant's overall difficulty using the headset and, consequently, the AURORA application. It was noted that individuals with previous experience with VR and AR applications (n = 12) did not necessarily correspond to their ability to interact with the HoloLens successfully, although overall, those with some experience appeared to pick up the gestures and training more efficiently. Five of the participants with previous experience satisfactorily finished the training protocol within five minutes where the remaining participants required more time to learn the AR gestures and perform the actions successfully to produce the desired outcomes. The subjective survey measures indicated that eighty percent of participants perceived their experience of learning to operate the application as moderately (n = 8) and extremely easy (n = 4) with similar results for learning the hand interaction gestures (Figure 3.1).

3.2.2 User Interface and Voice Commands

The immediate user feedback following distinct usability tasks within the application provided the most constructive diagnosis of specific usability issues. For example, early iterations of the application featured the start and stop assessment functionalities (i.e., buttons) contained within the assessment's instruction panel that was rendered stationary in the user's space (Figure 2.5). Participants could move the instruction panels with pinch selection or raycasting; however, this required the user's ability to perform the pinch correctly. Users found the buttons difficult to interact with due to difficulties in depth perception mentioned previously and the placement of the static instruction holograms relative to the user's own placement in the space.

To enable easier start and stop interactions, following iterations of the AURORA application featured the start/stop button collection incorporated in a hand menu (Figure 2.5). In this manner, the start/stop button collection is paired to the user and therefore allows for greater mobility and ease of selection when compared to the original semi-stationary instructions panel location. It is noted that two participants continued to struggle with perceiving the start/stop button locations even when these buttons were paired to the hand menu and situated in line with the hand. Five participants evaluated both methods for starting/stopping an assessment in a single testing session

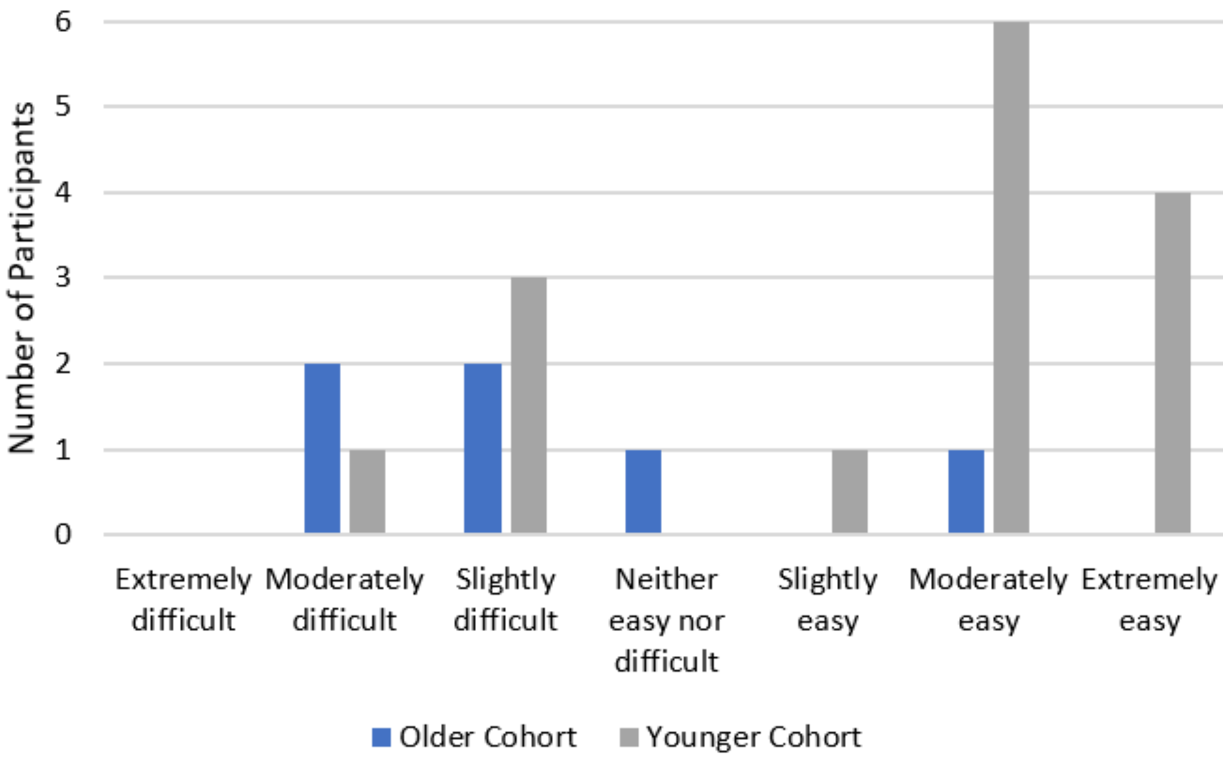

Figure 3.1: Perceptions of learning the hand gestures for both older and younger cohorts.

where half the assessments featured the hand menu interaction and the other half incorporated the stationary menu interaction. The new hand-constrained method was preferred.

Other usability issues identified throughout the testing procedures include the difficulty with reliable voice command recognition using the selected algorithm. Further investigation into the cause of this inconsistency is needed to ascertain whether the issue can be attributed to the headset's voice recognition algorithms, the selected words for voice activation, the application's voice command settings, or a combination of these issues.

3.2.3 Task Assessment Understanding

Other usability issues considered participants' difficulty (n=4) in understanding specific assessment tasks and goals, particularly for the star excursion assessment. Minor updates to the assessment animations and task descriptions were made to improve the users understanding of the tasks. In the post-test survey, the cumulative data suggest that eight participants indicated the terminology used throughout the system was moderately straightforward, six participants indicated that the terminology was extremely straightforward, and one participant (whose first language was Mandarin) indicated that the terms were somewhat confusing. The presentation of task instructions was reported as extremely and moderately straightforward for the majority of the usability participants (n=13). The animations and audio instructions were evaluated as moderately and extremely supportive in understanding task goals, where one participant reported these supportive measures as extremely distracting (Figure 3.2). The majority of the participants rated the information provided within the application as extremely, moderately, and somewhat sufficient if they were to self-administer the assessments in the future.

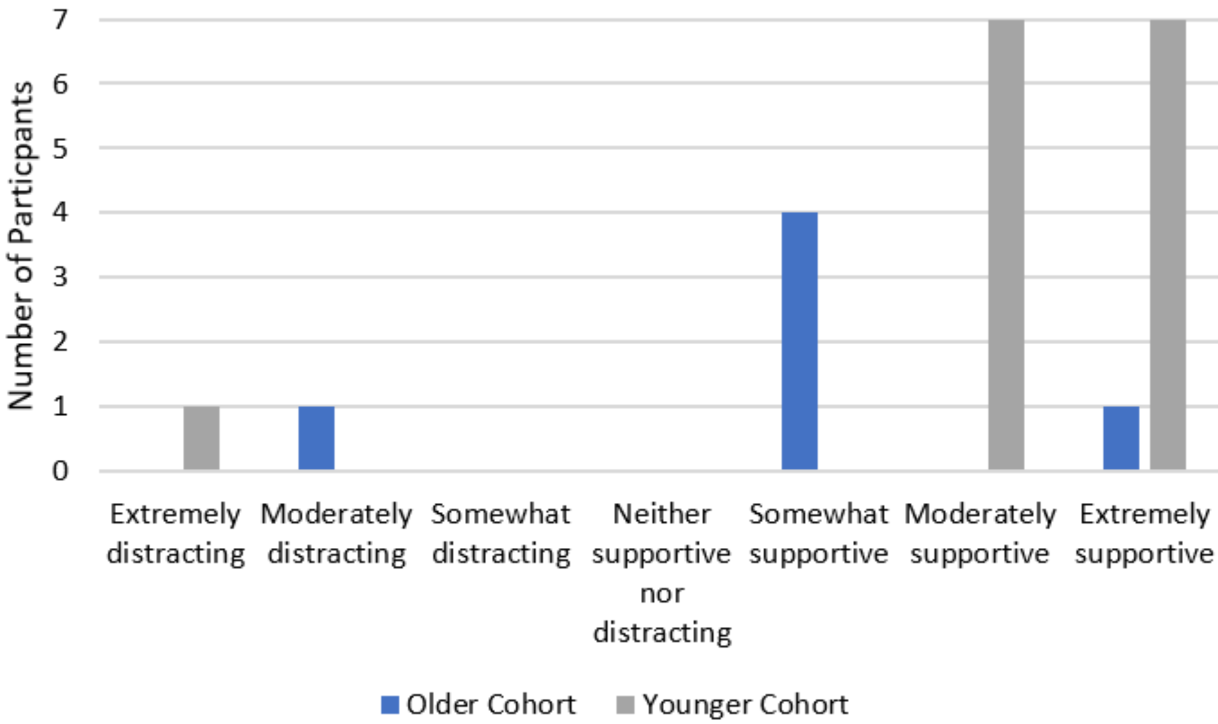

Figure 3.2: Perceptions from the younger and older cohorts for how supportive or distracting the assessment animations and audio instructions were.

3.2.4 Software

The final usability concern, revealed near the beginning of testing, was the headset's propensity to crash and exit when navigating from one assessment to the next. The post-test survey data verifies the frustration of the four participants that experienced crashes, with three users deeming the system navigation moderately unreliable and one user reporting the system navigation as extremely unreliable. The headset's crash issue was isolated to a concern with memory management when the inertial measurement unit (IMU) sensor was streaming and the assessment scenes were loaded and unloaded. While this issue was debugged after usability testing, for the remainder of usability testing participants (n = 10), the IMU stream was disabled to prevent system crashes. The disabled IMU stream resulted in no further system crashes and survey measures for system navigation reliability as slightly (n = 2), moderately (n = 6), and extremely reliable (n = 3) (Appendix B).

3.2.5 Effort and Future Use Survey Responses

The survey results indicate that most participants perceived their level of mental and physical fatigue as somewhat, moderately, and extremely low (Figure 3.3). Overall perceptions towards the use of the AURORA application were positive, with users reporting the system as moderately (n = 8) and slightly easy (n = 4) to use. Similarly, the participants' overall experience or attitude towards AURORA was positive, with nine reports of moderately good and two reports of extremely good (Appendix B). When asked to consider whether they would use the system again in the future for physiological assessments, no one selected never use. The majority of participants indicated that they would sometimes use (n = 8), almost every time use (n = 3), and always use (n = 1) the application in the future if it was available to them (Appendix B). The majority of the participants reported that they would strongly (n = 5), moderately (n = 4), and slightly recommend (n = 4) the AURORA application to others, where one user would moderately discourage and one user would neither discourage nor recommend the application (Appendix B).

3.3 Results: Older Cohort

3.3.1 Training, Learning, and Implementing Gestures

The time required to train the older cohort of older participants on the primary hand gestures associated with the headset was relatively consistent at roughly fifteen minutes in length. One participant continued to have difficulty interacting with holograms both inside and outside of AURORA due to difficulties in executing hand gestures. The participant lacked the physical dexterity to perform some of the gestures that required more intricate hand motions and relatively stable execution of gestures (e.g., raycasting). Two other participants were able to perform the gestures but had difficulty with hologram interactions due to an inability to perceive depth within the headset, according to the research team's observations and the participants' verbal feedback. It was discovered that in some cases, users would attempt to touch specific holographic buttons, but their fingers would be too far to touch the hologram's bounding box, so the system would not trigger the click. The survey measures indicated that four of the six participants reported difficulty learning the hand gestures (Figure 3.1). Participants perceived their experience of learning to operate the application as slightly (n = 2) and moderately difficult (n = 1), neither difficult nor easy (n = 1), slightly easy (n = 1), and extremely easy (n = 1).

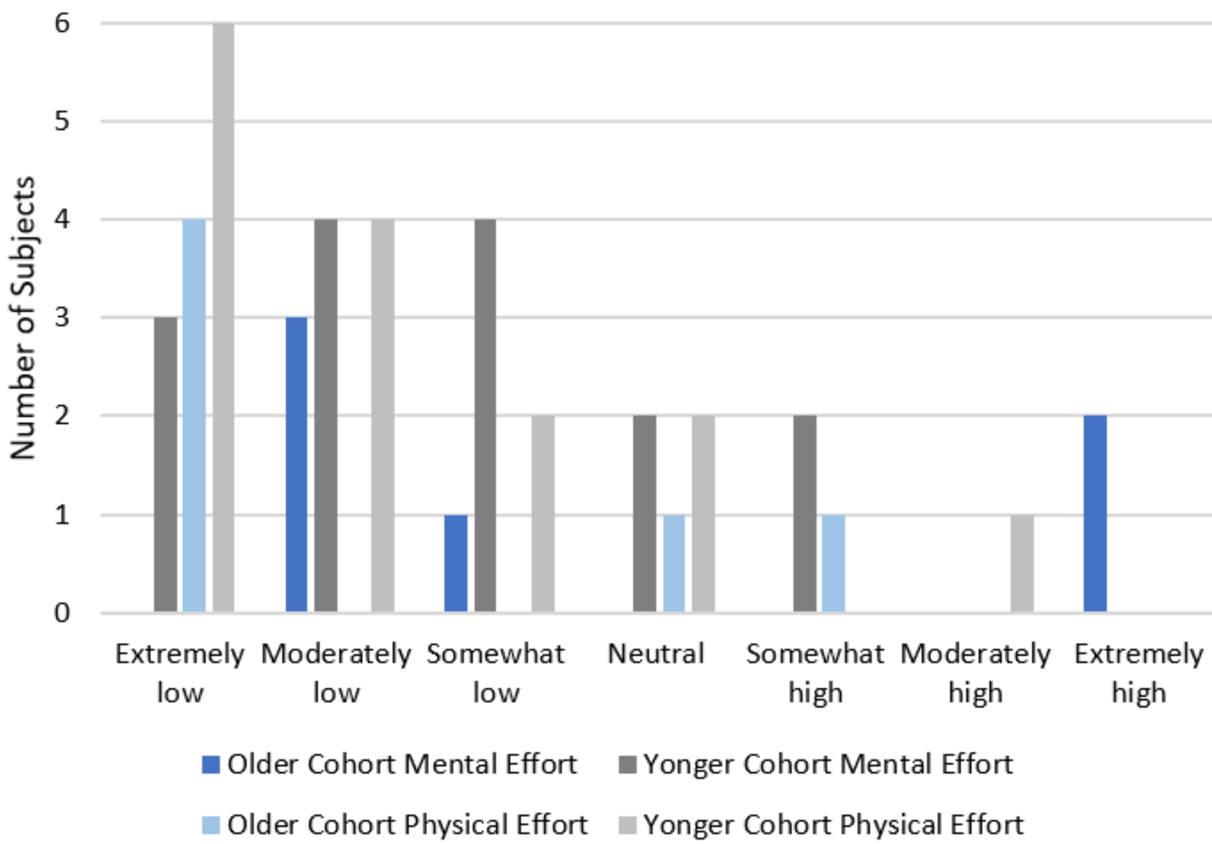

Figure 3.3: Perceived level of mental and physical effort for older and younger cohorts.

Following a variety of application-based usability tasks, immediate user feedback was used to diagnose specific usability issues. For instance, the left-hand-attached menu required the user's left palm to be positioned in front of the headset and selections to be made by the right hand. One participant who suffered from a tremor, affecting their ability to stabilize their right arm with respect to the left, had difficulties successfully interacting with the hand menus. The participant opted for an unanticipated tactic where she braced her hands on a table to suppress the tremor while attempting to acquire selections on the hand menu. This method was not reliably effective and required a great deal of effort to position the head and arms to support both visualization of the menu and articulation of the hands while being braced on the table.

3.3.2 Task Assessment Understanding

The cumulative results of the post-test survey indicate that the presentation of task instructions was deemed extremely (n = 1), moderately (n = 3), and somewhat straightforward (n = 1), with the remaining participant reporting the instructions as neither confusing nor straightforward. The human body animations and audio instructions were evaluated as somewhat (n = 4) and extremely supportive (n = 1) in understanding task goals, where one participant reported these supportive measures as moderately distracting (Figure 3.2). Many participants chose to read along while the audio instructions dictated the task; however, for tasks that required more information and could not be described completely in the written panels, the audio and written instructions differed slightly. The participant who reported the audio instructions as distracting noted that she preferred to read instructions rather than have them dictated at the same time. She felt she was unable to process the task instructions at her preferred pace when administered via audio instruction. In contrast, the participant who reported the audio and animations as extremely supportive felt she understood the assessment task instructions better with the visual and audio modalities. Most of the participants rated the information provided within the application as extremely (n = 1), moderately (n = 2), and somewhat sufficient (n = 1) if they were to self-administer the assessments in the future. The remaining participants rated the information as moderately (n = 1) and extremely (n = 1) insufficient.

3.3.3 Effort and Future Use Survey Responses

The survey results indicate that the older participants perceived their level of mental effort as somewhat (n = 1) and moderately low (n = 3), while two participants reported their mental effort as extremely high. Four of the six participants indicated their physical effort as extremely low, while the remaining two reported neutral and somewhat high (Figure 3.3). Overall perceptions towards the use of the AURORA application indicated that users perceived the system as moderately easy (n = 1), neither easy nor difficult (n = 2), slightly difficult (n = 2), and moderately difficult (n = 1). The participants' overall experience or attitude towards AURORA was positive with two reports each of somewhat, moderately, and extremely positive. When asked to consider whether they would use the system again in the future for physiological assessments, half of the participants (n = 3) selected always use. The remaining participants indicated that they would sometimes use (n = 1), almost every time use (n = 1), and never use (n = 1) the application if available. The participants reported that they would strongly (n = 2), moderately (n = 2), and slightly recommend (n = 1) the AURORA application to others, whereas one user would slightly discourage.

3.3.4 Comparisons Between Younger and Older Cohorts

While the older cohort was consistent with the time required to finish the training module, the younger cohort in the previous study conducted at the University of Michigan varied greatly from five minutes to over fifteen minutes for one participant. The majority of the younger cohort reported learning the hand gestures as easy; however, four out of the fifteen participants had difficulty with depth perception and performing the gestures as instructed (Figure 3.1). Two of the six older participants had difficulty with depth perception, whereas one other participant struggled with physically implementing the gestures due to a tremor. The younger cohort reported the audio sounds and animations as extremely and moderately supportive, whereas the older cohort had more diverse responses (Figure 3.2). Two older subjects responded that they would like further information to support self-administration, a request not made by the younger cohort. Mental effort was perceived to require a higher effort in older adults compared to the younger cohort (Figure 3.3). For physical effort, one participant from each cohort reported a rating above neutral, i.e., either slightly or moderately high. Most participants in both cohorts reported their physical effort as extremely low. Overall, both cohorts reported positive experiences towards the application as a whole and intentions to use the system again in the future if provided (Appendix B).

3.3.5 Discussion

The usability study focused on three primary outcomes: 1) how supportive is the AURORA application in self-administering assessments, 2) what is the ease of use and perceptions of usability for first-time users, and 3) what usability concerns need to be addressed for future iterations of the application before the integration into future evaluation studies. The results from the usability testing of twenty-one participants supported all three primary outcomes.

Future long-duration crews will be required to operate more autonomously from ground support; therefore, the ability to self-administer assessments is crucial. It was evident from the post-survey subjective usability measures that the terminology and task information provided (i.e., task animations paired with text and audio descriptions) in the application were sufficient for 13 of the 15 younger usability participants to feel comfortable self-administering assessments in the future. The remaining two younger participants felt the information provided was somewhat and moderately insufficient. These two participants, along with two additional participants (n = 4), reported difficulty in understanding specific assessment tasks and goals. The two participants who reported difficulty in both understanding the tasks requirements and the information being insufficient to support self-administration of assessments were not native English speakers. Anatomical terminology used throughout the assessment instructions (e.g., anterior/posterior and medial/lateral) were difficult to understand for these two users. Subsequent iterations of the task information included simplified terminology (e.g., forward/backward) to support users without human performance and anatomical knowledge. In AURORA, audio instructions are similar to the operations procedures for extravehicular activity on the International Space Station, where crew members carry minimal written reference procedures and flight operations support personnel provide immediate verbal instructions [133]. The simplified embedded instructions and overall user feedback indicate that the tool supports self-administration of assessments and may support astronaut performance assessments for future missions.

As augmented reality technology advances and more consumer-grade devices are available to the public, there is the potential for individually-owned headsets enabling remote self-care and clinically guided in-home treatment. An assessment system that supports self-administered tasks would be beneficial for in-home treatment, rehabilitation, and performance monitoring. Similar to the younger cohort, it was evident from the post-survey usability measures that the terminology and task information provided (i.e., task animations paired with text and audio descriptions) were sufficient for four of the six older usability participants to feel comfortable self-administering assessments in the future. The remaining two older subjects reported the information provided as moderately and extremely insufficient for self-administering; nevertheless, they were able to perform the tasks as instructed. These two participants preferred research staff support to administer the assessments as they were able to ask questions in real time about the task goals and how to interact with the application. Among the two older participants, one suggested an additional training module to support autonomous future use.

A distinct training module following gesture training and prior to usage of AURORA could outline important actions that were verbally administered by the research staff, such as navigating to assessments, starting/stopping assessments, shifting gaze to the right of the instructions panel to view the animations during the dictation of instructions. The second participant who reported the system's instructions as insufficient for self-administration reported a preference for external support by someone like the research staff. The findings suggest that while most of the participants felt comfortable using the system again in the future without additional guidance (i.e., external or through embedded training modules), the addition of such features may benefit users who were unable to attain the desired knowledge and usage of the system in the allotted time frame.

While usability for all users was desired, there were difficulties for some participants using the interaction modalities for both the younger and older cohorts. Generally, researchers select gestures for their optimum recognition rather than for their naturalness, resulting in arbitrary and unintuitive gesture interfaces [134]. However, extended use of an AR system has been shown to improve performance over time [135]. The users who had the most trouble interfacing with the system arose from their inability to perform the hand gestures as instructed and their inability to perceive the hologram locations correctly. Participants who struggled with raycasting used fists or all fingers to grab holograms. Participants who struggled with touch selection failed to isolate their index finger. These imprecise user behaviors impeded the headset's ability to detect the user's gestures and intentions. Participants' inability to conform to the established hand gestures contributed to difficulty using the headset and, consequently, the AURORA application. A majority of the participants (n = 11) perceived learning the hand interaction gestures and navigating the application as extremely to moderately easy. The ratings obtained from the perceptions of learning were consistent with the observed performance of the hand gestures. For the remaining participants who reported learning these aspects of the application usage as slightly and moderately difficult, other supportive measures such as interface design changes (e.g., less strict requirements of the gestures with additional permissible variability in the gesture postures) and improved training procedures may be warranted. In the future, custom gesture detection methods could be used instead of the HoloLens' traditional gesture detection algorithms. New training protocols that focus on constructive feedback and cueing will be implemented to support users with difficulty learning and executing hand gestures. Virtual demonstrations of proper hand placement and more practice interactions can be provided for further support. The amount of additional practice required will need to be tailored to each user's needs, as demonstrated by the wide range of training times required within this study.

The survey results indicate that most participants reported learning the hand gestures as slightly and moderately difficult, even though all participants were able to interact with the system. Older participants might have felt less confident about their abilities because they rated themselves as having difficulties even when researchers observed them interacting well. The younger cohort's perceptions were more in line with the researcher's observations. The younger and older users who had experienced difficulty perceiving hologram location correctly could benefit from future designs with additional depth cues. One older participant struggled with a tremor which impeded their ability to execute the gestures that required both hands. Although AURORA was initially intended for astronauts, the deconditioning and physiological changes observed in astronaut populations due to extended exposure to microgravity may be akin to aging populations and musculoskeletal disorders that affect one's ability to perform intricate controlled gestures. Further efforts to refine the voice command capabilities are expected to improve the application's usability overall by allowing users to bypass hand interactions in favor of verbal input. Ultimately, the development of future augmented reality systems and devices that leverage gesture control needs to consider the potential mobility limitations of the target users.

Although many of the hand gestures required for the augmented reality system are commonly used in everyday activities, such as choosing options on a touch screen or pinching a physical object, the lack of tactile feedback in the augmented environment may combine with limited depth perception to hinder performance. Users who struggled most found it difficult to distinguish the location of holograms correctly. Four participants found their ability to perceive hologram locations relative to global space exasperating. Users who interacted directly with the holograms were required to effectively perceive the hologram location to ensure direct contact with the index finger and the hologram's bounding box. The alternative method, i.e., distant manipulation with the use of raycasting, enabled users to interact with holograms at a distance, although users continued to struggle with depth perception of the extended hand ray relative to stationary holograms. This interaction technique is unique to augmented reality as humans do not naturally interact with the real world from a distance. However, practical depth perception is still required to interact with the holograms. Future studies aligned with depth perception and AR use may be warranted to determine if natural depth perception capabilities may predict capabilities within an AR environment.

Common methods in augmented reality systems for substituting tactile feedback are audio and visual cues to support hologram interactions [136]. With AURORA, after the user selects or interacts with a button, they receive immediate audio (e.g., tonal click sound) and visual (hologram surface color modulation) feedback, letting them know the interaction with the hologram has been initiated. For sensorimotor tasks, a holographic skeletal overlay of the user's hand can be used for target selection and touch interactions to provide visual feedback on the location of the tracked joints; however, when integrated into the system for pilot testing, it was observed that the systems' tracking capability was inadequate. As the hand moved faster, the discrepancy between the hand and the holographic overlay increased. It was preferred in the application to use a floating circle pointer over the index finger to enable the user to locate the system's tracked fingertip relative to their actual fingertip (Figure 2.3). The user was able to compare the location of the floating circle pointer to the relative surface location of instructions/button holograms for depth cues. Additionally, proximity lighting, where the surface shading of the button changes color as the finger approaches, aimed to improve depth perception. All the user interfaces evaluated by the participants featured proximity lighting, with the exception of the targets for the sensorimotor task. As these design choices were not enough to support all users in perceiving depth, there is an opportunity to explore additional visual, audio, and tactile cues to improve depth perception related interactions.

Although the device supports users with glasses and the device was properly calibrated to each user's pupillary distance, it was noted by the research staff that some of those who required corrective lenses (either glasses or contact lenses) may have had trouble visualizing the holograms. Thirteen of the fifteen participants reported wearing corrective lenses, however, only a handful of these participants (n=4) experienced depth perception issues. To view a stable hologram, the HoloLens uses stereoscopic rendering to provide depth cues. Each eye receives an image of the hologram, which requires effective convergence between the respective eyes to perceive depth [137]. Research from operations on the International Space Station suggests that crew members' visual-spatial perception of distances is generally underestimated in orbit compared to Earth [138]. Astronauts have also demonstrated greater depth perception instability after adjusting to weightlessness [139]. Perceptual instability may result from a lack of sensory information about spatial orientation relative to the vertical and from top-down processes implicitly developed with prior knowledge of Earth's gravity. While limited flight data has been collected with crew members and HoloLens use in orbit, no major issues have been explicitly reported with hologram perception

[133]. There is a need to further evaluate eye changes in space in context with AR systems to quantify how vision decrements may impact user performance.

The application interface underwent several iterations throughout usability testing to improve system use between each successive usability participant. The most prominent system-wide overhauls were the start and stop assessment interactions and the interaction for users to place ground-based assessment holograms on the floor with sliders rather than spatial mesh perception and ray-casting. In its current implementation, the voice command capabilities were considered unreliable and inadequate for use. Further efforts to refine the voice command capabilities are expected to improve the application's usability overall by allowing users to bypass hand interactions in favor of verbal input.

3.3.6 Limitations and Future Studies

Limitations of the usability studies include the small sample sizes and demographic background of the usability participants. Despite recruiting from a specific demographic (students with an engineering background), we acknowledge that the participants are not fully representative of the education, capabilities, experience, and other qualities of astronaut populations for which the system is intended. While the second usability study had an initial sample of older adults who reported high digital literacy, four of the participants experienced AR for the first time during the protocol. Although measures were taken to mitigate acquiescence bias through neutral wording selection, this bias to agree with researchers may also be a limitation of the usability studies. Nevertheless, critical feedback was provided and used to update the system. A new version of AURORA, informed by the usability results presented, was used to validate the AR tasks for the assessment metrics of interest (Chapters 4, 5, and 6).

Although usability was assessed for healthy non-disoriented participants, the intended use of AURORA also includes astronauts in disoriented or deconditioned states. Future studies can use the developed usability protocol for additional user scenarios, such as healthy individuals undergoing vestibular disruptions or those suffering from vestibular disorders to identify any usability concerns that may arise from disoriented or deconditioned users.

3.4 Conclusion

Due to the disruption of the sensorimotor and neurovestibular systems in microgravity, astronauts exhibit significant impairments in their functional abilities that require distinct adaptation timelines. Performance decrements can significantly affect operations during and shortly after gravity transitions when performance risks are greatest. In-flight assessment tools are needed to support crew autonomy and extravehicular activity readiness. These assessment tools require specific usability considerations to support self-assessment of user abilities, through intuitive interactions, visibility of displays, and consistent user actions to support the retention of how to use the device potentially months after initial training.

This chapter presented the human-in-the-loop usability testing of AURORA. The AR system may be a viable solution to the resource, volumetric, and time constraints of space operations by allowing holographic visual cueing to replace physical objects used in traditional Earth-based assessments. Based on the usability results, the overall user experience of the AURORA system shows promise for integration in astronaut assessment. Users felt the system was able to describe assessment requirements and support task completion effectively. There were several usability issues that were discovered and resolved throughout testing. For instance, the interaction to start and stop assessments was modified to a hand-constrained menu to support more mobility and ease of use. The most significant issue discovered was not software-based but rather how the users were interacting with the software. Some participants struggled to properly execute the trained hand gestures, thereby limiting their ability to interact effectively with the system. According to the findings, both younger and older users who had difficulty learning the required hand gestures may benefit from design changes to the interaction modalities (i.e., alternative or more permissive hand gesture postures) to support easier usage. Additional instructional or training protocols may also support users in learning and practicing the correct hand gestures to interface with the system effectively. Some users exhibited issues with depth perception, therefore training should also increase the exposure to holograms placed at a distance to provide more opportunities for users to correctly locate holograms within the application. Beyond training, improved interface design to support depth perception is warranted through audio or visual cueing (i.e., proximity lighting).

The results suggest that the augmented reality application is a potentially useful tool for supporting and facilitating the evaluation of astronauts' neurovestibular and sensorimotor performance throughout mission timelines. The findings also suggest that the augmented reality application could be a useful tool for assisting in the assessment of the neurovestibular and sensorimotor performance of aging populations or those who experience decrements in sensorimotor ability.

This chapter informs design guidance for AR mission display devices in NASA's Human Information Design Handbook (HIDH) [110] in Chapter 7, Section 7.2. Chapter 4 informs specific requirements for button size to support accuracy and efficiency to inform section 10.4, i.e., controls, for future versions of the NASA HIDH. Further evaluation of the quantifiable metrics and viability of the application for physiological evaluations is discussed in the following Chapters 4 , 5, and 6 where the AR-specific tasks were compared to gold-standard Earth-based measures.

CHAPTER 4

Evaluation of Augmented Reality Sensorimotor Assessments

The following chapter presents the comparative analysis of a hand-eye coordination assessment within AR compared to a touch screen device. This work informs the first and third research questions relating to the use of AR as a standalone assessment tool for sensorimotor evaluations by identifying the impact of holographic targets on task performance.

In this study, the performance of healthy adults on a multidirectional tapping task was evaluated within an AR environment and a traditional touchscreen device while standing. The research considered two main questions: (1) how performance differs with varying indexes of difficulty and (2) how virtually presented holographic targets impact users' accuracy, precision, movement time, throughput, and error rates when compared to a touchscreen device. It was hypothesized that the touchscreen modality would yield better accuracy, precision, movement time, throughput, and error rate performance when compared to the AR modality with the same task parameters. It was hypothesized that the larger target width utilized in the first index of difficulty (ID) would yield better performance across modalities. This study has important implications for future designers of AR applications, particularly for sensorimotor assessments that require hand-eye coordination, as it prompts designers to consider target size, movement distance, gesture interactions, and depth perception cues during the design process. In addition, the results inform the interpretation of hand-eye coordination within AR as compared to other interface modalities and support the use of AR as an assessment tool.

4.1 Methods

4.1.1 Participants

Thirty-two University of Michigan students participated in this experiment (13 female and 19 male, mean age 22.6 years ± 3.3 SD, min: 18, max: 31). All participants reported normal (n=16) to corrected-to-normal vision (n = 16) and reported no musculoskeletal, auditory, or vestibular disorders. Nine participants wore glasses while five participants wore contact lenses. Twenty-five participants were right-hand dominant, six left-hand dominant, and one participant reported being ambidextrous. The participants were informed of the procedures before participating and signed the informed consent approved by the University of Michigan Institutional Review Board for Health Sciences and Behavioral Sciences. All participants reported daily usage of a touchscreen device (i.e., tablet, smartphone, etc.) Twenty-one participants reported previous limited experience (between 1 and 10 hours) with extended reality devices (i.e., augmented reality or virtual reality); the remaining participants were considered novice users.

4.1.2 Software and Hardware Configuration

The experiment evaluated hand-eye coordination in the form of a two-dimensional multidirectional tapping task (Figure 4.1) within an AR environment in the Microsoft HoloLens 2 device and using a 23.8-inch touchscreen monitor. The AR environment projected the virtual array of targets 0.5 meters from the designated origin of the augmented space (the user's initial head position), which is set at the startup of the application. The virtual target array was aligned perpendicularly to the user's head position, with the array's center aligned to the user's eyes. In the augmented reality environment, the targets were holographically projected into the user's physical space, where world-anchoring ensures the array's position and size remain constant irrespective of the user's physical movements. In this manner, the participant is allowed to position themselves at an appropriate personalized arm-length distance to interact with the target space. The touchscreen monitor was positioned in a similar fashion where the center of the array was placed at the participant's eye height. Participants were permitted to move as closely as needed within the physical and augmented environments to complete the task.

4.1.3 2D Multidirectional Tapping Task

The 2D hand-eye multidirectional tapping task was adapted from the ISO9241-9 standard [89] and used 16 targets rather than 25 following protocols from previous literature [124, 140]. The 2D multidirectional task is an extension of the Fitts paradigm with the primary benefit of controlling for the effect of target direction [6]. The multidirectional task features 16 targets arranged equidistantly in a circular array with the principal aim of tapping the targets as quickly and accurately as possible. The sequence in which the participant acquires the targets follows a predefined tapping pattern that alternates the active target in a clockwise procedure across the full diameter of the array with the starting position at the apex of the array (Figure 4.1).

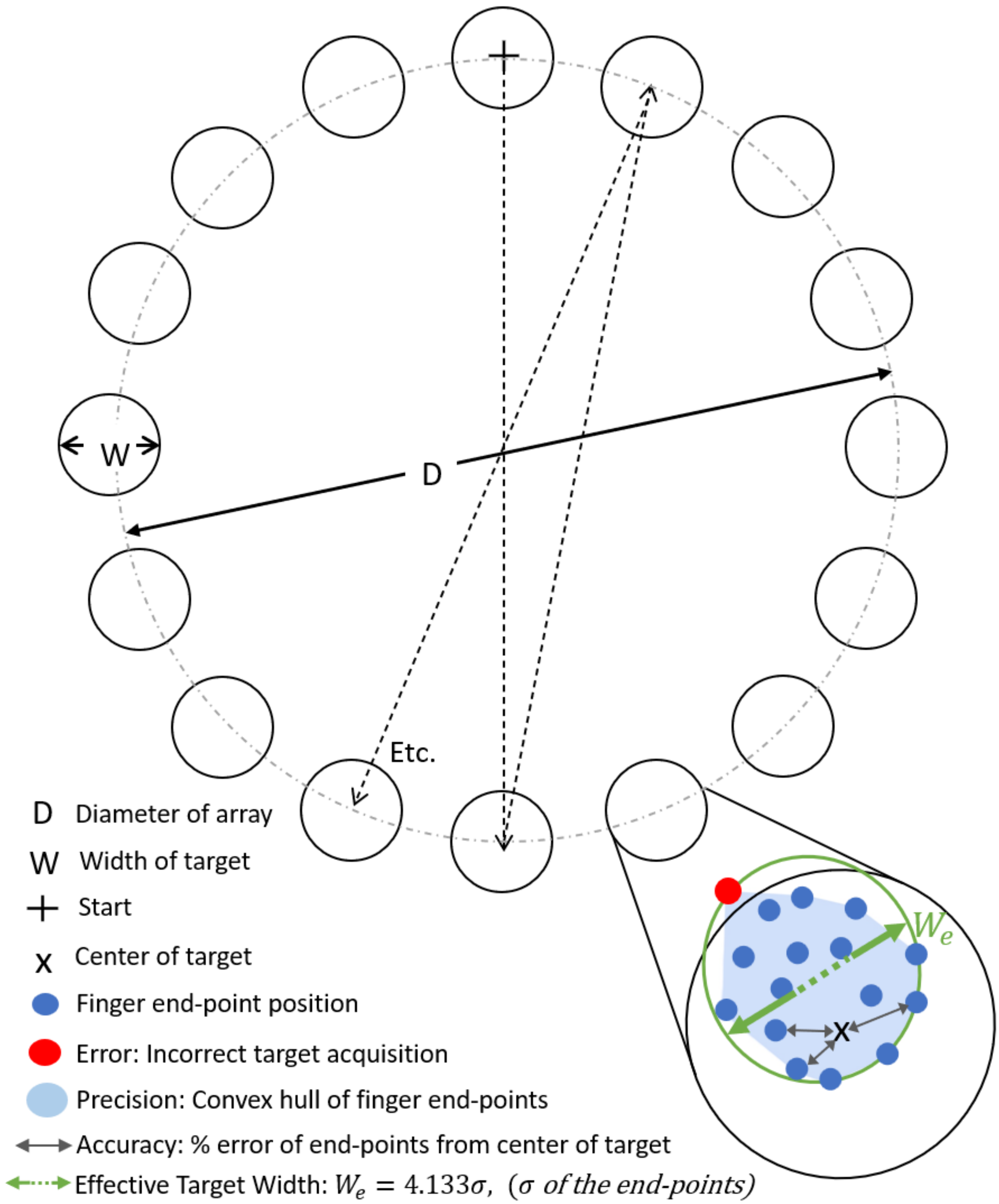

Figure 4.1: 2D multidirectional tapping task. The dashed line depicts the predetermined pattern of target acquisitions. The zoomed-in region presents a schematic of relevant metric information for an example target and is accompanied by the key to the left. Measures were referenced from [6]

.

Two indexes of difficulty (ID), defined by the movement distance (diameter of the array) and the target widths were implemented using the Shannon formulation [141, 6],

$$ID = log_2(\frac{D}{W} + 1) \tag{4.1}$$

where D is the movement distance distances and W is the target widths. The nominal index of difficulty incorporates an array diameter of 0.152 meters and a target width of 0.025 meters, yielding an index of difficulty of 2.824 bits. The target width was selected in accordance with NASA's Man-Systems Integration Standards (NASA-STD-3000) and the NASA Human Integration Design Handbook for push buttons [142, 107]. The more challenging index of difficulty comprises a diameter of 0.229 meters and a target width of 0.019 meters, yielding an index of difficulty of 3.700 bits. The selection of IDs is in line with the 2-8 bits range of IDs utilized within the literature [124, 125, 6].

Only pointing interactions with the index finger of the dominant hand were permitted. AR gesture interactions for distant manipulation were disabled within the augmented scene to enforce the use of touch selection and the pointing gesture. To ensure that the headset successfully tracked the finger joints and accurately registered target acquisitions, participants were instructed to keep their index finger positioned vertically (roughly 45 degrees above the horizontal) rather than perpendicular to the target (Figure 2.3). The tracking of the finger was indicated to the participant through a virtual white circular pointer at the estimated fingertip location. Target acquisition was dependent on the participant's tracked fingertip location relative to the bounds of the target. The next target within the sequence would be active after the current active target was pressed by the user's index fingertip. A press was registered when the fingertip-tracked position entered the outer boundary of the target mesh. The touchscreen monitor registered target acquisitions and progressed through the predefined target acquisition sequence when the fingertip contacted the touchscreen within the active target's bounds.

4.1.4 Protocol

The participants performed an hour protocol consisting of training and familiarization with the hardware and multidirectional tapping task, evaluative testing, and survey responses. Each participant was randomly assigned one of four treatment groups that specified the order in which they performed the hand-eye coordination task with respect to modality (AR vs. TS) and index of difficulty (ID1 vs. ID2). Eight participants were distributed within each of the four treatment groups. The 2D hand-eye multidirectional tapping task was organized into blocks, where one block consisted of five full rotations of the array, i.e., 16 target selections, otherwise referred to as a sequence.

The training period required the participants to perform one block (five sequences) for each index of difficulty and within each modality. To minimize learning effects, practice sequences were performed prior to testing. The training period familiarized the participant with the consistent predefined tapping pattern and the nuances of each index of difficulty and modality. The participant was instructed to use their dominant index finger to complete the tapping task.

The evaluative testing period required the participants to perform three blocks (fifteen sequences) for each index of difficulty and within each modality. Participants performed 240 target selections for each condition (960 selections in total) so that a central tendency of each participant's performance within each condition could be estimated. Participants were permitted to take breaks between sequences and blocks as needed.

Surveys were administered after the completion of each modality to gauge the user's perceived workload and fatigue irrespective of index order. The post-modality survey incorporated questions from the NASA Task Load Index [143] and a question regarding arm fatigue. Arm fatigue was captured with a five-point user perception question with the following ratings, no [0], mild [1], moderate [2], extreme [3], and unbearable arm fatigue [4].

4.1.5 Performance Measures

We considered standard measures for Fitts' law target acquisition tasks to evaluate our hypotheses, including accuracy, precision, movement time, throughput, and error rates [6]. Accuracy within target was defined as the percent difference in the distance between the tap position and the center of the target as

$$Accuracy = \frac{R - P}{R} * 100 \tag{4.2}$$

where R is the radius of the target and P represents the distance between the tap position and the center of the target. Zero accuracy represented the edge or outside of the target. Precision, the measure of variation between finger end-point locations within the target bounds, was evaluated within and across targets by finding the convex hull of the points within the targets (using the convhull function [144] to compute the 2-D convex hull of the points in column vectors x and y). Movement time (MT) refers to the amount of time required for the participant to make a successful attempt, which is the time difference between two successful taps. We do not include the time from the initial start to the first target as this period has added variability due to the initial posture of the participant at the start of the trial. Throughput, a measure that combines both movement time and accuracy, was calculated first for each participant and then the grand throughput [6] was evaluated by averaging across participants as:

$$TP = \frac{1}{y}\sum_{i=1}^{y}(\frac{1}{x}\sum_{j=1}^{x}\frac{IDe_{ij}}{MT_{ij}}) \tag{4.3}$$

where y is the number of participants, x represents the number of movement conditions, and IDe is the effective index of difficulty adjusted for accuracy by incorporating the effective target width and effective movement distance based on where the participants actually selected [6]. IDe is calculated as:

$$IDe = log_2(D_e/W_e + 1) \tag{4.4}$$

where D_e is the distance between the center of the effective target and W_e is the circle fit diameter of each effective target. We gathered the effective target by enclosing every tap position for a target by using a circle fit function [145]. The effective measures are smaller than the designed target widths and distances, resulting in an increase in the effective index of difficulty. By incorporating movement time and accuracy, throughput represents a unified measure for evaluating our hypotheses of the movement performance within each modality and index of difficulty. Error rates capture instances where the participant tapped outside the target boundary. For the touchscreen modality, the error was constrained to 2D; however, in AR the error was evaluated in 3D as the participant's fingertip may have been located within the target boundary along the horizontal axis, but they may not have intersected the target along the depth axis. In these cases, participants required additional movement of their finger closer to the target to successfully trigger a target acquisition. Error rates characterize the number of taps outside of the target with respect to the total number of taps as:

$$E = 1 - \frac{\sum_{i=1}^{C} S_i}{C} \tag{4.5}$$

where S is the result of whether it is a successful tap (0 for unsuccessful or 1 for successful), and C is the total number of attempts.

4.1.6 Statistical Analysis

Performance measures were assessed using a multifactor Analysis of Variance (ANOVA) with factors Participant (random), Index of Difficulty (fixed), Modality (fixed), and Task Order (fixed). If a significant effect was found in the ANOVA, post-hoc paired t-tests were performed. To address multiple comparisons (six post-hoc comparisons per metric), the level of significance was adjusted to $\alpha = 0.008$. The evaluated null hypothesis states there is no significant difference between AR and touchscreen (TS) modalities for all performance measures. Paired t-tests were performed to compare survey results between the AR and TS modalities. For these pairwise comparisons, level of significance was $\alpha = 0.05$. Effect size for t-tests was calculated using the Robust Cohen's d method and effect size for f-tests was calculated using the partial eta squared method.

4.2 Results

One participant's AR data was excluded from the analysis due to excessive noise and fluctuations in the recorded hand-tracking data. A separate participant had four trials removed as they did not follow directions. All participants were instructed to use their dominant index finger throughout the testing protocol; however, one participant switched hands during four AR trials. For this participant, these four sequences were not included in the analysis. Evaluations of performance across the time series of testing trials revealed no evidence of continued adaptation. This suggests the training protocol was sufficient for asymptotic performance.

4.2.1 Accuracy

Average accuracy decreased as the difficulty of the task increased (change from ID 1 to ID 2) $F(1, 81) = 159.35$, $p < 0.001$) and decreased when using AR compared to Touchscreen ($F(1, 81) = 1323.48$, $p < 0.001$) (Figure 4.2). There was no interaction effect for the Index of Difficulty with the Modality (Appendix C, Table C.1). The ANOVA model supported a significant interaction effect of the Modality and Task order. This interaction was further investigated, and it was found that within Order, the effect of Modality followed the same trend but had a differing magnitude. This interaction between Modality and Order did not affect the findings inferred from the main effects.

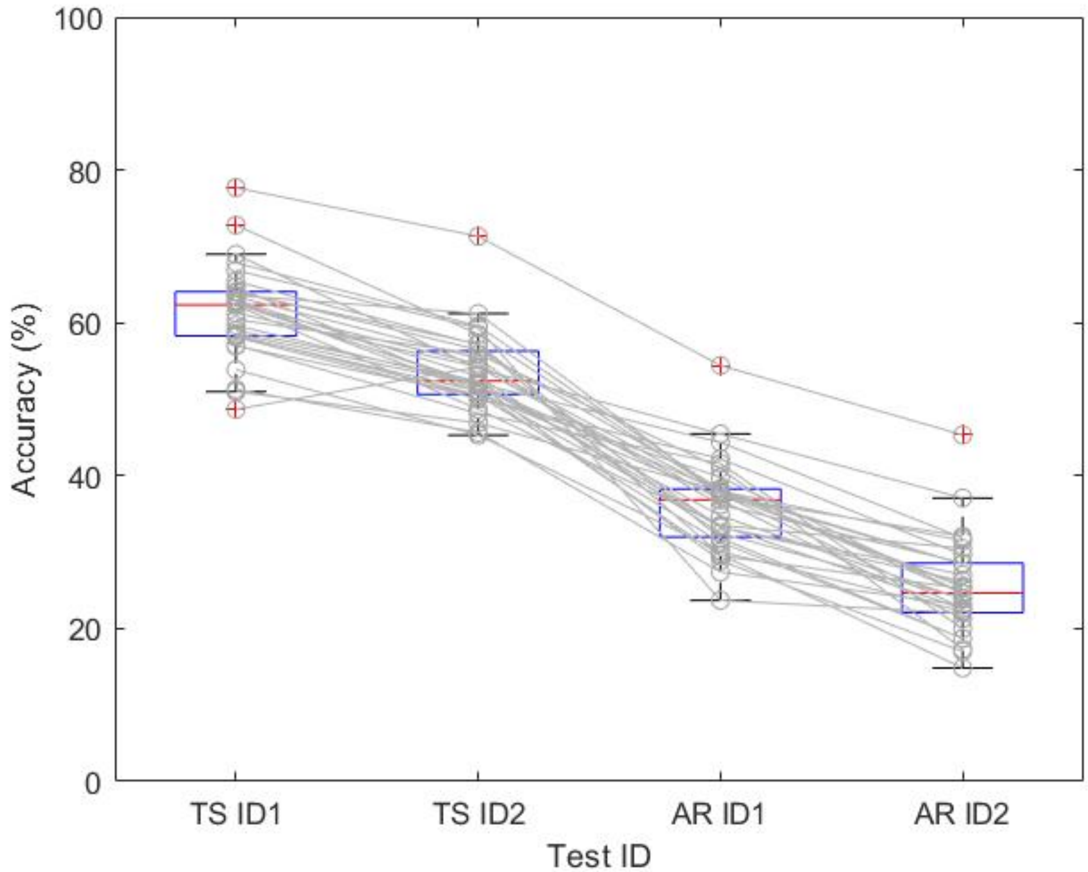

Figure 4.2: Box plots of the average accuracy of all the participants (n = 31) with respect to Index of Difficulty and Modality. All conditions were significantly different from each other (all of the p-values were smaller than 0.001.)

4.2.2 Precision

Precision was represented through the convex area of the data points. The convex area of the finger's end-point positions increased as the difficulty of the task increased (change from ID 1 to ID 2) under the AR environment ($t(30) = -3.42, p = 0.0018, d = 0.78$), while it remained similar under the Touchscreen environment ($t(30) = 0.405, p = 0.689, d = 0.28$). The overall convex area increased when using AR compared to Touchscreen for both ID1 ($t(30) = -4.78, p < 0.001, d = 1.87$) and ID2 ($t(30) = -5.71, p < 0.001, d = 1.30$) (Figure 4.3). There was a significant interaction effect for the Index of Difficulty with the Modality (Appendix C, Table C.2), and both Task Difficulty and Modality were significant main effects.

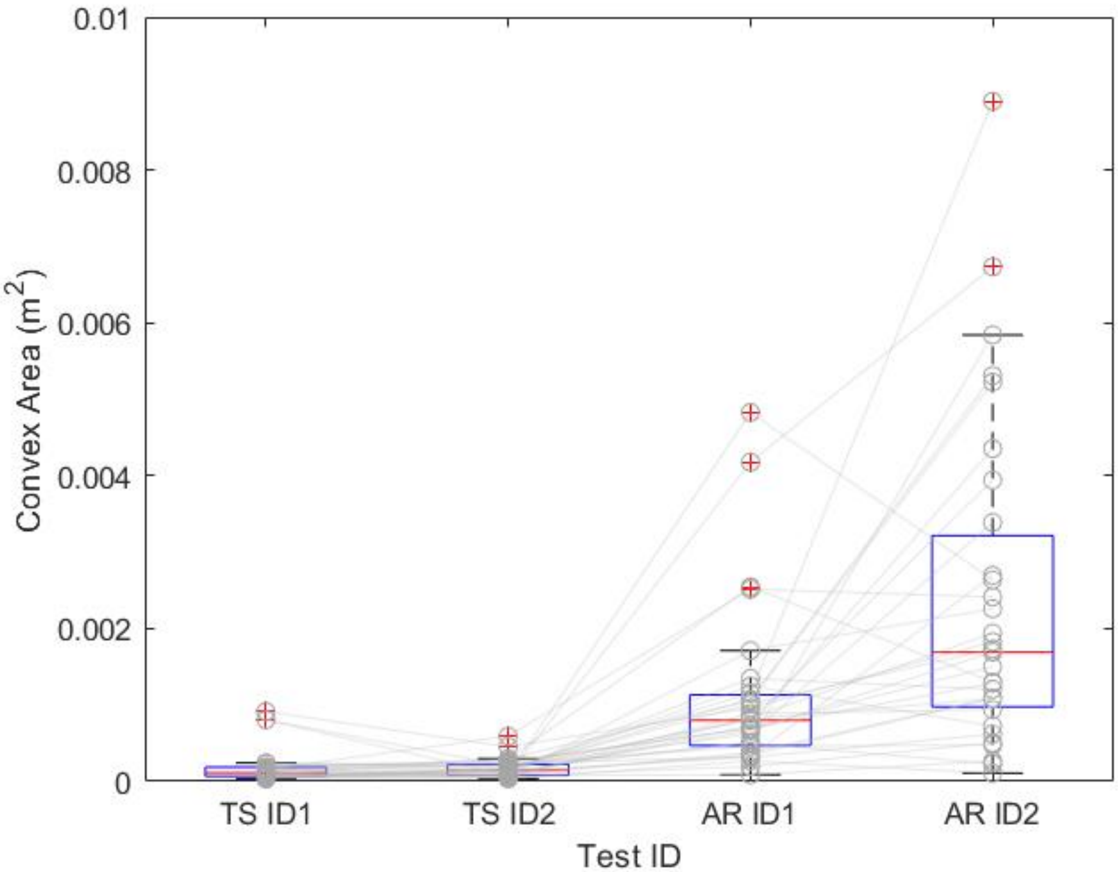

Figure 4.3: Box plots of the overall convex area of all the participants (n = 31) with respect to Index of Difficulty and Modality. TS ID1 was not significantly different from TS ID2 ($p = 0.689$). All other comparisons were significantly different from each other (AR ID1 vs. AR ID2, $p = 0.0018$, all other comparisons $p < 0.001$).

4.2.3 Movement time

Movement time increased as the difficulty of the task increased (change from ID 1 to ID 2) for AR ($t(30) = -11.44, p < 0.001, d = 3.70$) and for the TS ($t(30) = -18.73, p < 0.001, d = 2.78$). For both ID 1 ($t(30) = -24.63, p < 0.001, d = 5.81$) and ID 2 ($t(30) = -19.07, p < 0.001, d = 7.86$), movement time increased when using AR compared to Touchscreen (Figure 4.4). There was a significant interaction effect for the Index of Difficulty with the Modality (Appendix C, Table C.3), and both Task Difficulty and Modality were significant main effects.

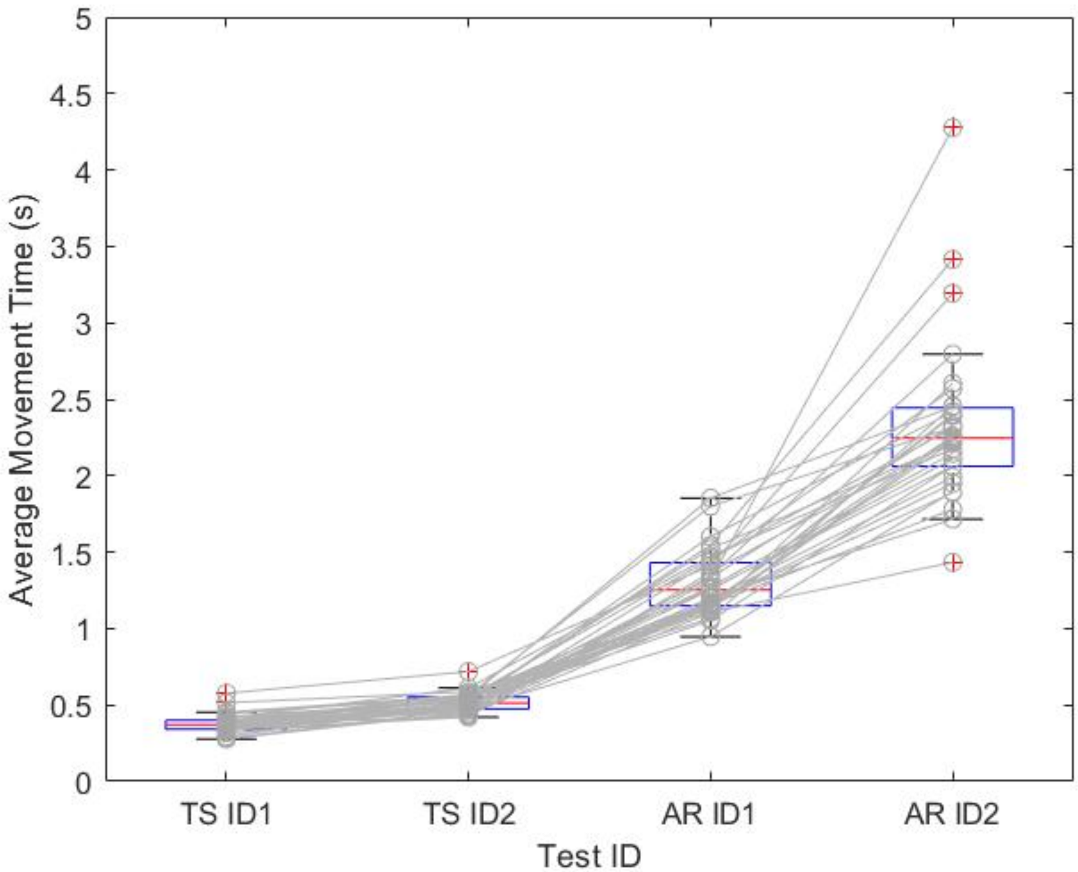

Figure 4.4: Box plots of the average movement time of all the participants (n = 31) with respect to Index of Difficulty and Modality. All conditions were significantly different from each other (all of the p-values were smaller than 0.001.)

4.2.4 Error rates

Error rates increased as the difficulty of the task increased (change from ID 1 to ID 2) for both TS ($t(30) = -5.614, p < 0.001, d = 1.22$) and AR ($t(30) = -9.633, p < 0.001, d = 1.40$). For both ID 1 ($t(30) = -15.67, p < 0.001, d = 3.28$) and ID 2 ($t(30) = -18.76, p < 0.001, d = 3.96$) error rates were higher when using AR compared to Touchscreen (Figure 4.5). There was a significant interaction effect for the Index of Difficulty with the Modality (Appendix C, Table C.4), and both Task Difficulty and Modality were significant main effects.

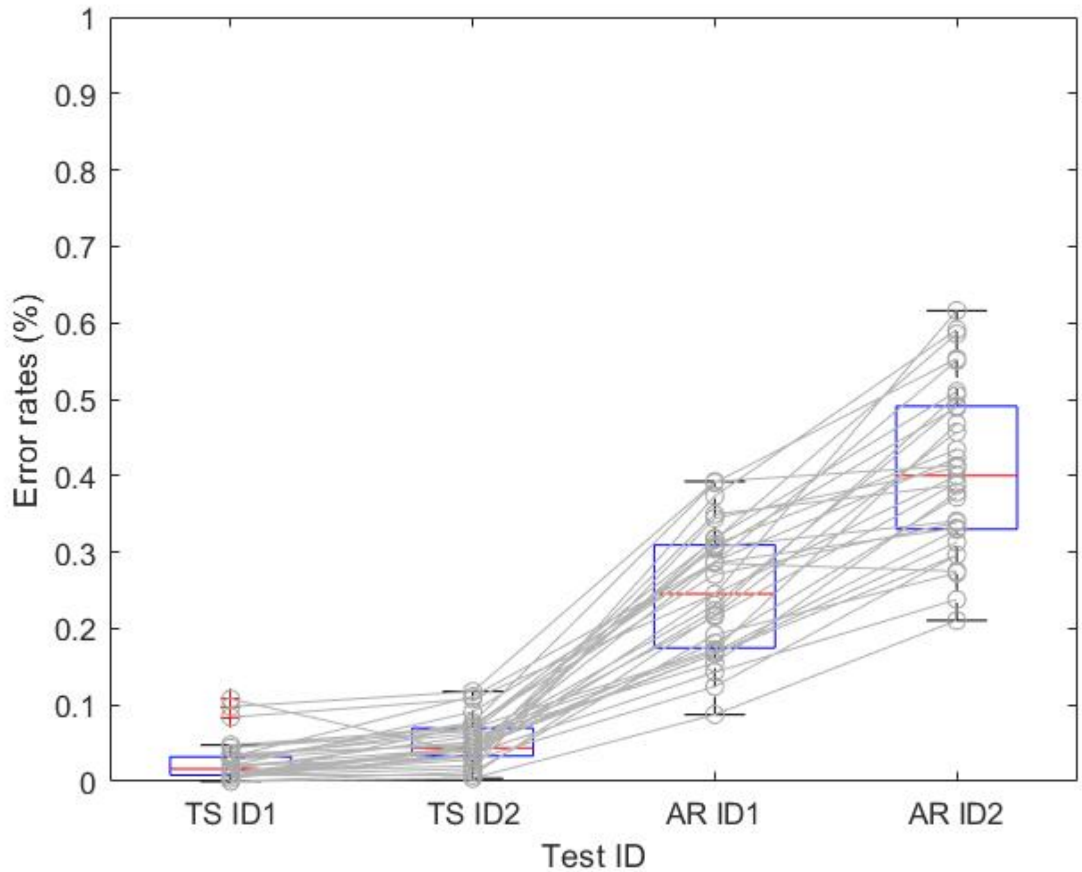

Figure 4.5: Box plots of the error rates of all the participants (n = 31) with respect to Index of Difficulty and Modality. All conditions were significantly different from each other (all of the p-values were smaller than 0.001.

4.2.5 Throughput

Throughput decreased as the difficulty of the task increased (change from ID 1 to ID 2) for both TS ($t(30) = 6.82, p < 0.001, d = 1.01$) and AR ($t(30) = 9.226, p < 0.001, d = 1.66$). For both ID 1 ($t(30) = 27.08, p < 0.001, d = 7.16$) and ID 2 ($t(30) = 37.48, p < 0.001, d = 8.44$) throughput decreased when using AR compared to Touchscreen (Figure 4.6). There was a significant interaction effect for Index of Difficulty with the Modality, the Index of Difficulty with the Task Order, and the Modality with the Task Order (Appendix C, Table C.5), and both Task Difficulty, Task Order and Modality were significant main effects.

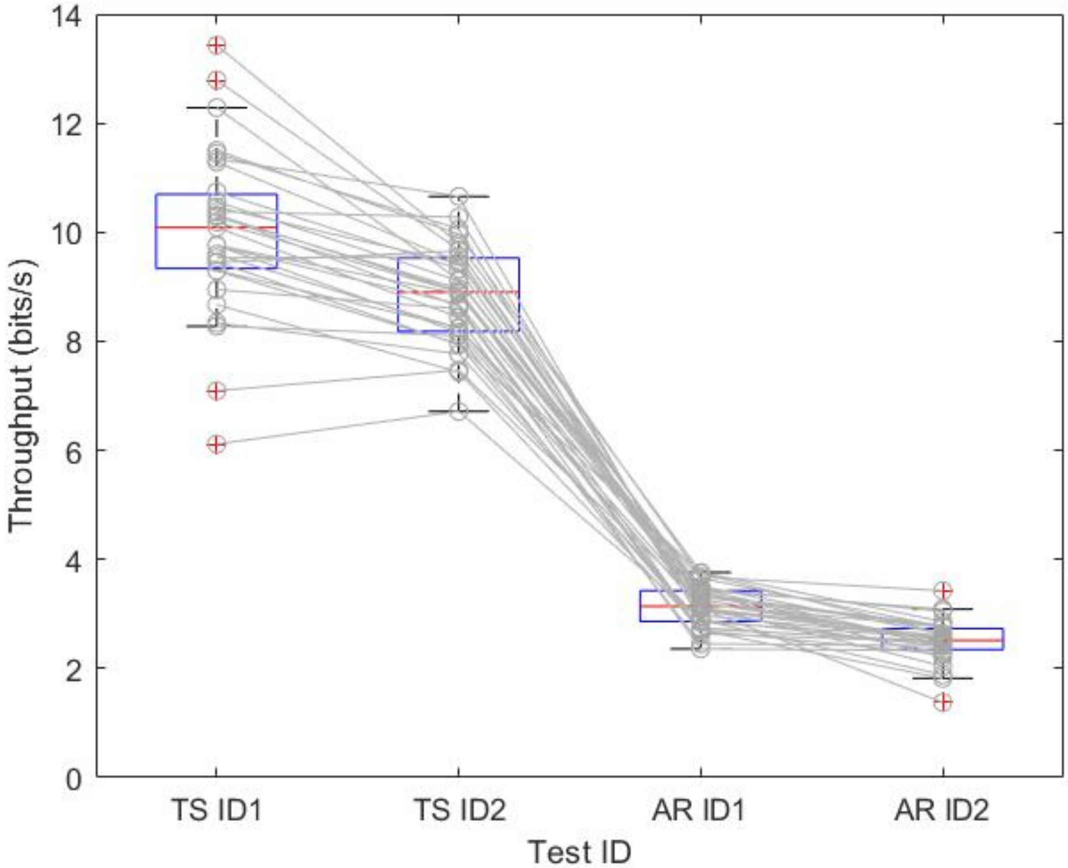

Figure 4.6: Box plots of the throughput of all the participants (n = 31) with respect to Index of Difficulty and Modality. All conditions were significantly different from each other (all of the p-values were smaller than 0.001.

4.2.6 Survey Result

There was a significant difference in perceived physical demand between AR ($M = 42.13$, $SD = 24.07$) and TS ($M = 21.53$, $SD = 14.98$); $t(31) = -6.02$, $p < 0.001, d = 1.00$), where M is the mean and SD is the standard deviation (Figure 4.7). There was a significant difference in perceived mental demand between AR ($M = 32.69$, $SD = 24.27$) and TS ($M = 21.59$, $SD = 17.85$); $t(31) = -2.86$, $p = 0.007$, $d = 0.51$, $d = 0.51$). There was a significant difference in perceived performance between AR ($M = 71.23$, $SD = 17.54$) and TS ($M = 90.77$, $SD = 8.94$); $t(30) = 7.65$, $p < 0.001$, $d = 1.37$). Due to an incorrect interpretation of the scale, one participant's data was removed for perceived performance. There was a significant difference in perceived effort between AR ($M = 51.97$, $SD = 23.97$) and TS ($M = 26.81$, $SD = 19.86$); $t(31) = -6.71$, $p < 0.001$, $d = 1.08$). There was a significant difference in perceived frustration between AR ($M = 22.75$, $SD = 24.82$) and TS ($M = 7.75$, $SD = 9.17$); $t(31) = -4.40$, $p < 0.001$, $d = 0.75$). There was a significant difference in perceived arm fatigue between AR ($M = 1.63$, $SD = 0.61$) and TS ($M = 0.84$, $SD = 0.51$); $t(31) = -7.27$, $p < 0.001$, $d = 1.32$). For AR use, participants rated no (n = 1), mild (n = 10), moderate (n = 19), extreme (n = 1), and unbearable arm fatigue (n = 1). For TS use, participants rated no (n = 7), mild (n = 23), and moderate arm fatigue (n = 2).

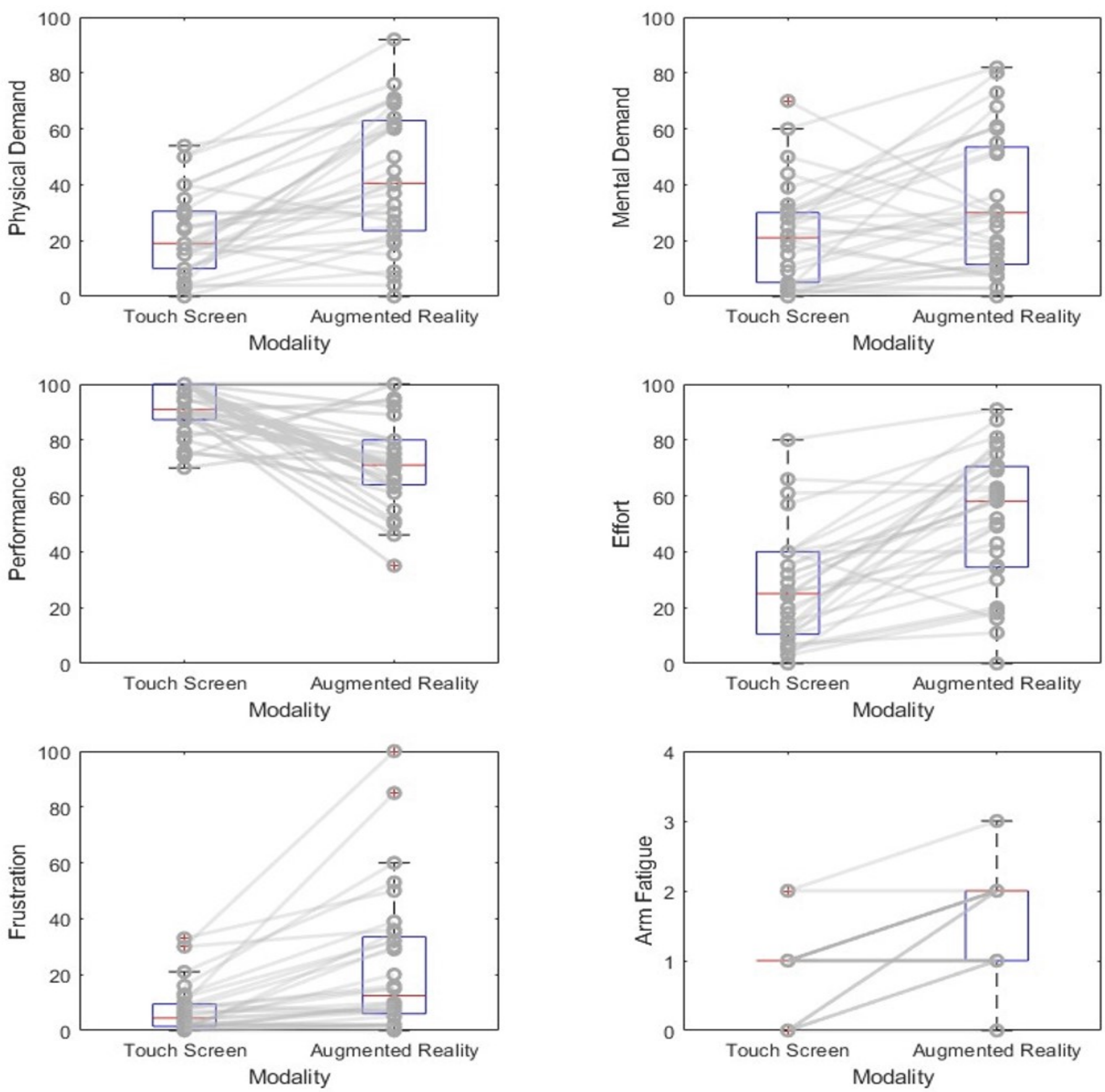

Figure 4.7: NASA TLX survey results depicting user perceptions of physical demand, mental demand, performance, effort, and frustration while performing the task across the two modalities. Bottom right: Perceptions of arm fatigue. All measures were significantly different between the AR and TS modalities.

4.3 Discussion

The research considered two main questions: (1) how performance differs with varying indexes of difficulty, and (2) how holographic targets impact users' accuracy, precision, movement time, throughput, and error rates when compared to a touchscreen device. We hypothesized the larger target width utilized in the first index of difficulty (ID) would yield better performance across modalities. We also hypothesized the touchscreen modality would yield better performance across all dependent measures when compared to the AR modality.

The results support our first hypothesis that the lower index of difficulty (i.e., larger targets with smaller movement distance) would yield better performance. Within both modalities, the average throughput, accuracy, and precision of participants decreased while the error rates and movement time increased when the difficulty of the task increased (change from ID1 to ID2). The larger targets, coupled with the shorter movement distances, resulted in higher accuracy, precision, and throughput as well as a decrease in error rates and movement times, consistent with the literature [146, 125, 147]. For the Touchscreen, the convex area was consistent across IDs, indicating that participants used a smaller region within the button, even when space was available with a larger target width.

The results also support our second hypothesis that the touchscreen modality would yield improved performance across the dependent measures. Results support that AR has an increased movement time, decreased accuracy, and increased error rates as compared to the touchscreen device. Although the targets were arranged along a 2D plane in both modality environments, the added complexity of the three-dimensional selection in AR was revealed to be more difficult for users, thereby leading to higher error rates and the emergence of multiple taps per target in some cases. Participants modulated movement time to achieve a level of accuracy to support target activation within both modalities; however, the magnitude of this modulation was different across modalities. Although AR increased movement time, results did not support an improvement in accuracy, precision, or error rate when compared directly to the touchscreen. The AR environment was more challenging and required modulation of the user's strategy to maintain performance due to difficulty with hologram interactions, gesture recognition, and visual perception. The self-reported perceptions of the participants, captured in the post-test survey, aligned with the direct measures analyzed. Notably, there was a higher perceived effort, physical demand, mental demand, frustration, arm fatigue, and a decrease in perceived performance while using AR. Several participants anecdotally reported that their frustration, arm fatigue, and physical demands increased due to difficulty interacting with and acquiring the holograms, resulting in multiple taps per target. We found stronger negative ratings for AR in this study as compared to our previous usability study [148], where the majority of participants reported low overall mental and physical demands. The multidirectional tap task was one of eight assessments evaluated by the participants in the previous usability study. It is possible that the usability participants' perceptions of effort may have differed because they did not have a touchscreen device to anchor their comparison of performance and personal experience. Although the participants in this study were permitted to take rest breaks between blocks, many chose to continue even though arm fatigue was likely to increase.

Throughput, which combined speed and accuracy, was significantly lower within the AR conditions for both indexes of difficulty when compared to TS. The grand throughput while performing the nominal index of difficulty (ID1) yielded 10.12 bit/s and 3.11 bits/s for TS and AR, respectively. The grand throughput while performing the more difficult index of difficulty (ID2) yielded 8.90 bit/s and 2.49 bits/s for TS and AR, respectively. A previous AR study [88] with an estimated index of difficulty of 2.6 (target width of 46 cm, and distance of 237.6 cm) observed a lower throughput of 1.52 bit/s even though their study had a slightly lower index of difficulty than the nominal index in this study. These differences may arise from the AR devices used as the Magic Leap One device utilized by [88] has been experimentally shown to have a smaller field of view and poorer spatial accuracy (i.e., the mean distance between a target and the sensed fingertip location) than the HoloLens 2 device [149] With respect to other input modalities, our results suggest that performance on the 27-inch monitor (10.12 and 8.90 bits/s) was higher than that observed on a mobile touchscreen (ID = 1.14 bits to ID = 3.17 bits) for a multidirectional tapping task with 20 targets [90]. Our results suggest that the AR modality yielded throughput values higher than that of joystick (1.6-2.55 bits/s), touchpad (0.99-2.9 bits/s), and mouse input (3.7-4.9 bits/s) [6]. Although a true comparison cannot be made due to the different indexes of difficulty and the potential difference in participants, the results can still provide insight into the range of performance for distinct input modalities.

The tactile feedback and physical touch limits offered by the touchscreen monitor may have supported the improved accuracy over AR where the finger was free to travel through the target. Research has shown that haptic feedback, relating to the sense of human touch, has improved accuracy with typing on touchscreen devices [150], and mid-air haptic feedback in a Virtual Reality (VR) context was found to increase grasp accuracy when dealing with small targets [151]. As with accuracy, a similar trend was observed where precision performance decreased when AR was used. The precision decrement was greater between indexes of difficulty for the AR modality than TS. The convex hull of the endpoint positions suggests users pressed closer to the center of the targets while using the TS device than while using the AR device. Hand gesture interactions, tracking capabilities, and gesture speeds may have contributed to the increased movement time, lower precision, lower accuracy, and higher difficulty with the ID increase observed in AR. The results suggest that for rapid-aimed movements in AR, target sizes should be larger than 0.019

meters if the accuracy and efficiency in AR are desired at the level seen with the touchscreen. For interface push buttons with fingertip activation, NASA specifies a minimum diameter of 0.010 meters when barehanded and 0.019 meters when gloved [107]. Based on our results, AR interactions could be achieved under NASA's gloved thresholds, which align with our more difficult index of difficulty. However, we would anticipate greater errors and less accuracy than physical button hardware as a holographic button size of 0.019 meters degraded performance on all dependent measures. In order to achieve the same level of performance as with TS, we would recommend a larger button size than that explored in this study if accuracy and speed are desired. In general, AR interfaces should have wider targets and shorter movement distances than physical buttons, which corresponds to lowering the index of difficulty. Our levels of selected difficulty would still be appropriate for measuring user abilities in applications relating to sensorimotor performance where higher difficulty is desired to evaluate hand-eye coordination.

A previous AR usability study [148] observed that participant performance in AR was influenced by (1) their ability to perceive hologram locations correctly, (2) their ability to maintain fingertip visibility to support the tracking system, and (3) the speed at which the task was performed, which may have been a constraint to the tracking systems capabilities. The study observed a user perceptual offset error with a subset of users, where perceptions of hologram locations fell short of the actual hologram locations, thereby limiting the ability to acquire targets. This depth perception issue is consistent with the higher error rates observed in AR in the present study as in some cases, multiple taps per target were required to generate a successful target acquisition in AR. For interface design considerations, improved surface illumination and other visual or audio cueing may support depth perception. In AR, there is no tactile feedback that provides the user with depth information, and therefore, the task becomes more difficult when compared to touchscreen devices. Additionally, the headset's tracking system was observed both in the previous and current study to be constrained by the user's hand posture and gesture execution. If the user held their index finger parallel to the ground, the back of the user's hand occluded the fingertip resulting in tracking inconsistencies (Figure 2.3). While the participants in this study were instructed to hold their fingers vertically rather than perpendicular to the target, their hand postures may not have always supported the headset's tracking abilities. As noted both anecdotally throughout testing and within the post-test surveys, the magnitude of arm fatigue was perceived as higher while performing the task in AR. Participants were permitted to take breaks between each block of target selections; however, there was no order effect on error rates indicating that fatigue throughout testing did not significantly increase successive taps per target. Designers may consider visual cueing to remind users of hand postures or indicate when the fingertip is occluded. Finally, the Microsoft HoloLens 2 relies on the articulated hand tracking capabilities that function at the highest frame rate of 45 fps for near-depth sensing [152]. Discrete user motions may have been executed too fast for near-depth sensing and the tracking algorithms despite an increase in average movement times within AR compared to TS. Anecdotal accounts report the white fingertip pointer (Figure 2.3) lagging behind the fingertip at increased speeds. High movement speeds and poor user depth perception combined with imperfect hand gesture posture may have limited overall user performance as measured by the dependent variables.

4.3.1 Future Work and Limitations

The findings of this study are limited to two devices, the HoloLens 2, and a 23.8-inch touchscreen monitor. Characterizing performance on additional devices is warranted to support design decisions. While a training period was provided and deemed sufficient for the testing protocol, users may still improve performance with extended use of AR, which may influence results. As noted, both anecdotally throughout testing and within the post-test survey, the magnitude of arm fatigue was perceived as higher within AR. Although, participants were allowed to take breaks between blocks and sequences, most opted not to, and there was no order effect on error rates, indicating that fatigue did not significantly increase errors throughout testing. Several studies have documented arm fatigue during AR [153, 154] and VR [155] interactions that require prolonged arm postures, with a recent investigation revealing an increase in error rates and perceptions of physical effort and arm fatigue when aiming overhead compared to shoulder height during target acquisition in VR [156]. Future work should consider the added fatigue of using AR for user interface interactions and similar investigations may consider presenting the target array at an angle below shoulder height. While this study had an initial sample of younger adults, further exploration into older or sensorimotor-impaired populations is warranted to validate the task as a potential sensorimotor assessment tool and to uncover potential differences across populations. The lack of tactile feedback in the AR environment may have impacted performance. In addition to improved design recommendations discussed for user perception, superimposing the holographic array on a flat physical surface may improve this aspect of interaction. Even though performance varies from other modalities, the AR tool could be used as an assessment method to evaluate trends in performance. Other use cases may include the evaluation of user performance for the same target acquisition task under vestibular disorientation as an analog for sensorimotor impairment.

Future investigations may include alternative assessment tasks such as random target acquisition and the extension of the interaction space into three dimensions similar to recent research with virtual reality systems [157]. As an alternative to measuring performance, Fitts' law models have been extended to improve predictive power for complex 3D reaching tasks through the addition of target angle with respect to the center of the interaction space and target depth [158, 157, 159]. The current study task was more challenging when using AR than a touchscreen for the same multidirectional tapping task configuration, suggesting that the effective index of difficulty differs and could provide information for developing new predictive models in 3D. Although the targets were presented within a two-dimensional format, the increased complexity required to perform movements in three dimensions introduces additional variability in performance that is not accounted for in traditional Fitts' law predictive models. Ultimately, further research is required to extend these predictive models to account for domain-specific attributes in extended reality systems.

4.4 Conclusion

The research evaluated user performance on a 2D target acquisition task under two display modalities, augmented reality and touchscreen, and two indexes of difficulty. Dependent measures for performance included accuracy, precision, error rates, movement time, and throughput. As hypothesized, user performance declined as the task difficulty increased. The touchscreen yielded better performance on all dependent measures. Results suggest that although performance was limited within augmented reality, the environment may still be a viable tool for the evaluation of hand-eye coordination with the benefit of hand and eye tracking capabilities. From an interface design perspective, designers should consider target size, movement distance, gesture interactions, and depth perception cues during the design process for AR applications in context with the task needs for accuracy and movement times.

The current chapter evaluated the use of AR for hand-eye coordination assessments by determining the impact of holographic targets on user performance to inform the viability of AR as a sensorimotor standalone assessment tool. The following Chapters, 5 and 6, evaluate the performance of healthy adults on holographically-administered balance and operational assessments to address the feasibility of using AR as a standalone assessment tool for vestibular assessments.

CHAPTER 5

Evaluation of Augmented Reality Neurovestibular Dynamic Balance Assessments

In the following chapter, AURORA's dynamic balance assessments are compared to traditional physical implementations used in clinical care to inform research questions one and three pertaining to the use of AR as a standalone tool for vestibular assessments by quantifying the impact of holographic objects on task performance.

In this study, the performance of healthy adults on two dynamic balance tasks was evaluated within an AR environment and while using traditional physical objects. The selected assessments include the Four Square Step Test and the Tandem Walk task. The research considers how virtually administered holographic balance tasks impact users' postural control and task completion strategies. We hypothesized that as a result of the visual field limitations posed by the AR headset, there would be differences in (1) task completion time, (2) step characteristics, and (3) head and torso kinematics. This study has important implications for future designers of AR applications, particularly for postural control and balance assessments, as it prompts designers to consider the presentation and impact of holographic content on user performance. To ensure users' safety, it is necessary to also know if AR use affects one's balance to a concerning degree. Finally, these results inform the interpretation of user performance within AR as compared to traditional clinical-based assessments of the Four Square Step Test and Tandem Walk tasks.

5.1 Methods

5.1.1 Participants

Twenty University of Michigan students participated in this experiment (11 female and 9 male, mean age 25.20 years ± 8.7 SD, min: 18, max: 60). All participants reported normal ($n = 14$) to corrected-to-normal vision ($n = 6$) and reported no musculoskeletal, auditory, or vestibular disorders. The participants were informed of the procedures before participating and signed the

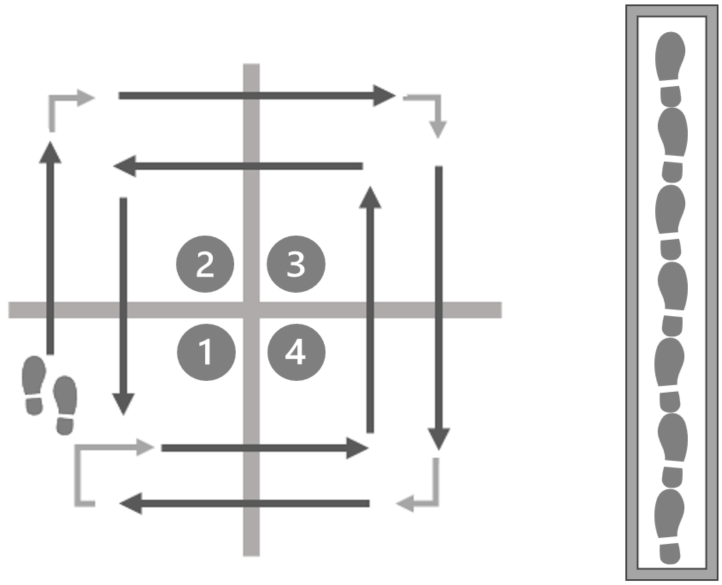

Figure 5.1: Left: Four Square Step Test. Right: Tandem walk task with hollowed-out beam for augmented reality implementation.

informed consent approved by the University of Michigan Institutional Review Board for Health Sciences and Behavioral Sciences. Fifteen participants reported previous limited experience (between 1 and 10 hours) with extended reality devices (i.e., augmented reality or virtual reality); the remaining participants were considered novice users.

5.1.2 Protocol

This study is part of a larger protocol that evaluated a suite of dynamic assessments and functional tasks designed to be conducted within a 2x2 meters space based on requirements for use in a space capsule. This paper presents the Four Square Step Test (FSST) and Tandem Walk (TW). The participants performed an hour-and-a-half protocol consisting of AR system calibration and training, task familiarization, evaluative testing, and survey responses. The participants donned eleven inertial measurement units and 36 motion capture markers at specific anatomical locations (Figure 5.2). For this research, the motion capture marker set did not include heel markers to prevent interference with proper heel-to-toe contact during the TW. The Microsoft HoloLens 2 device, which weighs roughly 1.25 lb [160], was worn for both the AR and physical conditions to maintain consistent weight distribution on the head across the modalities. The AR visual display was active during the AR trials and inactive for the physical trials. Users could see their physical environment whether the device was active or inactive due to the see-through holographic lenses.

The HoloLens 2 utilizes onboard eye-tracking capabilities to automatically compute eye posi-

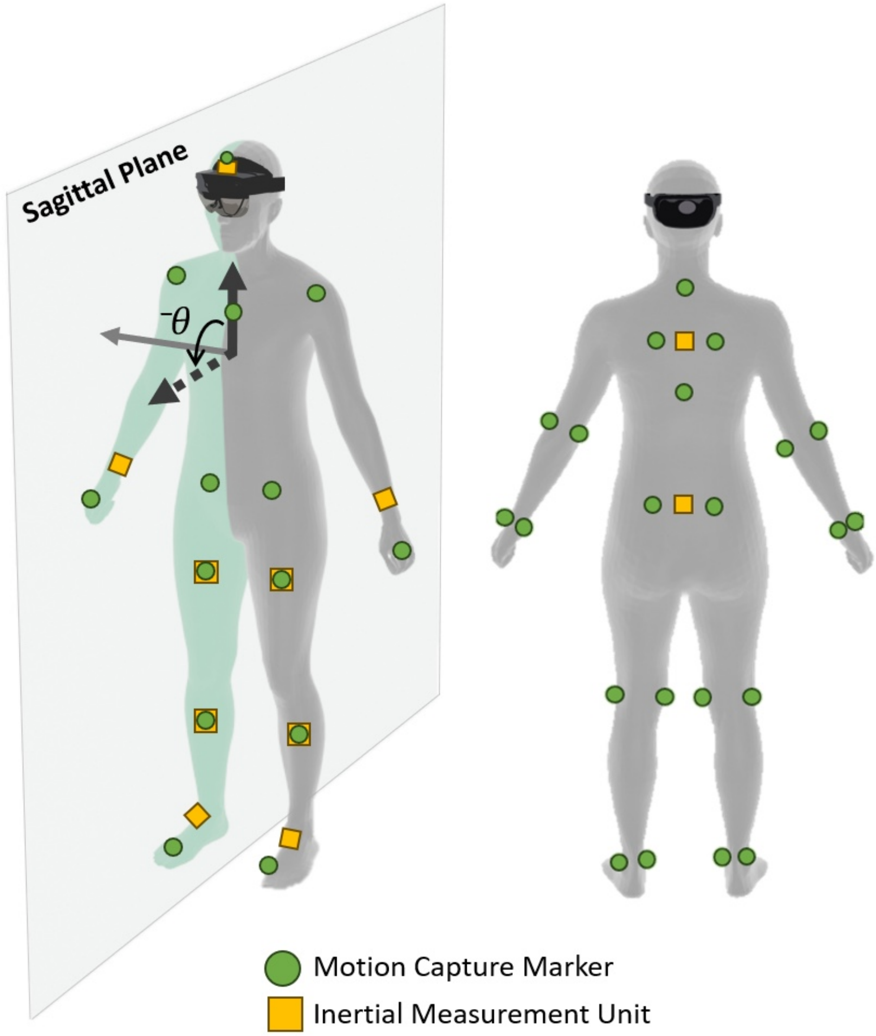

Figure 5.2: Body-Worn sensor configuration comprised motion capture markers, inertial measurement sensors, and the Microsoft HoloLens 2. Depiction of torso pitch, as estimated along the sagittal plane. Head and torso pitch angles were calculated relative to a neutral standing posture with a right-handed ISB coordinate system. Positive pitch corresponds to leaning back.

tions to enable accurate hologram positioning, improved display quality, and comfortable viewing experiences across users [152]. Following the eye calibration, each participant was trained on the hand gestures required to operate and interact with the system (e.g., touch selection with the index finger, a hand ray for distance manipulation, and a pinch gesture to manipulate holograms [127]. Task familiarization was performed for each assessment and within each modality. The task instructions were consistent across modalities. Participants were permitted to practice each task twice before proceeding to the testing trials of the task.

The participants performed the FSST followed by the TW for each modality, irrespective of their modality order. Three trials were conducted for each task and modality, totaling six trials per task. Following the completion of three trials per modality for a single task, a post-task survey was administered to obtain user perceptions of workload and fatigue. The post-task survey incorporated questions from the NASA Task Load Index [143] and a question regarding the user's preferred modality to perform the task.

5.1.3 Vestibular Assessments

In this analysis, users' balance performance was evaluated for the FSST and the TW. Participants performed the assessments within an AR environment and using a traditional clinical-based methodology with physical objects. Each participant was randomly assigned one of two treatment groups that specified the order in which they performed the assessment tasks with respect to modality (AR holographic objects vs. physical objects).

The TW task required participants to walk heel-to-toe along a linear beam-like object that was 2 meters in length. During the physical condition, the task axis was indicated by masking tape placed on the ground, while in the AR environment, a holographic beam was displayed. The HoloLens 2 does not feature foot tracking capabilities, and therefore, the feet could be occluded by the holographic content placed on the floor. To mitigate occlusions, the TW beam was hollowed out to present a skeleton-like structure where the feet can be placed with minimal hologram obstruction (Figure 5.1). Each participant was instructed to align their left foot's heel with the object's edge at the start of the task. As the participants moved along the beam, they were instructed to maintain heel-to-toe contact for each step. While participants were allowed to look down when necessary, their hands were required to remain on their hips during the task.

The FSST features a cross-shaped object placed on the floor (Figure 5.1). Following a predefined stepping pattern, participants stepped over the axes of the floor object in all directions. While facing forward, the participant performs the pattern clockwise and subsequently counterclockwise (Figure 5.1). For the physical condition, the quadrants were constructed from masking tape placed on the floor. The AR environment featured a holographic cross-shaped object denoting the quadrants. With this particular task, the emphasis was not on accuracy within quadrants, but rather on completing the task quickly and safely. The participants were instructed to keep their heads neutral facing forward, and to limit looking down at the floor.

5.1.4 Performance Measures

Using the literature and clinical practice, dependent measures were derived from the inertial measurement unit (IMU) sensors and the motion capture data. Using IMUs, postural sway and orientation metrics were assessed, and motion capture was used for task-specific metrics requiring accurate position estimations, such as foot placement relative to task objects. While motion capture can provide metrics of postural sway, IMUs offer advantages over motion capture as they do not require a laboratory setting. The study included both methods due to the wide availability and use of IMUs outside the laboratory. The primary dependent measure was task completion time. Additional dependent measures were extracted from the IMUs worn on the head and the torso to evaluate postural sway and user strategies for task completion. The IMU-based measures extracted from the raw accelerometer and gyroscope signals were based on the L2-norm of the signals, defined at each time point k as:

$$||\mathbf{S}_k|| = \sqrt{{S_{x,k}}^2 + {S_{y,k}}^2 + {S_{z,k}}^2} \tag{5.1}$$

where S_x, S_y, and S_z are the signal values with respect to the X, Y, and Z axes. From the L2-norm, metrics included the root mean square of the norm vectors and the magnitude (i.e., the absolute value of the peak norm vector) across the entire time series. These dependent measures have been widely implemented across the literature as principal parameters for measuring standing postural balance through IMU sensors [112].

Sensor-to-segment anatomical alignment and orientation decomposition were conducted on the IMU measures to estimate the orientation of the head and torso in the sagittal, coronal, and transverse planes (Figure 5.2). In this algorithm, pitch, roll, and yaw were calculated relative to a neutral standing posture. The orientation of the inertial measurement unit was first estimated using an extended Kalman filter, which serves as a sensor fusion algorithm for the accelerations and angular velocity signals [161, 162]. The remaining algorithmic processes include defining the local task frame given by the IMU sensors, defining the anatomical frames, projecting the relevant global vectors into the anatomical planes [163], and computing the orientation angles and the sign of the angles using the right-handed ISB coordinate system [164]. Pitch, as estimated along the sagittal plane, was the primary metric of interest for the FSST and TW. Roll was defined as the side-to-side lean in the coronal plane and yaw as the side-to-side twist in the transverse plane. A positive pitch occurs when the participant leans back (nose up), a positive roll occurs when the participant leans to the right, and a positive yaw occurs when the participant twists to the left.

Task-specific analyses that align with task goals were conducted using motion capture data to provide accurate position estimations with respect to task objects. Lower-limb metrics were extracted from markers located on the feet, i.e., the toe markers and medial and lateral malleoli. Dependent measures were averaged within the trial and finally across the three trials for each modality to yield a single measure of performance per participant for the AR and physical conditions.

For the FSST, foot clearance (i.e., step height) was calculated and averaged across the feet for each trial. Foot clearance is seldom reported within the literature; however, this metric may inform different user strategies for task completion with holographic and physical objects. In

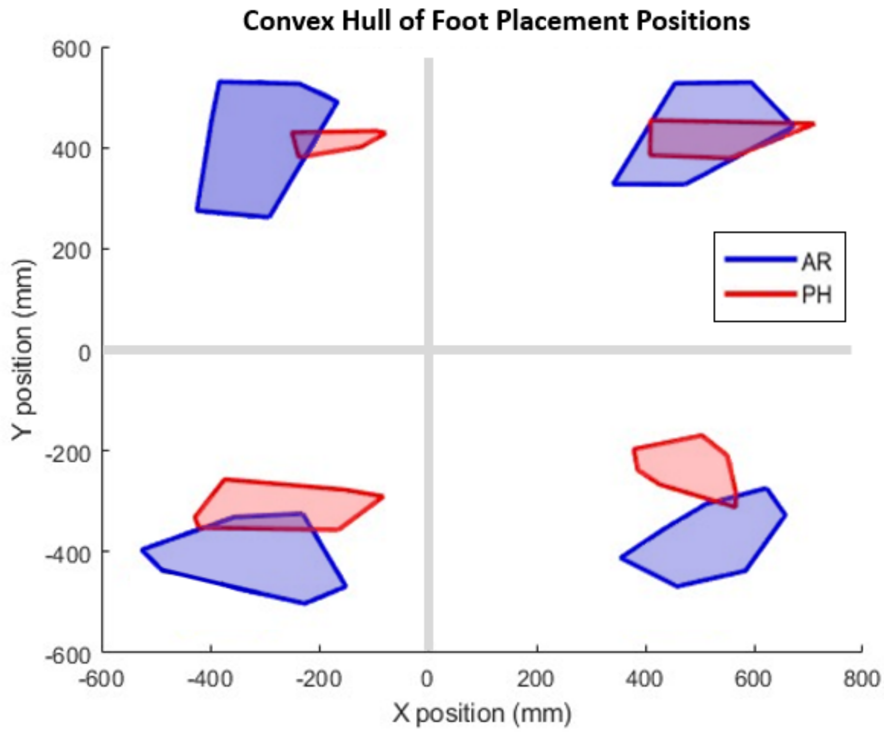

Figure 5.3: Convex hull of the left and right foot placements as estimated by the motion capture toe markers within each quadrant, averaged across modality and trials. The image is an example of the compiled data for a single participant.

line with other recent literature, temporal decomposition analysis was implemented to identify the portion of the total task time transitioning between quadrants and time spent within a double stance [165, 166]. Further temporal decomposition was conducted to evaluate the total time spent transitioning to each task axis (i.e., forward, backward, right, and left), normalized by total task time. Although foot placement accuracy was not emphasized in the task goals and instructions, the variability in stepping positions within each quadrant was calculated to inform user strategies between modalities. During instances of double stance foot placement within each quadrant, the convex hull of the end-point positions of the toe markers was measured (Figure 5.3). The average convex hull for each square and the grand average for the task was quantified.

For the TW, the primary dependent measure used within the literature is the number of correct consecutive tandem steps as defined by appropriate heel-to-toe contact [21]. It was observed that all participants performed consecutive tandem steps throughout all trials without loss of balance, extra footfalls, or excessive space between the feet. This analysis quantified the correct number of steps through a surrogate measure, step length distance. Step length distance was measured from the motion capture markers of each foot at each successive step. The distance between the two toe markers of each foot at each step was measured and averaged across trials. The average and standard deviation of step lengths were calculated to determine the degree to which the participant performed the task with consistent heel-to-toe contact. Based on the assumption that the feet were aligned relatively straight along the task axis, larger step lengths may indicate improper heel-to-toe contact. Similarly, step height and step height variability were calculated from the toe markers and averaged across feet and trials. Increased step height variability may suggest more instability during tandem walk cycles. In addition to the standard measures of task performance, we explored behavioral strategies between the two modalities in the analysis of the TW task. The additional measure included the foot clearance, i.e., step height. In previous literature, foot clearance has been calculated from the heel displacement [167]; however, the toe motion capture markers were used for this research to determine foot clearance during the tandem walking cycles.

5.1.5 Statistical Analysis

Performance measures were assessed using paired t-tests to compare group means for the AR and Physical modalities. Paired t-tests were also performed to compare the survey results between the AR and Physical modalities. For these pairwise comparisons, the level of significance was $\alpha = 0.05$. Data from one participant was excluded in the analysis of the FSST measures that relied on motion capture markers from the foot (i.e., step height, step height variability, and the convex hull of the foot placements) as one toe marker was occluded for the majority of the FSST task. The effect size was calculated using Cohen's method for paired samples.

5.2 Results

5.2.1 Four Square Step Test

The FSST yielded a significant difference in the total task time between AR ($M = 12.04$ s, $SD = 1.64$) and Physical conditions ($M = 10.92$ s, $SD = 1.42$) ($t(18) = -4.35$, $p < 0.001$, $d = 0.70$) (Figure 5.4). Significant differences were found in the decomposition of total task time spent in double stance (AR: $M = 0.15$ s, $SD = 0.09$; PH: $M = 0.11$ s, $SD = 0.06$; $t(18) = -2.36$, $p = 0.029$, $d = 0.47$), spent transitioning quadrants (AR: $M = 0.85$ s, $SD = 0.09$; PH: $M = 0.89$ s, $SD = 0.06$; $t(18) = 2.36$, $p = 0.029$, $d = 0.47$), and spent while transitioning to the right task axis (AR: ($M = 0.12$ s, $SD = 0..03$; PH: $M = 0.11$ s, $SD = 0.02$; $t(18) = -2.67$, $p = 0.016$, $d = 0.57$) (Appendix D). The convex hull of the foot placement within the four task quadrants was larger within AR ($M = 3315.80$ mmˆ 2, $SD = 2445.88$) than within the Physical condition ($M = 1926.23$ mmˆ 2, $SD = 1333.08$) ($t(18) = -2.20$, $p = 0.041$, $d = 0.68$) (Figure 5.4). No significant differences were found between modalities for the duration of single

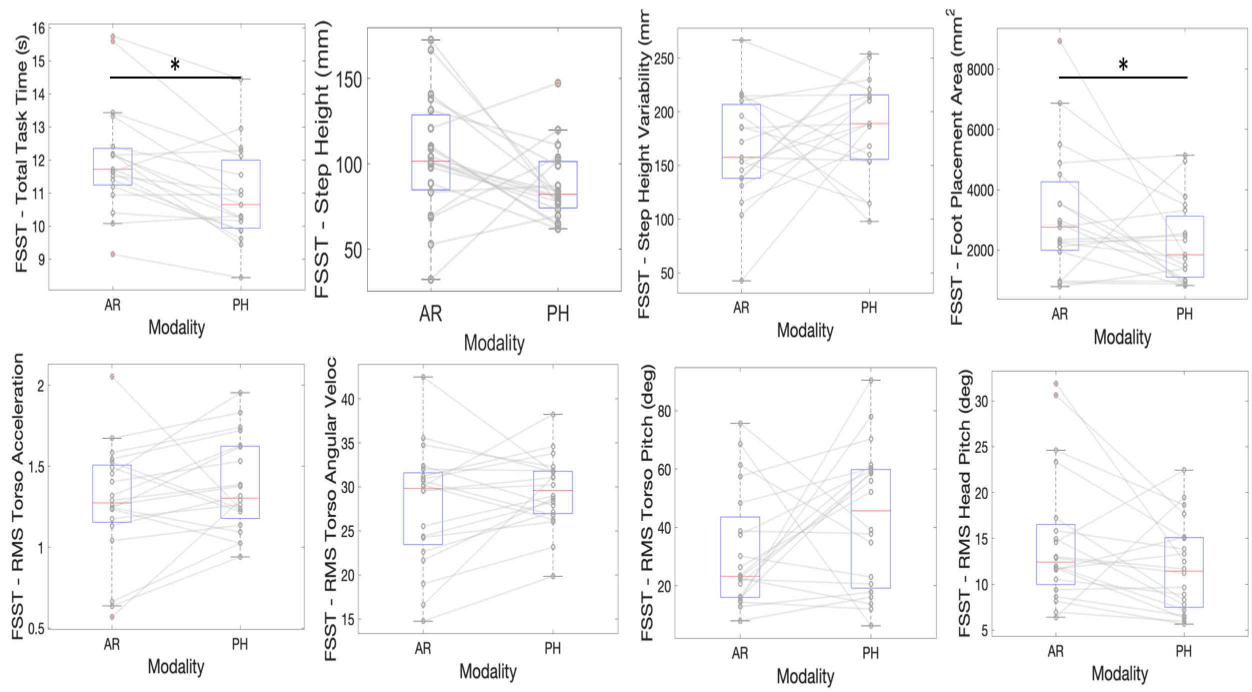

Figure 5.4: Four Square Step Test dependent measures: total task time (p = 0.009), step height, step height variability, convex hull of foot placement (p = 0.019), RMS torso acceleration, RMS torso angular velocity, RMS head pitch, and RMS torso pitch. Significant differences are indicated by the black line and asterisk.

leg stance periods, the transitions forward, backward, or leftward, average step height, and step height variability (Appendix D).

No significant differences were identified between the modalities as measured by the IMU accelerations and angular velocities for the head and the torso. Additionally, no significant differences were discovered for the head downward pitch in the AR condition (M = 14.80 deg, SD = 7.35) than in the Physical condition (M = 11.86 deg, SD = 4.99; $t(19) = -1.88$, p = 0.08, d = 0.45) or the torso pitch for AR (M = 31.95 deg, SD = 20.07) and Physical (M = 43.37 deg, SD = 24.50; $t(19) = 1.56$, $p = 0.13$, $d = 0.49$) (Figure 5.4).

5.2.2 Tandem Walk

The TW task yielded a significant difference in the total task time between AR (M = 11.12 s, SD = 4.75) and Physical (M = 9.07 s, SD = 3.22) ($t(19) = -2.88$, $p = 0.016$, $d = 0.48$) (Figure 5.6). In the TW, average step height exhibited a higher clearance within AR (M = 180.98 mm, SD = 35.00) than within Physical (M = 165.24 mm, SD = 25.59) ($t(18) = -4.28$, $p < 0.001$, $d = 0.49$) (Figure 5.6). The step height variability, average step length, and step length variability exhibited no significant differences between modalities (Appendix D).

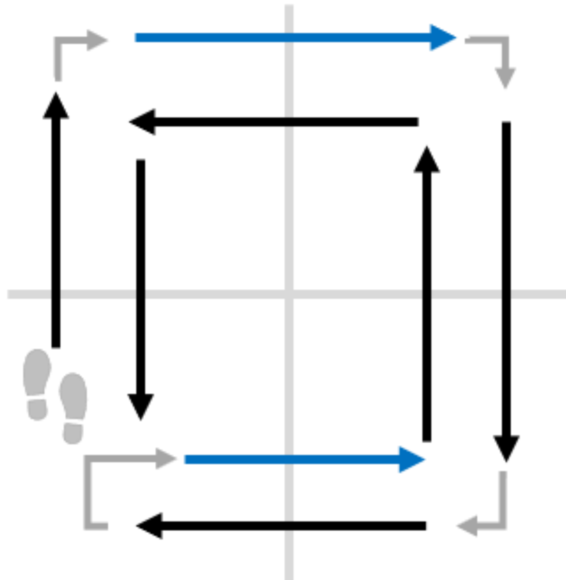

Figure 5.5: The image depicts the rightward transitions within the FSST stepping sequence. The modalities were significantly different from each other for the rightward step duration (p = 0.016, d = 0.57).

For the tandem walk, there were small significant differences between AR and Physical Modalities in the RMS of the torso acceleration signals (AR: $M = 0.42$ m/s$\hat{2}$, $SD = 0.11$; PH: $M = 0.45$ m/s$\hat{2}$, $SD = 0.12$; $t(19) = 2.30$, $p = 0.033$, $d = 0.27$), the RMS of the torso angular velocity signals (AR: $M = 15.81$ deg/s, $SD = 3.92$; PH: $M = 17.51$ deg/s, $SD = 5.11$; $t(19) = 3.53$, $p = 0.002$, $d = 0.36$) (Figure 5.6), the amplitude of the torso acceleration (AR: $M = 1.75$ m/s$\hat{2}$, $SD = 0.53$; PH: $M = 2.03$ m/s$\hat{2}$, $SD = 0.82$; $t(19) = 2.61$, $p = 0.017$, $d = 0.39$), and the amplitude of the torso angular velocity (AR: $M = 0.67$ deg/s, $SD = 0.19$; PH: $M = 0.74$ deg/s, $SD = 0.25$; $t(19) = 3.11$, $p = 0.006$, $d = 0.29$) (Appendix D). No significant differences were identified for the root mean square and amplitude of the head accelerations and angular velocities or the head downward pitch (AR: $M = 43.21$ deg, $SD = 17.15$; PH: $M = 36.36$ deg, $SD = 11.26$; $t(19) = -1.74$, $p = 0.098$, $d = 0.45$, and the torso pitch (AR: $M = 34.57$ deg, $SD = 21.40$; PH: $M = 34.04$ deg, $SD = 22.04$; $t(19) = -0.35$, $p = 0.73$, $d = 0.02$) (Appendix D).

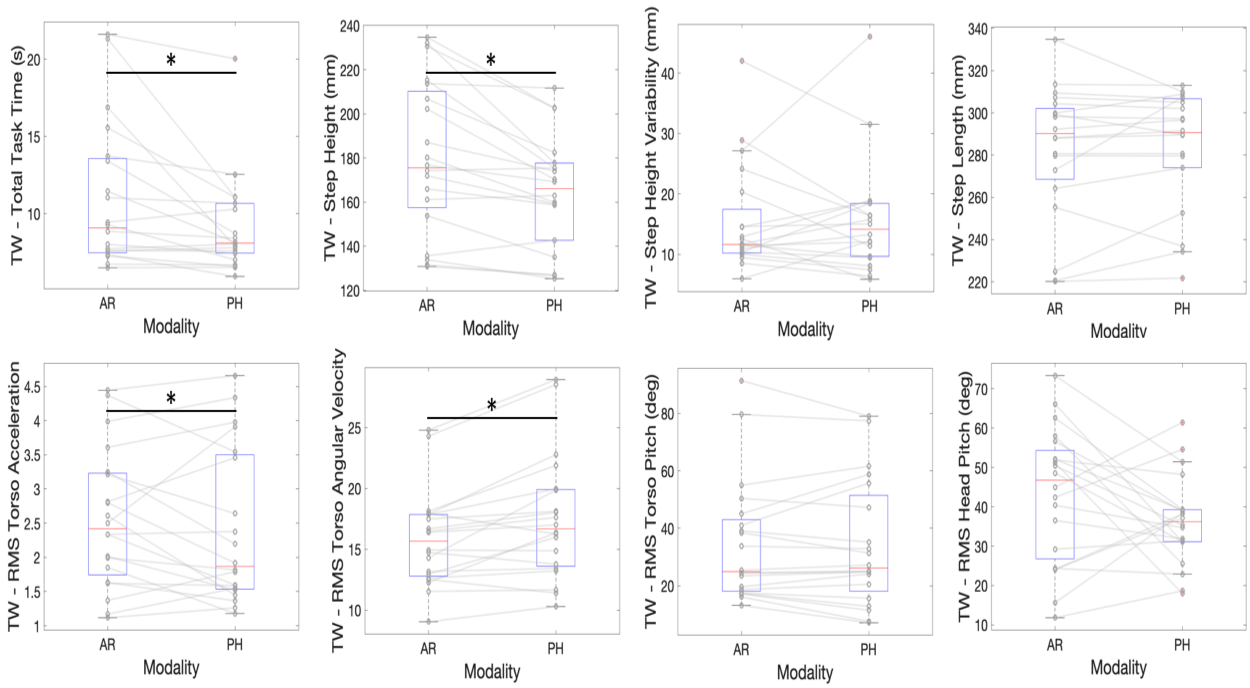

Figure 5.6: Tandem Walk dependent measures: total task time (p = 0.016, d = 0.48), step height ($p < 0.001$, d = 0.49), step height variability, step length, RMS torso acceleration (p = 0.033, d = 0.27), RMS torso angular velocity (p = 0.002, d = 0.36), RMS head pitch, and RMS torso pitch. Significant differences are indicated by the black line and asterisk.

5.2.3 Survey Result

For the FSST, there was a significant difference in perceived mental demand between AR ($M = 15.84$, $SD = 10.06$) and Physical ($M = 8.89$, $SD = 13.62$) ($t(19) = -2.17$, $p = 0.04$, $d = 0.56$) (Figure 5.7). There was also a significant difference in perceived task success between AR ($M = 79.74$, $SD = 21.62$) and Physical ($M = 91.58$, $SD = 10.07$) ($t(19) = 2.74$, $p = 0.01$, $d = 0.67$) (Figure 5.7). There were no significant differences in perceived physical demand, effort, or frustration between the two modalities for the FSST (Appendix D). There was no significant difference in perceived physical demand, mental demand, success, effort, or frustration between the two modalities for the TW (Appendix D). For modality preferences, all twenty participants preferred the Physical condition for the FSST. For the TW, thirteen participants preferred the Physical condition, two participants preferred AR, and the remaining five participants had no preference.

5.3 Discussion

In this study, twenty healthy adults were evaluated on two dynamic balance tasks, the Four Square Step Test and Tandem Walk, using AR and traditional physical objects. The research

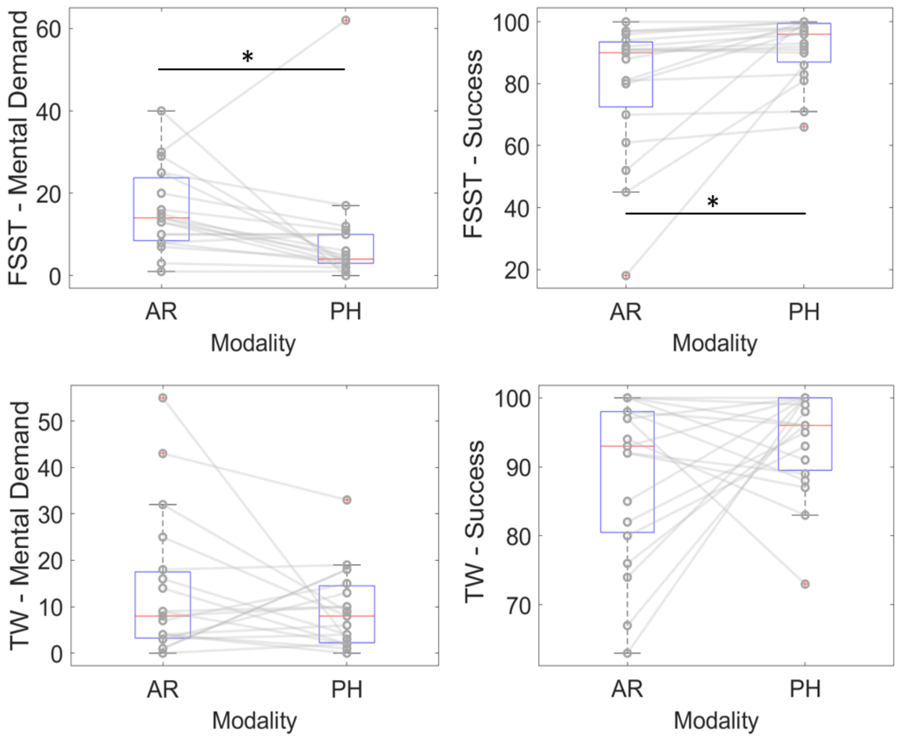

Figure 5.7: Survey results for perceived mental demand and success. For the FSST, the modalities were significantly different from each other (p = 0.04, d = 0.56; p = 0.01, d = 0.67). Significant differences are indicated by the black line and black asterisk..

explored how virtually administered holographic balance tasks impact users' balance control and task completion strategies. We hypothesized that as a result of the visual field limitations posed by the AR headset, there would be differences in (1) task completion time, (2) step characteristics, and (3) head and torso kinematics.

The results support our first hypothesis that the participants would exhibit differences in task completion within both dynamic balance tasks. The total task completion times were found to be significantly different across modalities. For both assessment tasks, the AR environment resulted in longer completion times. In the FSST, participants using AR (mean 12.04s) and the traditional physical condition (mean 10.92s) had slower completion times than normative data for healthy young adults 26 years of age (mean performance time 6.91s $\pm$ 0.49s) [168]. According to the literature, numerous cutoff scores have been identified for increased risk of falls in various populations such as older adults ($>$ 15 seconds) [121], individuals with vestibular disorders ($>$ 12 seconds) [120], amputees ($>$ 24 seconds) [169], and stroke patients ($>$ 15 seconds) [170]. Although the participant's average age within this study varies with respect to the previous literature, the majority of the healthy participants would be correctly identified as non-fallers within both conditions with the given cutoff thresholds. With the cutoff thresholds for older adults and stroke patients, two participants would be deemed at risk of future falls because they performed over 15 seconds within the AR condition. While using AR, five participants shifted from no risk to at-risk with the cutoff for vestibular disorders ($>$ 12 seconds). One participant's performance in the physical condition shifted from no risk to at-risk with the 12-second cutoff. Four participants performed above 12 seconds in both modalities. Even though participants were instructed to perform the FSST as fast and safely as possible, the results suggest they may have been capable of performing the task quicker across both modalities. Based on the healthy participants' performance in the current study, it will be necessary to conduct further research to determine whether the cut-off thresholds should be modified when AR is used to determine fall risk.

Tandem walk total task time cannot be compared with literature due to differences in protocols and walking distances; however, other connections can be made concerning cutoff scores. In clinical care, TW cutoff values discriminating between fallers and non-fallers have been evaluated using the number of missteps [171] and composite measures summing the task time and number of incorrect steps [172]. [171] identified a cutoff score of 2 missteps out of 10 total correct step attempts. [172] reported composite cutoff values of 11.6 and 13.7 to discriminate between no, moderate, and severe mobility restrictions. In this study, the walk distance was held constant, which meant the number of steps varied across participants, making comparisons with the composite cutoff not feasible. In addition to extra footfalls, missteps for astronauts are characterized by step durations longer than three seconds and step gaps from heel to toe larger than ten centimeters [1]. All participants performed tandem steps quicker than 3 seconds within both modalities. Step gaps larger than ten centimeters, as measured from heel to toe, were not directly evaluated with the protocol, but could be determined in the future using step lengths with measurements of individual foot lengths. Future work may consider these additional TW measures for comparison across AR and Physical modalities.

The motion capture analysis supported our second hypothesis that participants' step characteristics would differ between modalities. In the FSST, participants exhibited greater variability in foot placement in AR, longer stepping duration when transitioning in the right direction, and longer total stance duration in AR than in the Physical condition. Within the FSST stepping sequence, the rightward step is the second and fifth transition in the sequence. The fifth transition from the bottom left quadrant back to the bottom right quadrant is a unique dynamic movement within the sequence. This transition marks the progression from clockwise rotation to counterclockwise rotation. As a result, the rightward translation is expected to take longer regardless of modality because momentum must be overcome to return from the direction previously translated. According to [165], this section of the FSST was particularly difficult for those with a left amputation but could be improved with consecutive tests. The post-task survey indicated that participants were less confident in their FSST performance while using AR, which aligns with the increased variability observed in foot placement and increased rightward translation time. Although precise foot placement was not emphasized as a task goal, the increased variability in foot placement suggests the influence of limited perceptual abilities in AR. The difference in foot placement was not driven by instability as there were no significant differences found between the modalities for the head and torso acceleration and angular velocity signals. The survey results also exhibited higher magnitudes of perceived mental demand for the AR condition, supporting the different perceptual and cognitive processing to complete the task in AR. Although the HoloLens currently has one of the widest fields of visual display on the market, it is still significantly constrained compared to human vision [113]. In this case, the physical condition afforded peripheral vision where users could see the tape on the floor even with their heads held neutral and facing forward, resulting in a better mental spatial mapping of the interaction space, which decreased foot placement variability. In contrast, the holograms were more difficult to view with peripheral vision in AR since the head must be pointed at the object to fully render. When performing the task within AR, users felt less confident because they had to rely on their mental model mapping of the quadrant locations without peripheral input. The lack of confidence within AR is consistent with all participants preferring the physical condition. A potential follow-on study could leverage the visual cueing capability of AR and provide the user with a top-down view of the interaction space and an icon that traverses in accordance with user movement to facilitate the user's spatial mapping. While visual cues may support peripheral vision, it is important to note that this would modify the task from the physical traditional condition.

The presence of holograms influenced the participants to step higher in TW. It is hypothesized that these step height characteristics are a result of the interaction between one's lower limbs and the holograms, particularly within the TW task where users were permitted to look at their feet. According to the results, participants pitched their heads more during the TW task (43.21 and 36.36 degrees) than during the FSST (14.8 and 11.86 degrees). While the TW object was hollowed out (Figure 5.1) to mitigate this aspect of AR interactions, there may still have been occlusions at the edges, which caused participants to instinctively step higher to clear the perceived obstacle or to enable improved viewing of the object. Even though physical collisions with holograms were not feasible, participants were more conservative in stepping over the holographic content, thus supporting their perception of the holographic objects in relation to themselves and the surrounding environment. Assessments that require participants to step higher may be of interest in clinical settings for increased task difficulty or training to encourage higher step heights for distinct populations. Hennah and Doumas [173] found that older adults were less able to adapt their step widths and step heights when walking in dual-task situations than younger adults, which may make them more vulnerable to tripping or falling. For dynamic tasks such as tandem walking, AR may provide a novel engaging experience for assessments and rehabilitation in various populations.

The IMU orientation analysis revealed mixed support for our third hypothesis related to head and torso kinematics. The FSST revealed no differences in torso and head kinematics since participants were instructed to look forward and did not have to restrict torso movement to view the holograms. Small significant differences in body kinematics emerged within the TW due to the AR environment and field-of-view restrictions. Significant differences were found between the accelerometer and gyroscope data for the torso. The results suggest that the participants' overall torso movement was restricted within AR when compared to the physical condition. While higher magnitudes of kinematic measures, such as torso angular velocity, are usually interpreted as more instability when comparing controls to patients (e.g., peripheral neuropathy [174]), in this study the difference in torso kinematics does not reflect more instability. Here users selected strategies with a restriction of torso movement in AR to complete the task as a compensatory action to maintain a stable head position for viewing the holograms. Depending on where the user focuses their attention, only portions of the TW beam could be seen. In contrast, the entirety of the physical tape was perceived. Although the addition of visual cues would make the task less similar to the physical condition, future designers may consider a top-down mini-map view of the user position or larger field-of-view headset capabilities for improved performance. Ultimately, the use of AR did not affect the primary indicator of performance, the correct number of steps, therefore AR could still be used to assess TW performance. The restrictions of participant torso motion within the AR condition to keep their heads more stable is similar to compensatory mechanisms observed in astronauts postflight and vestibular patients, who restrict head movement to mitigate nausea symptoms [3]. Therefore AR would not exacerbate or cause additional stimulating motions to negatively affect TW performance in these populations, and the presence of IMUs could be useful in detecting these adaption strategies.

The use of holographic objects in place of traditional physical objects influenced user performance in both of the dynamic balance tasks. While participants were able to perform both tasks successfully, the strategy by which they completed the tasks differed across modalities, with higher step heights in the TW and higher foot placement variability in the FSST leading to longer task performance times. The few participants who rated the AR environment as the preferred modality for the TW noted that the AR environment was more engaging. The AR-administered assessments were useful for obtaining postural control and task completion measures. Although this study combined motion capture with embedded sensors, a standalone tool for assessing balance could be created using embedded sensors only. For clinicians and researchers interested in using AR to administer postural control assessments, the interpretation of task performance must be handled carefully, especially when comparing performance to non-AR administered assessments.

5.3.1 Future Work and Limitations

This study explored and presented several motion-based and time-based measures; however, future studies should examine these findings further with a larger sample size. While this study had an initial sample of healthy younger adults, a future exploration into older or sensorimotor-impaired populations is warranted to evaluate the influence of AR-based balance assessments on participant performance across diverse population samples. Additional measures of performance could be applied to characterize medial-lateral torso sway through double integration of the IMU accelerations and evaluating the area under the displacement curve [167, 175]. Future studies may also consider leveraging the visual cueing capability of AR to improve perception to support performance on the FSST and TW tasks. Previous work has used dynamic holographic eye gaze objects for static balance assessments to inform task actions, such as instructing head pitching timing [127]. While the results suggest that overall task performance was slower in AR, the evaluation of the effect of practice and repetitive use on task performance is needed. Although this study was part of a larger protocol that evaluated six distinct dynamic assessments and functional tasks, additional evaluations into other validated and widely used assessments in clinical care are warranted. Previous research has evaluated the use of AR for other sensorimotor assessments such as hand-eye coordination tasks as compared to a touch screen device [176]. As AR technology continues to advance and headsets become more affordable, there is a potential to bring AR into in-home care for self-administered assessments or training for vestibular patients, aging populations, athletes, and even astronauts to support clinical decision-making and patient empowerment. Ultimately, it is critical to understand how holographic content influences user strategy to correctly interpret and score user performance for a given assessment task.

5.4 Conclusion

In this research study, holographic objects and augmented reality were evaluated for their influence on participant performance in the Four Square Step Test and Tandem Walk. The performance of twenty healthy adults was compared while using augmented reality and traditional clinical-based physical objects. Although participants were able to perform both tasks successfully, the strategy by which they completed the task differed in AR as measured by inertial measure unit sensors and motion capture software. Overall task completion time was longer for both assessment tasks when administered in AR. For the Four Square Step Test, participants took more time to translate in the rightward direction and exhibited higher foot placement variability within the task quadrants. For the Tandem Walk, participants stepped higher on average in AR. Additionally, in an effort to maintain head stability for viewing the holographic content, participants restricted their torso movement slightly more in the AR environment for the tandem walk. The participants were able to complete both tasks in AR and meaningful measures of postural control were obtained, indicating that AR may be a useful instrumentation solution with embedded sensors to evaluate a variety of populations for postural control.

The current chapter evaluated the impact of holographically administered dynamic balance assessments on user performance to inform the use of AR as a standalone assessment tool that observes spaceflight resource constraints. Chapter 6 assesses the impact of AR on operationally-informed assessments to further inform the first and third research questions relating to the impact of AR and the feasibility of use in spaceflight assessments.

CHAPTER 6

Evaluation of Augmented Reality Neurovestibular Operational Balance Assessments

The following chapter presents the comparative analysis of AURORA's operational balance assessments to physical implementations of the tasks to inform research questions one and three pertaining to the use of AR as a standalone tool for vestibular assessments by quantifying the impact of holographic objects on task performance. To determine the impact of AR on astronauts during sensorimotor impairment, the results from a healthy user population are compared to astronaut performance post-flight on analogous assessment tasks.

In this study, the performance of healthy adults was compared for two dynamic operationally-informed balance tasks between an AR environment and with physical objects. The selected assessments comprised an ingress and egress task and an obstacle weave task. Due to the limited field of view of the headset, we anticipated that there would be differences in (1) task completion times, (2) step characteristics, and (3) head and torso accelerations and angular velocities. This study carries significant implications for AR application designers, especially those focused on postural control and balance assessments. The research underscores the importance of considering how holographic content is presented and how the use of AR may affect user performance. Understanding whether the use of AR has a significant impact on users' balance is essential for ensuring safety and informing decision-making. Furthermore, the findings shed light on the potential utilization of AR as an inflight assessment tool for future exploration-class missions.

6.1 Methods

6.1.1 Participants

Twenty University of Michigan students participated in this experiment (11 female and 9 male, mean age 25.20 years $\pm$ 8.7 SD, min: 18, max: 60). All participants reported normal (n = 14) to corrected-to-normal vision (n = 6) and reported no musculoskeletal, auditory, or vestibular disorders. The participants were informed of the procedures before participating and signed the informed consent approved by the University of Michigan Institutional Review Board for Health Sciences and Behavioral Sciences. Fifteen participants reported previous limited experience (between 1 and 10 hours) with extended reality devices (i.e., augmented reality or virtual reality); the remaining participants were considered first-time users.

6.1.2 Procedure

The study is part of a larger protocol evaluating a set of clinically informed dynamic assessments and operationally-informed functional tasks. To observe potential inflight volumetric restrictions, the tasks were constrained to a 2x2 meter circular volume (Figure 1). In this paper, we present the ingress and egress task (IET) and the obstacle weave (OWT) task. The participants completed an hour-and-a-half protocol, which included calibration and training of the AR system, task familiarization, evaluative testing, and survey responses. Each participant donned eleven IMUs and thirty-six motion capture markers at specific anatomical locations and was randomized into one of two orders in which they performed the tasks (holographic AR objects vs. physical objects first). To ensure consistent weight distribution on the head across the modalities, a Microsoft HoloLens 2 device weighing approximately 1.25 lbs was used across all conditions. With the see-through holographic lenses, participants could see their physical environment regardless of whether the device was active or inactive during the AR and physical trials.

For accurate hologram positioning, higher display quality, and a comfortable viewing experience, participants first underwent eye calibration with the HoloLens 2, which computes eye positions automatically by using onboard eye-tracking capabilities [18]. After the eye calibration, participants were trained for the hand gestures necessary to interact with the system (e.g., touch selection with the index finger, a hand ray to manipulate distance, and a pinch gesture to manipulate holograms [127]). Prior to the testing portion of the protocol, participants were provided with task demonstrations and the ability to practice each assessment task twice within each modality. Instructions for the task were consistent across both conditions. The participants performed the ingress and egress task followed by the obstacle weave, irrespective of their modality order. Three trials were conducted for each task and modality, totaling six trials per task. A post-task survey was administered after participants completed three trials for each modality for a single task. The post-task survey incorporated questions from the NASA Task Load Index [143].

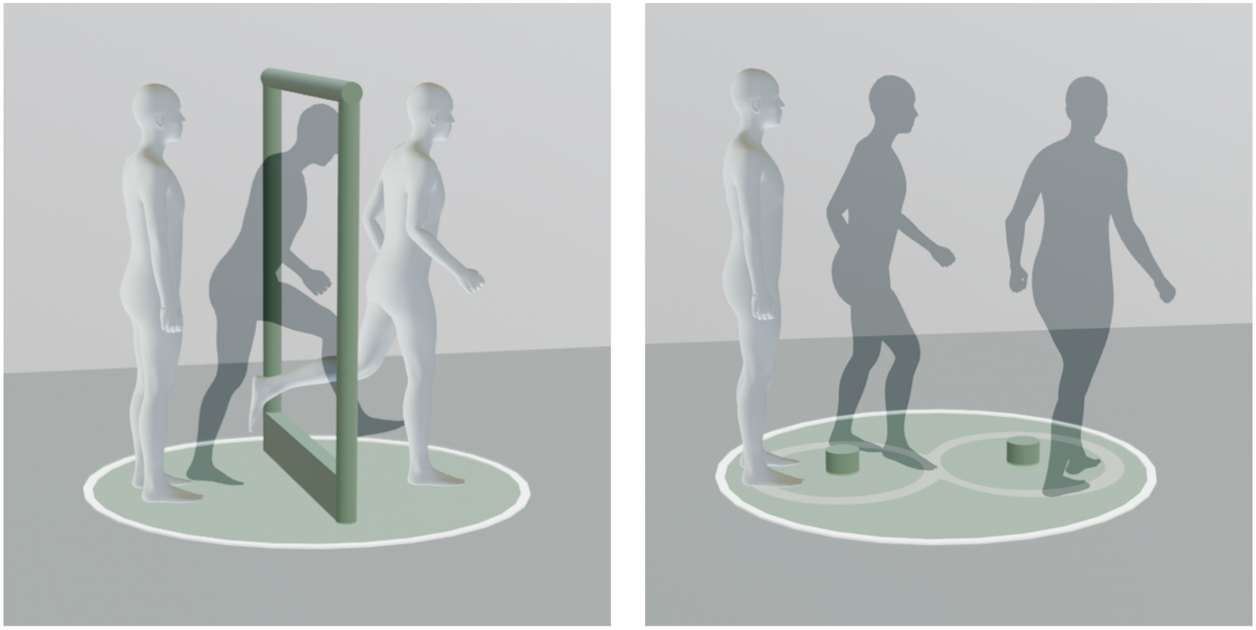

Figure 6.1: Ingress and egress task (left). Obstacle weave task (right). Internal portal dimensions were 1.70 m high by 1 m wide and a step height of 0.2 m. The obstacles were 0.15 m wide, placed 1 m apart, and 0.5 m from the edge of the interaction space.

6.1.3 Operationally Relevant Sensorimotor Tasks

An evaluation of dynamic balance and task performance was conducted for the two operationally-informed assessments, the Ingress and Egress Task (IET) and the Obstacle Weave Task (OWT). The assessments were completed both in an AR environment with holographic content and with physical objects with corresponding dimensions. The operational balance assessments were derived from the FMT task [20] and were chosen due to their operational relevance and their functional characteristics that have proven difficult for crew members after long-duration missions [75, 1]. The intention is that crewmembers will be better prepared for their operational EVA tasks with the additional complexity of spacesuits if they are able to perform these dynamic tasks prior to EVA without an increased concern of falls or functional immobility.

The Ingress and Egress Task simulates an astronaut entering or exiting their spacecraft or airlock upon landing on a planetary body, a potentially difficult task immediately following a gravity transition from extended exposure in microgravity. The dynamic task required participants to ambulate through an airlock portal while minding their posture to allow clearance for their head with the top of the airlock as well as their feet with the step portion of the airlock (Figure 6.1). To evaluate the relative stability of both legs as support legs during the dynamic motion, participants are instructed to step through the airlock, turn 180 degrees, and step through the airlock again with the opposite leg as the leading leg. During transitions through the airlock, participants were permitted to maintain movement in the frontal plane or turn sideways as egress strategies may vary when a crewmember wears a spacesuit.

The Obstacle Weave Task simulates aspects of astronauts ambulating on a planetary surface which will require dynamic directional changes in movement to avoid large geology samples, surface structures, or craters. Participants were required to maneuver around two aligned circular obstacles placed on the floor 1 meter apart (Figure 6.1). The participants were instructed to weave in a figure-eight pattern to allow for both rotations about the right and left side, i.e., clockwise and counterclockwise motions. The starting position was located behind one of the objects, positioned axially along the midplane of the obstacles. After crossing the midplane of the two obstacles at the center of the interaction space, participants walked around the second obstacle before returning to the starting point. The obstacle weave is most akin to the slalom segment of the FMT, where crew members must maneuver around multiple foam columns. Task performance of the obstacle weave task can also be compared to segments of the obstacle walk, another postflight test in which crew members navigate around a cone [1].

6.1.4 Performance Measures

IMU sensors and motion capture data were used to derive dependent measures based on literature and clinical practice for dynamic balance tasks. The use of IMUs enabled the assessment of postural sway and orientation metrics to inform participants' postures while performing the assessment tasks. The motion capture system was used for deriving task-specific metrics requiring precise position estimation, such as collisions with the holographic content. Despite the ability to measure postural sway with motion capture, IMUs are more practical as they do not require a laboratory setting. The study included both methods due to the wide availability and use of IMUs outside the laboratory. In addition, the HoloLens 2 device features an embedded IMU sensor which can be leveraged to enable future standalone assessments.

The primary dependent measure was task completion time as is customary with the FMT task [20]. The IMU-based measures extracted from the raw accelerometer and gyroscope signals were based on the L2-norm of the signals. Using the L2-norm, metrics included the root mean square (RMS) of the norm vectors and the magnitude (i.e., the maximum value of the rectified norm vector). The use of these dependent measures for measuring standing postural balance through IMU sensors is common in the literature [177]. Orientation decomposition and sensor-to-segment alignment were performed on the IMU measurements to estimate the orientation of the head and torso in the sagittal, coronal, and transverse planes. A neutral standing posture was used as a zeroing method for computing alterations in pitch, roll, and yaw across the time series of each task. For the accelerations and angular velocity signals, an extended Kalman filter was used as a sensor fusion algorithm to estimate the orientation of the IMU. The remaining algorithmic process includes defining the local task frame with respect to the IMU sensors, defining the anatomical frames, projecting the relevant global vectors into the anatomical planes [163], and computing the orientation angles using the right-handed ISB coordinate system [164]. The primary metric of interest for the IET was pitch, as measured along the sagittal plane. Roll was defined as the side-to-side lean in the coronal plane. Yaw was considered the side-to-side twist in the transverse plane. A positive pitch was defined when the participant leaned back (nose up), a positive roll occurred when the participant leaned to the right, and a positive yaw occurred when the participant twisted to the left.

Task-specific analyses that aligned with task goals were conducted using motion capture data. Dependent measures were averaged within the trial and then across the three trials for each modality to yield a single measure of performance within AR and physical conditions. Although participants could not feel any physical contact with the holograms, the level of clearance achieved from the holograms was investigated during post-processing. Lower-limb metrics were extracted from markers located on the feet, i.e., the toe markers and medial and lateral malleoli. Upper-body metrics relating to the head were extracted from the motion capture marker located on the top forehead portion of the headset.

For the ingress and egress task, foot stepping heights were captured during each aspect of the task providing measures for the first leading leg to step through the airlock, the first following leg, the second leading leg after the 180 turn, and the second following leg closing out the task. The head clearance was estimated by calculating the distance between the head and the ground during the most prominent head downward pitch under the airlock structure. The head clearance height is interpreted with respect to the height of the airlock within the discussion. Finally, temporal decomposition was conducted to determine the total time to complete each portion of the task, i.e., the translation through the airlock on the first and second attempts as well as the 180-degree turn within the middle of the task.

For the obstacle weave, foot clearance (assessed by step height) and step height variability (assessed by standard deviation) were calculated and averaged across the feet for each trial. The enclosed walking area was measured using a conforming boundary function around the 2D planar positions from the toe markers with a shrink factor of 0.9 providing a compact boundary that envelopes the data. The enclosed walking area characterized the overall space utilized by the participants, while similar analyses were conducted to capture the inner foot boundaries to understand the area used around each obstacle. A larger enclosed walking area indicates further distance from the task obstacles as does a larger inner boundary area.

6.1.5 Statistical Analysis

Performance measures and survey responses were assessed using paired t-tests to compare group means for the AR and Physical modalities. The level of significance for these pairwise comparisons was 0.05. The effect size was calculated using Cohen's method. Data from two participants were excluded in the analysis of the obstacle weave for measures that relied on motion capture markers from the feet (i.e., step height, step height variability, and enclosed walking area) as one toe marker was occluded for most of the task. The inner boundary area measures for the obstacle weave were unaffected as the inner foot marker was consistently present across participants.

6.2 Results

The compiled results and associated p-values for the ingress and egress task and the obstacle weave task are presented in Tables 6.1 and 6.2. Subject-specific data, the median, and interquartile ranges are presented in Figures 6.2, 6.3, and 6.4. Supplemental figures are provided in Appendix E.

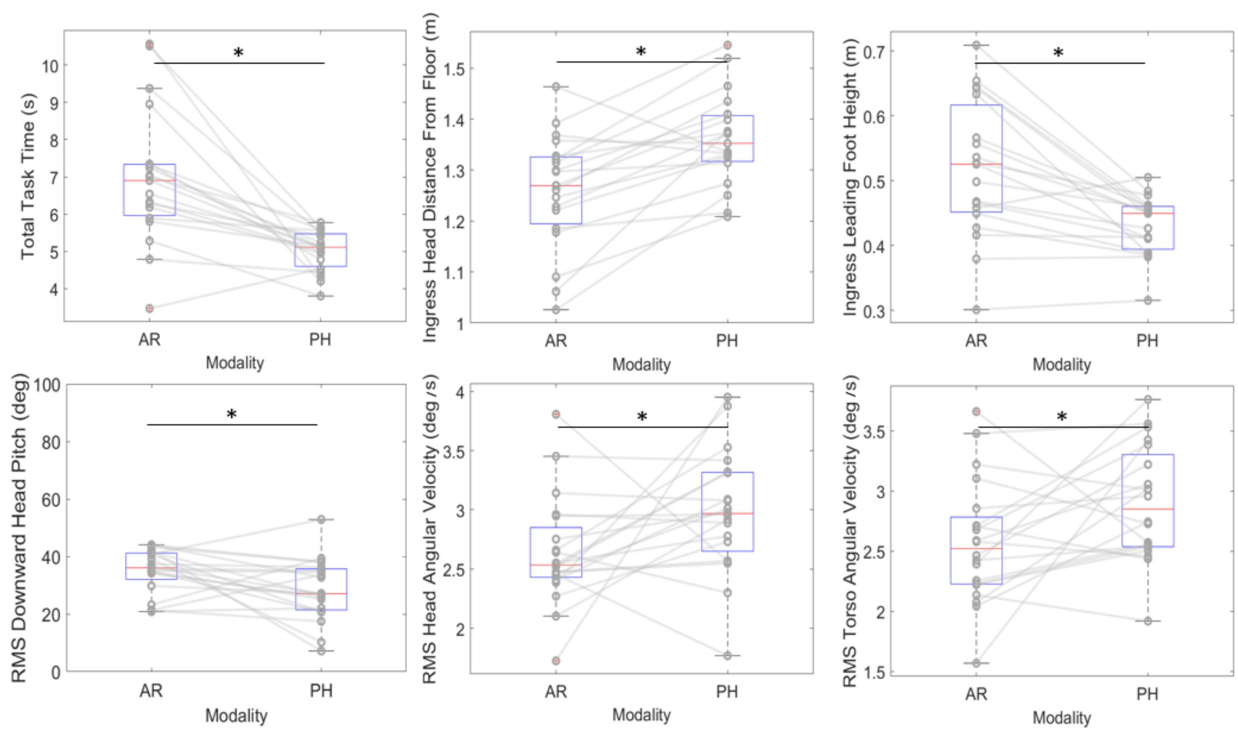

Figure 6.2: Ingress and egress task measures include total task time, head distance from the floor, second leading leg height, RMS of downward head pitch, and RMS of head and torso angular velocity.

Ingress & Egress Task				
	d.f.	**t**	**P-Values**	**Effect Size**
Total Task Time (s)	19	-5.05	$p < 0.001^*$	$d = 1.40$
First Leading Leg Step Height (m)	19	-1.99	$p = 0.061$	$d = 0.54$
Second Leading Leg Step Height (m)	19	-2.78	$p = 0.012^*$	$d = 0.73$
First Following Leg Step Height (m)	19	-2.02	$p = 0.058$	$d = 0.56$
Second Following Leg Step Height (m)	19	-2.31	$p = 0.032^*$	$d = 0.64$
First Head Pitch Height (m)	19	5.43	$p < 0.001^*$	$d = 0.92$
Second Head Pitch Height (m)	19	5.26	$p < 0.001^*$	$d = 0.89$
First Step-Through Duration (s)	19	-4.19	$p < 0.001^*$	$d = 1.05$
Turn Duration (s)	19	-4.51	$p < 0.001^*$	$d = 1.25$
Second Step-Through Duration (s)	19	-4.67	$p < 0.001^*$	$d = 0.99$
Obstacle Weave				
	d.f.	**t**	**P-Values**	**Effect Size**
Total Task Time (s)	19	-3.00	$p = 0.008^*$	$d = 0.53$
Right Footstep Height (m)	17	1.14	$p = 0.269$	$d = 0.15$
Left Footstep Height (m)	17	2.99	$p = 0.008^*$	$d = 0.32$
Average Step Height (m)	17	2.24	$p = 0.039^*$	$d = 0.23$
Step Height Variability (m)	17	4.26	$p < 0.001^*$	$d = 0.60$
Foot Trajectory Enclosed Area (m^2)	17	1.18	$p = 0.252^*$	$d = 0.31$
Combined Inner Boundary Foot Area (m^2)	18	-0.11	$p = 0.912$	$d = 0.029$
Inner Boundary Area Obstacle One (m^2)	18	-1.99	$p = 0.062$	$d = 0.46$
Inner Boundary Area Obstacle Two (m^2)	18	-0.61	$p = 0.55$	$d = 0.14$

Table 6.1: Task-specific measures of performance and related P-values and Cohen's effect size (d). The threshold of significance (< 0.05) is denoted with an asterisk (*).

6.2.1 Ingress and Egress Task

The ingress and egress task yielded a significant difference in the total task time between AR and Physical conditions, with increased overall task duration while using AR (Table 6.1 and Figure 6.2). The ingress and egress task segments (first and second attempts through the portal and turn duration) exhibited increased step-through duration for AR compared to the physical condition.

While there was no significant difference between the step heights during the first time through the airlock, the step heights were higher during AR with the second leading leg and second following leg during the return traverse (Figure 6.2). During both the first and second traverses, there was a significantly lower head height measured from the floor for AR (Table 6.1, Figure 6.2), with a significantly greater RMS head pitch angle (Table 6.2). AR resulted in significantly lower RMS of the head and torso angular velocity (Figure 6.2). There was no significant difference between groups found for the RMS of the head acceleration, RMS of the torso acceleration, the magnitude of the acceleration and angular velocity signals for the head and torso, or the roll and yaw angles of the head and torso (Appendix E).

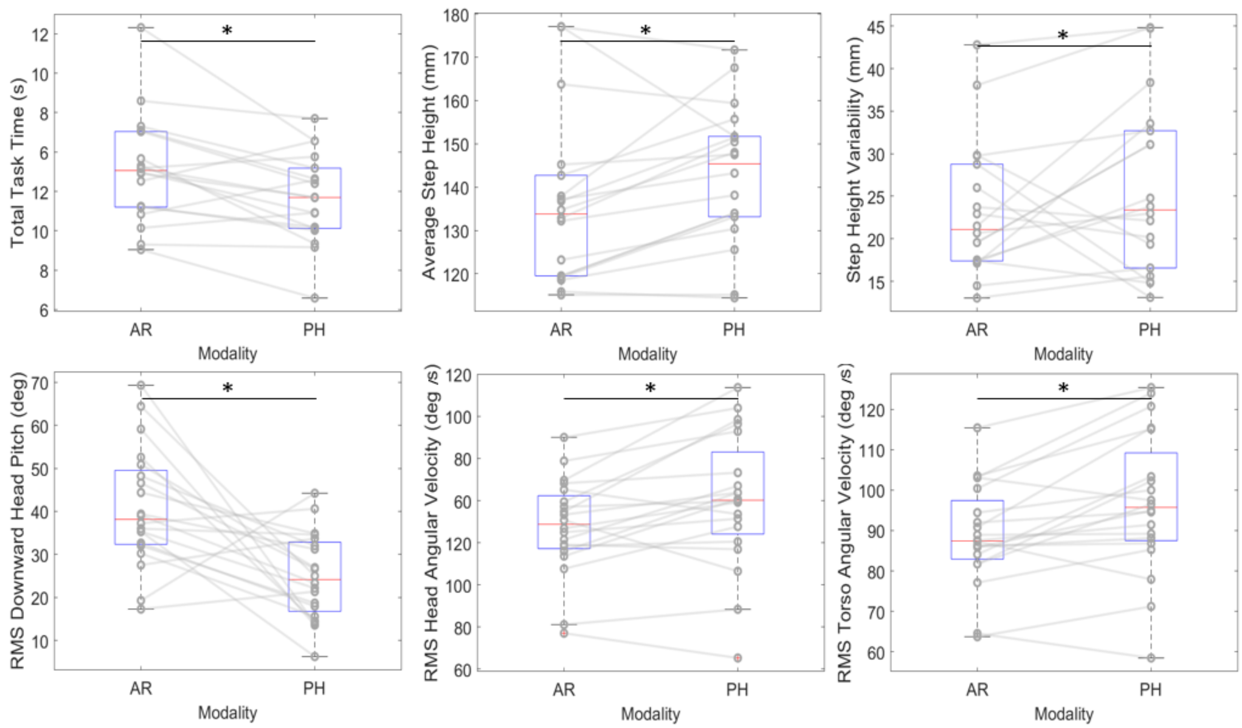

Figure 6.3: Obstacle weave task measures include total task time, average distance to the first obstacle, step height variability, RMS of the head pitch angle, and RMS of the head and torso angular velocity.

6.2.2 Obstacle Weave Task

Participants in the AR condition completed the obstacle weave task slower than when performed in the Physical condition (Table 6.1). In the Physical condition, participants stepped higher and with greater variability when performing the task. The results yielded no significant differences between the combined and separate inner boundary areas.

For the IMU-derived measures (Table 6.2) participants in the AR condition had significantly lower RMS of head acceleration, head angular velocity, torso acceleration, torso angular velocity, and the magnitudes of the head acceleration and torso angular velocity (Figure 6.3). No significant difference was observed for the magnitude of the head angular velocity or torso acceleration

IMU Measures	Ingress & Egress Task			Obstacle Weave		
	t	*P-Values*	*Effect Size*	*t*	*P-Values*	*Effect Size*
RMS Head Acceleration (m/s^2)	1.52	$p = 0.144$	$d = 0.21$	4.09	$p < 0.001^*$	$d = 0.55$
RMS Head Angular Velocity (deg/s)	2.80	$p = 0.011^*$	$d = 0.67$	3.07	$p = 0.006^*$	$d = 0.43$
RMS Torso Acceleration	1.79	$p = 0.089$	$d = 0.43$	3.31	$p = 0.004^*$	$d = 0.25$
RMS Torso Angular Velocity (deg/s)	2.63	$p = 0.017^*$	$d = 0.63$	3.66	$p = 0.002^*$	$d = 0.51$
Magnitude Head Acceleration (m/s^2)	-0.18	$p = 0.856$	$d = 0.04$	4.94	$p < 0.001^*$	$d = 0.96$
Magnitude Head Angular Velocity (deg/s)	0.62	$p = 0.543$	$d = 0.14$	1.38	$p = 0.182$	$d = 0.29$
Magnitude Torso Acceleration (m/s^2)	-0.79	$p = 0.437$	$d = 0.18$	0.63	$p = 0.539$	$d = 0.09$
Magnitude Torso Angular Velocity (deg/s)	-0.10	$p = 0.925$	$d = 0.02$	2.40	$p = 0.027^*$	$d = 0.54$
RMS Head Pitch (deg)	-2.28	$p = 0.034^*$	$d = 0.66$	-3.73	$p = 0.001^*$	$d = 1.28$
RMS Torso Pitch (deg)	0.03	$p = 0.977$	$d = 0.01$	-0.94	$p = 0.357$	$d = 0.20$
RMS Head Roll (deg)	-0.98	$p = 0.338$	$d = 0.30$	-4.94	$p < 0.001^*$	$d = 1.51$
RMS Torso Roll (deg)	-0.72	$p = 0.483$	$d = 0.17$	-2.19	$p = 0.041$	$d = 0.54$
RMS Head Yaw (deg)	1.64	$p = 0.118$	$d = 0.39$	6.74	$p < 0.001^*$	$d = 1.33$
RMS Torso Yaw (deg)	1.02	$p = 0.323$	$d = 0.30$	0.65	$p = 0.527$	$d = 0.13$

Table 6.2: IMU-based measures of performance and related P-values and Cohen's effect size (d). The threshold of significance (< 0.05) is denoted with an asterisk (*). The degrees of freedom were 19 for all comparisons.

(Appendix E).

The RMS of the head pitch angle and the head roll angle were significantly greater while using AR (Figure 6.3). In contrast, the RMS of the head yaw angle was significantly reduced within AR. For the RMS of the torso roll angle, participants exhibited higher magnitudes while performing the task within AR. The results yielded no significant difference in the RMS of the torso pitch or torso yaw angles (Appendix E).

6.2.3 Survey Results

Participants reported higher perceived mental demands while using AR to perform the ingress and egress task and the obstacle weave (Figure 6.4). The perceived task performance was rated higher when the tasks were completed in the physical condition. While the perceived physical

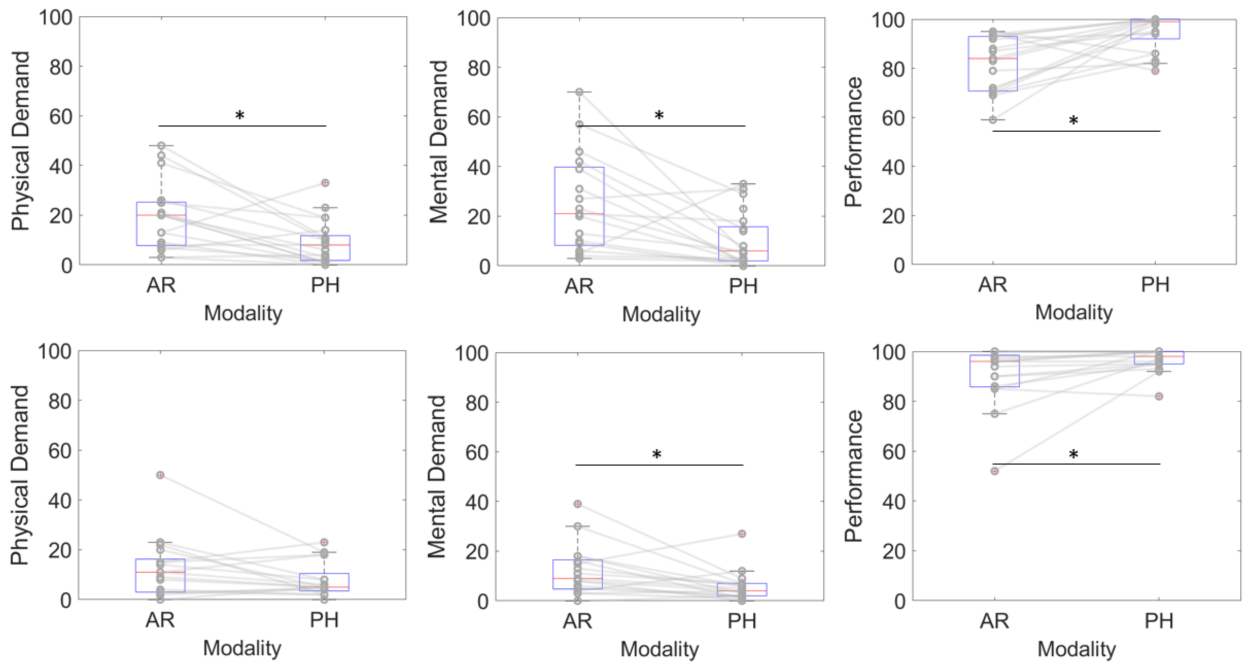

Figure 6.4: NASA TLX measures for the ingress-egress task (IET) and the obstacle weave task (OWT). Top: perceived physical demand ($p < 0.001$, $d = 0.66$), mental demand ($p = 0.001$, $d = 0.71$), and performance ($p < 0.001$, $d = 1.29$) for IET. Bottom: perceived physical demand ($p = 0.099$, $d = 0.42$), mental demand ($p = 0.023$, $d = 0.72$), and performance ($p = 0.034$, $d = 0.63$) for OWT.

demand, effort ($t(19) = -3.53$, $p = 0.002$, $d = 0.83$), and frustration ($t(19) = -2.80$, $p = 0.012$, $d = 0.39$) were rated higher in AR for the IET, there were no differences in physical demand, effort, or frustration observed for the obstacle weave (Appendix E).

6.3 Discussion

In the present study, healthy adults performed two operationally informed functional assessment tasks within an AR environment and while using physical objects. The selected assessments comprised an ingress and egress task and an obstacle weave task. Due to the limited field of view of the headset, it was hypothesized there would be differences in (1) task completion time, (2) step characteristics, and (3) head and torso kinematics as measured by accelerations, angular velocities, and orientations.

The results affirm our first hypothesis that participant performance would reveal differences in task completion times. The use of the AR environment led to prolonged completion times for both assessments. According to postflight research, crewmembers have altered locomotor function after long-duration space travel (163 to 195 days on the ISS), requiring a median of 48 percent longer time to complete the entire FMT courses one-day postflight relative to preflight [20]. Using analogous portions of the FMT, we found that the use of AR increased the ingress and egress task time by 39 percent and obstacle weave time by 12 percent. For the ingress and egress task, all aspects of the functional movement, such as the initial step-through, 180-degree turn, and second step-through, were performed slower while using AR, contributing to the overall extended task duration. AR produced similar results in terms of directional change along the temporal dimensions of the task as microgravity-induced locomotion, suggesting that AR use may increase task time further while one is also experiencing vestibular disorientation. An increase in the time required to complete a discrete task successfully is not inherently dangerous or unfavorable. The increased time is likely to be beneficial to someone experiencing motion sickness. However, to inform decision-making regarding EVA operations, a new threshold of performance is required to incorporate the longer baseline completion times observed in AR. In this study, the effects of continued use of AR on time completion were not examined; however, the times observed may serve as preliminary thresholds for healthy users. The development of thresholds for users with vestibular alterations, such as astronauts, requires further research.

The second hypothesis regarding step characteristics was supported. Participants exhibited greater caution during the IET in the AR environment by lifting their feet higher to avoid the bottom portion of the portal during the egress attempt following the 180-degree turn. While all participants maintained proper clearance of their feet with respect to the step of the airlock, one participant narrowly cleared the holographic step (0.02 m) with their following leg, which may suggest that their perception of the object's height may have been diminished and would have increased fall or tripping risk if the object were real. For the OWT, no differences were found in the enclosed walking area or the inner boundary area measures, suggesting that participants maintained consistent distances from the obstacles across both modalities. The consistency in how participants positioned their feet across modalities could be attributed to the limited task space and the need to circumnavigate the obstacles. Participants took lower steps during the OWT with respect to the ground and were less variable in their step heights within AR, suggesting a more cautious or deliberate approach for the obstacle weave task. The perception of task performance supports these findings, with higher perceived success within the physical condition and more mental demand within AR for both tasks. Additional investigations could explore alternative metrics, such as the average distance of the user's feet to the center of the obstacles; however, it is critical to ensure an accurate representation of the foot locations in the holographic reference frame to compare obstacle locations for distance-based analyses in AR. Existing AR systems lack the capability to track foot movements, necessitating efforts to incorporate aspects of foot-tracking features into new AR systems to support step characteristics for task analysis.

The third hypothesis was supported by differences in head and torso kinematics. With reduced peripheral vision of the holograms in AR, participants increased head pitch for both the IET and OWT. In addition to increased head pitch for the OWT, participants increased their head and torso roll angles to support consistent visualization of the obstacles while navigating in the figure-eight pattern. For the IET, participants lowered their heads further below the top portion of the airlock and adopted a greater downward head pitch. In contrast, torso posture remained consistent across both modalities. All participants were able to duck their heads low enough to clear the top of the holographic portal, although the perceptions of task performance were discordant with the direct measures analyzed. This discrepancy suggests that participants believed they had not sufficiently lowered their heads, potentially due to a constrained perception of the entire portal structure. Within the FMT, although most crewmembers have exhibited an ability to avoid the obstacles completely, a subset failed to avoid contacting the foam obstacles in testing sessions one day following their return to Earth [20]. While foam obstacles are generally safe, AR may provide a contactless and safe approach to assessing and training obstacle avoidance during periods of significant vestibular dysfunction both inflight and postflight on Earth.

The acceleration and angular velocity data also supported our third hypothesis regarding differences in head and torso motions, with decreased head and torso angular velocity for both tasks and decreased torso accelerations for the obstacle weave compared to the physical conditions. A modulation of head and torso angular motion may have been used to stabilize holographic images within the users' view. In the physical condition for the OWT, head and torso accelerations and angular velocities were higher, potentially indicating a greater confidence in performing the task and perceptions of object locations relative to their walking trajectories. The results suggest that participants were more deliberate in their movements while interacting with holographic objects, despite there being no risk of collisions or penalties. In operations, the reduction of motion sickness symptoms through reduced body movements in astronauts is likely to take precedence over holographic collisions. Therefore, metrics of successful task completion, such as object collisions, should be incorporated to track incidents of error especially if the goal is to determine EVA readiness where task precision is desired for safety. The constrained field of view of the headset, which required adjustments in head posture to keep objects in view, played a role in users' selected strategies. The reduced head and torso angular velocities observed in this study are similar to the crew performance observed while turning around a cone, which was three times lower on landing day than preflight [1]. The use of AR presents similar strategies as the reduction of postflight angular rate due to motion sickness and vestibular impairments [1]. Therefore, performing tasks in AR would not be expected to exacerbate these symptoms. In operational terms, when astronauts conduct extravehicular activities in spacesuits, they must contend with a reduced visual field of view and maintain a mental map of obstacles and objects for safe navigation. The increased perceptions of mental demand within AR may also be desired in an assessment task to better capture the cognitive demand present in operational tasks. Thus, the aspect of AR promoting more deliberate motion and mental demand could prove beneficial, in addition to other advantages of the technology that observes spaceflight resource constraints.

Although the decrease in accelerations and angular velocities resembled those of astronauts in deconditioned states, the adjustment in head pitch differed from crew performance, which could have either positive or negative implications. When astronauts transition between different gravity environments, increased head pitch can pose a challenge for reducing motion sickness symptoms. Postflight research demonstrates astronauts primarily reduce head pitch during locomotion, a strategy also utilized by vestibular deficit patients [178]. Ultimately, minimizing head movements serves as a rapid short-term behavioral countermeasure to remedy motion sickness [179]; however, crewmembers are encouraged to administer progressively larger self-paced head movements to facilitate readaptation to Earth's gravity [17]. The operational tasks in AR may require a larger head pitch due to the field of vision, which may encourage a controlled approach to promoting more head movements to assist in adaptation or generate increased symptoms of motion sickness. Crewmembers are advised to remain within their personal threshold of motion tolerance after landing to minimize motion sickness symptoms because performing head tilts too rapidly or with too much amplitude may aggravate symptoms [17]. Ultimately, user performance under vestibular and sensorimotor disorientation will differ and future research should examine how AR will impact these situations.

While the study explored several motion-based and time-based measures for a sample of twenty healthy participants, future studies should examine these findings across larger population samples and sensorimotor impaired populations such as recently returned astronauts. The obstacles chosen for the obstacle weave task were more representative of the cone dimensions used within other postflight assessments such as the walk-and-turn [180] rather than the dimensions of foam pillars used within the FMT [20]. Considering the AR headset's field of view constraints, shorter obstacles were used to allow users to see the entire object when directing their gaze. However, future investigations could consider the impact of obstacles that exceed the user's rendered field of view within AR, thereby potentially requiring more head movements to view the object. While this study was part of a larger protocol that evaluated six different dynamic assessments and functional tasks, additional evaluations of other relevant operational tasks to assess EVA readiness are needed to understand the potential benefit of using AR as a standalone assessment tool within spaceflight operations.

The current study was able to successfully use head measures to assess performance, a capability that could be extracted from the embedded IMU sensor of the HoloLens 2 to enable self-administered assessments with real-time data feedback to inform readiness. Nevertheless, additional body-worn sensors would need to be integrated to characterize torso and lower body performance. While no instances of collisions with task objects were observed in this study, potential future applications could introduce an audio alert if a user's head nearly makes contact with the top of the holographic airlock, particularly if this alert is deemed valuable for instructing crew members to adopt postures that optimize egress performance. The AR application could serve as a potential countermeasure to support the rehabilitation and readaptation of the crewmembers by encouraging gradually more extreme body postures that they will eventually need to perform in more difficult circumstances such as while wearing a spacesuit.

6.4 Conclusion

This research evaluated the effect of holographically administered operationally informed functional tasks on participant performance in an ingress and egress task and an obstacle weave task. The performance of twenty healthy adults was compared within augmented reality and physical object environments. As measured by inertial measure unit sensors and motion capture software, participants performed both tasks successfully, however, task completion strategies differed across modalities. Task completion time was longer, and the head and torso angular velocities were reduced for both assessment tasks when administered in AR. More extreme head pitch was observed across the tasks as well as increased average head roll orientation during the obstacle weave. Participants were more deliberate and careful with the task execution by stepping higher and lowering their heads further for the IET. The participants were able to complete both tasks in AR and meaningful measures of postural control were obtained, indicating that AR may be a useful instrumentation solution with embedded sensors to evaluate a variety of populations for postural control. Participants adopted strategies similar to sensorimotor impaired crewmembers, with increased task time and reduced head and torso angular velocities. However, the participants also had increased head pitch angles, which is typically reduced in astronauts postflight. From these analyses, AR may be a useful instrumentation solution for evaluating inflight performance with embedded sensors and could be part of a countermeasure toolset.

CHAPTER 7

Conclusions, Recommendations, and Future Work

This dissertation examined the viability of using augmented reality (AR) in spaceflight operations to evaluate astronaut crews' sensorimotor and neurovestibular performance after gravity transitions to determine extravehicular activity readiness. Using augmented reality and wearable sensors, preliminary evaluations of the use of AR for physiological assessment were performed by comparing performance against gold standard Earth-based implementations to determine how holographic content impacted performance. The research set out to address the following research questions:

1. Can augmented reality be used as a standalone assessment tool for neurovestibular and sensorimotor evaluations in astronauts?
   1.1 What dependent measures can be derived from the sensor capabilities of AR systems in evaluating balance and postural control?
   1.2 What dependent measures can be derived from the sensor capabilities of AR systems in evaluating hand-eye coordination?

Chapters 2, 4, 5, and 6 directly addressed questions 1.1 and 1.2 in order to aid in the evaluation of question 1 as a whole. Various dependent measures derived from the AR system and external systems were evaluated across the sensorimotor and balance assessments to quantify user performance. Hand and eye tracking metrics were facilitated by the AR headset and can serve as an embedded capability for future applications. In order to inform decision-making for telerobotics and other tasks requiring fine-tuned hand-eye coordination, finger-tip location and eye gaze can be measured with respect to task objects. Body-worn inertial measurement units were used to quantify differences in the AR and physical tasks, however, metrics from the IMU located on the head, such as accelerations, angular velocity, and orientation, can be obtained directly from the IMU within the headset for future applications. A motion capture system was utilized to evaluate AR-based task performance by extracting task completion time, step characteristics, and head position relative to the ground. Task completion time and head position with respect to global space can be captured from the AR headset while step characteristics are not supported by the current technology. The use of two-dimensional matrix barcodes, such as QR codes, on the feet could assist in identifying foot positions, but the user would need to be looking at their feet in order for the headset sensors to calculate foot position. Nonetheless, integrating the dependent measures for the eyes, head, and hands could enable a standalone assessment tool.

2. For self-assessment of physical abilities, what usability features for design and user interactions need to be considered?

    2.1 What usability concerns are discovered with younger healthy adults?

    2.2 What usability concerns are discovered with older healthy adults?

Chapter 3 identified usability concerns from both younger and older users to evaluate questions 2.1 and 2.2. Emergent usability concerns relate to the difficulty of learning and executing the AR hand gestures correctly, proper depth perception of holograms, and the ability to perform system navigation for autonomous use without external support. With consideration for application design and training protocols, recommendations are provided to improve usability across these user difficulties in Section 7.2.

3. What impact does the use of holographic objects in assessments traditionally performed with physical objects have on overall performance?

    3.1 When compared to traditional balance assessments, how does AR impact postural control and task completion measurements?

    3.2 When compared to traditional hand-eye coordination assessments, how does AR impact hand-eye coordination and task completion measurements?

By comparing performance within the AR modality to physical implementations of the assessment with real objects, Chapters 4, 5, and 6 evaluated questions 3.1, and 3.2. As compared with a touchscreen device, AR performance was lower across the dependent measures that were analyzed in Chapter 4. Performance within AR for the dynamic balance assessment tasks resulted in longer task completion times and alterations to step characteristics (Chapter 5), while performance in AR for the operational assessments also exhibited longer task completion times and reductions in head and torso angular velocities (Chapter 6).

7.1 Summary of Results

The purpose of this section is to summarize the main results obtained from each chapter in this thesis and contextualize these findings in light of the research questions being examined.

7.1.1 Can augmented reality be used as a standalone assessment tool for neurovestibular and sensorimotor evaluations in astronauts?

During the development of the AURORA application (Chapter 2), a strong emphasis was placed on (1) integrating traditional physical assessments with holographic configurations, (2) providing visual cues and user interfaces to support self-administration, and (3) making use of the embedded sensor streams to extract performance metrics. The comparative analyses (chapters 4, 5, and 6), utilized metrics of performance extracted from chapter 2 to inform the impact of holographic content on user performance. However, these measures can be used to strictly evaluate performance within AR. Through these design aims and evaluative testing, we were able to address the viability of AR as a standalone assessment tool across various domains of utility, namely, holographically administered assessments, usability for autonomous use, and data output capabilities. In order to support spaceflight operations and resource constraints, these domains of system utility are essential.

Chapter 2 introduced the successful integration of six widely used sensorimotor and vestibular assessments within an AR environment, namely, single leg and tandem balance, the four square step test, the star excursion, tandem walking, and a multidirectional tapping task with various indexes of difficulty. In addition to these gold-standard clinical assessments, new operationally-informed assessments such as the capsule ingress and egress task, the obstacle weave, and the geology sampling task were also successfully integrated into AR. Physical floor-based objects traditionally used within these assessments were replicated into holographic content thereby rendering the headset a standalone tool for administering these tasks. The evaluation of task completion performance was conducted by utilizing clinically-based tasks and performance metrics, comparing the results to relevant literature with similar protocols as appropriate. The research established the groundwork for well-informed criteria for assessing performance in AR, however, more extensive sample sizes and populations are required to build readiness thresholds.

The development of AURORA demonstrated the potential for AR-based performance assessments that are easily deployable and observe resource and volumetric constraints of future spacecraft and mission operations. All the tasks within AURORA were executable within a minute or less as supported by evaluative testing and task duration measures (Chapters 4, 5, and 6). Throughout usability and evaluative tests, all seventy-three participants completed each task and remained within the limited 2x2 meter area, proving that these assessments could be conducted within a confined space such as a capsule. As AR allows users to visualize their physical environments, participants did not sustain any falls or injuries while using the system. The benefits of holographic content include safe collisions with objects, i.e., crew members are not likely to trip or collide with objects, which could result in unsafe evaluations, especially when they are experiencing vestibular disorders following gravity transitions.

The key attributes that render AR a prospective solution for future in-flight assessment tools are the data output and sensor capabilities seamlessly integrated into the HoloLens architecture. Chapter 2 presented the available sensor streams that enable hand and eye tracking capabilities, head orientation and kinematics tracking, and head positional data with respect to global space. As a result of the evaluations of sensorimotor hand-eye coordination and neurovestibular postural control, meaningful performance measures were obtained, suggesting that AR is a useful instrumentation solution with embedded sensors. Hand tracking data enabled the quantification of task accuracy, precision, movement time, error rate, and derived measures such as throughput for the multidirectional tapping task in chapter 4. While eye tracking data was captured, direct evaluations of this dataset were not conducted within this thesis; however, future studies may examine this aspect of performance. Inertial measurement units enabled quantification of head downward pitch, roll, and yaw angles, accelerations, and angular velocities in chapters 5 and 6. As a result of the embedded sensors, unique performance across all assessment tasks was revealed when compared to traditional physical implementations of the tasks. Although these measures were used to quantify and compare performance in AR to physical objects for the current research, they can also be used to develop performance thresholds across mission duration, i.e., preflight, inflight, and postflight.

In addition to the sensor capabilities, the HoloLens headset features onboard computing and storage to enable real-time data feedback and storage of relevant performance metrics which are both vital for future inflight assessment tools. In both usability and evaluative testing, augmented reality was found to be an effective tool for administering neurovestibular and sensorimotor assessments, thereby providing strong evidence for the viability of AR use in future inflight evaluations. Nonetheless, when making comparisons across modalities, care should be taken, considering the performance variations that emerged when using AR that were partly attributed to the restricted field of view.

7.1.2 For self-assessment of physical abilities, what usability features for design and user interactions need to be considered?

As space operations become increasingly autonomous, empowering crew members to self-administer assessments becomes a vital capability of future assessment tools. In chapter 2, design affordances were introduced specifically for AURORA to enhance autonomous use, including task animations accompanied by both verbal and written descriptions. Chapter 3 evaluated the extent to which usability participants felt confident in their ability to self-administer assessments following a single training session amounting to an hour of uninterrupted use. In future operations, astronauts may use the application several months after training, and therefore intuitive use and the ability to relearn the system is imperative. The majority of participants in both the younger and older cohorts positively reported confidence in their ability to self-administer the assessments in the future and rated the animations and audio instructions as supportive. Although ratings of mental and physical demand were slightly higher when comparing AR to task completion in the physical implementations in Chapters 4, 5, and 6, ratings were low when participants were only presented with the AR environment in the usability testing (Chapter 3). In the absence of the ability to compare performance to the execution of the task with physical objects, participants perceive demand to be low, which is promising for future use of the application.

The usability results demonstrated that self-administration of the assessments was most likely to be affected by the ability to learn and correctly perform the AR hand gestures (chapter 2). Although the vast majority of participants were able to learn and interact with the system through the selected hand gestures, a subset of users struggled, suggesting that further training and flexibility in permissible hand postures is warranted for broad use across all user demographics. Several users had difficulties interacting with the system because they were unable to perform the pinch gesture as instructed and had difficulty determining hologram location with respect to depth for touch selections. Design recommendations to support depth perception, accuracy, and the execution of head gestures are presented in the Section 7.2. Ultimately, the research demonstrated the foundation for future app design and training features necessary for self-administered AR assessments to support a standalone performance evaluation system.

7.1.3 What impact does the use of holographic objects in assessments traditionally performed with physical objects have on overall performance?

Chapter 4 evaluated how virtually presented holographic targets impact users' accuracy, precision, movement time, throughput, and error rates when compared to a touchscreen device for a multidirectional tapping task. The use of augmented reality yielded decreases in performance across the measures of accuracy, precision, and throughput while exhibiting increases in movement time and error rate. Participants rated stronger negative ratings of physical and mental demand, frustration, effort, performance, and arm fatigue within AR when compared to the touchscreen device. Across all the dependent measures, the touchscreen modality yielded improved user performance compared to AR in line with the hypotheses set forth. Despite lower performance for the same index of difficulty when compared to a touchscreen, AR may still be a useful tool for the longitudinal assessment of hand-eye coordination with the benefit of embedded hand and eye tracking. The levels of difficulty selected within this research would still be appropriate for measuring user abilities in applications relating to sensorimotor performance where a higher difficulty is desired

AR vs. Gold Standard	Dependent Measures							Participant Perceptions		
Assessment Task	*Time to Complete*	*Task Variability*	*Error*	*Head Pitch*	*Head Motion*	*Torso Pitch*	*Torso Motion*	*Perceptions of Performance*	*Mental Demand*	*Physical Demand*
Multidirectional Tapping Task	↑ (red)	↑ (red)	↑ (red)					↓ (red)	↑ (orange)	↑ (red)
Four Square Step Test	↑ (orange)	↑ (orange)	–	–	–	–	–	↓ (orange)	↑ (orange)	–
Tandem Walk	↑ (green)	–	–	–	–	–	↓ (green)	–	–	–
Capsule Ingress and Egress	↑ (red)		–	↑ (orange)	↓ (orange)	–	↓ (orange)	↓ (red)	↑ (orange)	↑ (orange)
Obstacle Weave	↑ (orange)	↓ (orange)	–	↑ (red)	↓ (orange)	–	↓ (orange)	↓ (orange)	↑ (orange)	–

↑ Increased in AR
↓ Decreased in AR
– Consistent across modalities
Not applicable, i.e., did not assess

↑ (red) - Large effect (≥ 0.8)
↑ (orange) - Medium effect (0.5 ≤ d < 0.8)
↑ (green) - Small effect (0.2 ≤ d < 0.5)

Figure 7.1: Summary table of the comparative analysis studies across the dependent measures. The results represent the impact of AR use on participant performance and user perceptions. The results are presented as directional change (increased, decreased, no change) and are color-coded by the relative effect size (green, orange, and red corresponding to small, medium, and large effect sizes).

to evaluate hand-eye coordination. The embedded sensors provide supplemental information that can provide greater insight into sensorimotor performance than touchscreens alone. In order to assess how sensorimotor performance might be affected by extended exposure to microgravity, hand tracking allows evaluations of users' hand trajectory efficiency and eye tracking provides information on gaze metrics. The fusion of these two sensor streams allows one to characterize hand-eye coordination beyond simple temporal or accuracy measures.

Chapters 5 and 6, evaluated the influence of holographic objects and augmented reality-administered assessments on participant performance and task completion through comparisons with physical objects. Although participants were able to perform all of the assessment tasks successfully, task performance differed in AR as measured by inertial measure unit sensors and motion capture software. The use of holographic objects in place of traditional physical objects influenced user task completion time, step characteristics, and head and torso movements in each of the dynamic postural control tasks. A high-level summary of the impact of augmented reality on task performance across the dependent measures is presented in Figure 7.1.

Two clinically-informed assessments, the Four Square Step Test and the Tandem Walk were evaluated in chapter 5. The results yielded increased task completion time across both assessments when administered within AR. For the Four Square Step Test, participants took more time to translate in the rightward direction and exhibited higher foot placement variability (i.e., less precision) within the task quadrants. Perceptions of performance revealed participants were less confident in their FSST performance while using AR, which aligns with the increased variability observed in foot placement. The difference in foot placement was not driven by instability as there was no significant difference found between the modalities for the head and torso acceleration and angular velocity signals. The reduced field of view of the headset may have contributed to these findings since peripheral vision was not well supported in the headset.

The Tandem Walk also showed distinct user strategies that were influenced by the reduced field of view of the headset and the presence of holographic content that partially occluded the lower limbs. Participants stepped higher on average in AR. Additionally, in an effort to maintain head stability for viewing the holographic content, participants restricted their torso movement more in the AR environment. While none of these user strategies discovered influenced the successful completion of the task, the results indicate that performance should be assessed carefully across modalities for the same task. In the case of spaceflight performance assessments, AR-based thresholds of performance would need to be developed to inform changes in performance preflight and postflight.

Two operationally-informed assessments, the ingress and egress task, and the obstacle weave task were evaluated in Chapter 6. AR exhibited increased task completion times and decreased magnitudes of head and torso angular velocity across both tasks. To promote visualization of the holograms with reduced visual field of view in AR and to promote higher clearance with the task objects, more extreme head pitch angle magnitudes were observed. By stepping higher and lowering their heads further for the ingress and egress task and maintaining further distances from the obstacle weave objects, participants were more deliberate and careful with task execution in AR. The increased task time and reduced torso angular velocity were similar to those strategies utilized by crewmembers with sensorimotor impairments, except in cases of extreme head pitch angles that could either promote gradual readaptation to gravity-rich environments or exacerbate their symptoms. Therefore, performing these tasks in AR would not be expected to exacerbate motion sickness symptoms if crew members remained within their personal threshold of motion tolerance. Even though AR had no impact on the successful completion of the task, thresholds should be developed to assess EVA readiness, especially in populations with vestibular impairments.

7.2 Augmented Reality Interface Design Recommendations

This section provides design guidance for AR mission display devices in NASA's Human Information Design Handbook (HIDH) [110], which currently lacks guidance for Extended Reality

(XR) technologies. A single mention of augmented reality is made in the handbook and the note relates to frame rate and latency constraints of 60 Hz and 16 ms, respectively. The current recommendations provide guidance for section 10 within the HIDH which considers crew interfaces.

7.2.1 Button Size Recommendations

- In general, AR interfaces should have wider targets and shorter movement distances between successive targets than physical buttons, which corresponds to lowering the index of difficulty for reaching movements.
- For rapid-aimed movements in AR, target sizes should be larger than 0.019 meters if the accuracy and efficiency in AR are desired at the level of performance seen with a touchscreen.
- While AR task performance in rapid-aimed movements improved with 0.025 meters target width compared to 0.019 meters, errors persisted suggesting interface buttons should exceed these dimensions, especially with consideration for the sensorimotor challenges faced by crewmembers.

Larger buttons have been assessed as a countermeasure for spaceflight-induced dysmetria [142] where crew members may overshoot when reaching for switches, buttons, pointing, or applying forces to objects [9]. Considering AR interactions may be inherently more difficult than physical interactions, providing larger buttons should be especially considered to support altered human performance capabilities in microgravity.

7.2.2 Hologram Placement and Interface Recommendations

- When possible, holographic interface displays should be placed close enough to support crewmember's perceptual abilities while remaining within the field of view of the headset's rendering capabilities.
- World-fixed holograms should be set at an initial distance of 0.5 meters from the user's head for legibility.
- Where ambulation is permitted, 0.5 meters is still recommended as the environment allows the user to reposition themselves closer in relation to the hologram if needed.
- Hand-attached menus for frequently used actions are recommended because touch selection with buttons located on the opposite hand was easier than holograms located at a distance.

- Holograms located at a distance required sufficient depth perception and the execution of hand gestures correctly. Proximity lighting, in which the surface color of the hologram adjusts as the index finger approaches can support the user's depth perception and button interactions.
- In addition to visual cues, audio cues are also recommended to provide the user with auxiliary feedback for successful target acquisitions.
- For interface displays and holographic information panels, it is suggested to use surface and text colors with adequate contrast, such as blue panels with white writing which is used throughout AURORA.
- For panels with a large quantity of text and details, such as assessment instructions, designers should consider providing audio dictation for users who prefer to listen to instructions.
- AR designers who wish to build assessment applications and tools should consider pairing instructional content with visual representations or demonstrations of the task.
- Consistency should be maintained across sequences of actions but also across prompts, menus, interface colors, fonts, layouts, and symbols.
- During the design of interfaces, designers should consider the prevention of errors and easy reversal of actions.
- Current AR technology does not support lower body tracking, therefore holograms placed on the floor may partially occlude the user's lower limbs. For floor-based holograms, especially intended for user interaction with the lower body, designers should consider hollowed-out or wireframe-like structures to minimize hologram and user occlusions.

The suggested AR viewing distance recommendation is in accordance with the NASA HIDH [110], NASA-STD-3000 [142], and the Human Factors Ergonomics Society standard [181] for the viewing distance of displays. The recommendation is also supported by the best practices offered by Microsoft for designers where holograms placed closer than 40 cm are discouraged. The optimal zone for hologram placement is stated between 1.25 and 5 meters [182]. According to research regarding augmented reality text integration, white text displayed on a blue holographic background has been described as the best color combination, especially when the user's background is not plain [183, 184]. Current user interfaces such as computers uphold consistency with common actions and support the prevention of errors with dialog boxes requiring user input. It is essential to uphold common interface attributes found in frequently used devices within augmented reality (AR) to facilitate transferability through familiarity, ultimately enhancing usability for first-time users of AR systems.

7.2.3 Hand Gesture Recommendations

- Pointing selection is recommended across interface interactions as this AR gesture was the most intuitive because the gesture resembles interactions with touchscreen devices.
- Raycasting, the method by which one can interact with holograms at a distance with pointing and pinching, can also be used for instances where the user cannot reach a hologram. Although further training and practice may be required to perform this gesture correctly.
- It is important to note that training protocols should remind users that camera-based hand-tracking AR devices will not be able to sustain effective finger tracking if the hand is out of range of the cameras or the finger is occluded.
- This research determined that keeping the tip of the index finger within the tracking system's field of view was best achieved by maintaining hand postures above the horizontal during pointing gestures to avoid occlusions with the back of the hand.

For pointing selection with the index finger, researchers and application developers should be cognizant of the system and user limitations. Namely, the finger must remain visible within the AR systems tracking cameras to support gesture recognition and the successful completion of input execution. Similarly, participants should be instructed to minimize occlusion of their fingertip with the back of their hand by holding the hand more vertically, especially when the interaction space is located at eye height where the tracking sensors are located. A lower interaction space, such as holograms at chest height, allows horizontal hand postures to be supported since the back of the hand no longer obstructs the finger joints that are tracked by the AR system. Microsoft presents an optimal placement region of holograms at zero to thirty-five degrees below the horizon measured at eye height [182]. The upper region to support user comfort is zero to ten degrees above the horizon while the lower region is thirty-five to sixty degrees below the horizon. Lowering the interaction space may also mitigate arm fatigue which has been shown to increase with long duration in overhead postures within extended reality technologies [155, 153, 154, 156]. While speech input was not evaluated in the current research, future designers may consider replacing repeated gesture input with speech input to further mitigate arm fatigue.

7.2.4 Assessment Recommendations

- For operational assessments, designers should consider common functional requirements of EVA activities to develop meaningful measures of performance.

- Assessment tasks should be developed with the intended use case in mind. In other words, if the task is to be conducted in a space capsule, the assessment should be scaled appropriately to enable safe and effective utilization in a confined space.
- Extracting the embedded inertial measurement unit sensor data should be a built-in AR application capability for monitoring head orientation, acceleration, and angular velocity throughout task completion.
- Eye calibrations should be conducted prior to use to ensure hologram clarity and viewing.
- Designers may consider extracting eye-tracking or hand-tracking sensor data to support data metrics for the quantification of hand-eye coordination. It is important for designers to establish which hand joints are of interest for tracking to inform task performance. For example, the index finger can quantify the user's selection finger with respect to buttons and targets while the palm joint can give an overall position of the hand throughout the interaction with potentially fewer occlusions.
- Training of AR systems should include AR gesture demonstrations, system navigation, where to access system information, and how a user can perform reversal of actions if mistakes are made.
- For the placement of floor-based holograms, designers may implement sliders to fine-tune the positioning of the holograms or enable the headset's scene-understanding capabilities to magnetize objects to physical boundaries such as the floor or wall.
- In future AR analyses, researchers should consider measuring depth perception and visual dominance in addition to handedness to determine these aspects of user capability and performance.

Emergent findings from the research demonstrated user difficulty with depth perception of holographic content. While depth perception and visual dominance were not captured for the present population samples, future studies should consider adding pretest screening protocols to assess user's perceptual abilities. Demographic data associated with perception may aid in identifying users who struggle with depth perception and may lead to design changes that can assist these users.

7.2.5 Connections to Current Literature and Analogous Industries

The impact of interface display on operator performance has been studied extensively in analogous operational settings such as the aviation industry. Over the past two decades, Commercial aircraft have started to incorporate computer-based flight instrumentation with touchscreen displays thereby replacing traditional physical controls including switches, knobs, levers, dials, and keypads [185]. These new touch-based displays pose potential advantages and disadvantages to operations. The following section offers connections from aviation flight display research to the current augmented reality research relating to the impact of display modalities and emergent design recommendations.

Touchscreen displays have been identified as advantageous alternatives to traditional physical input modalities by offering ease of use with direct interaction, modifiability and upgradability, and reduced weight [186]. However, vibration, primarily introduced through turbulence, can be a factor diminishing operator performance. Degradations to human performance in motion environments can result in task inefficiency, task inaccuracy, fatigue, and motion sickness. Numerous studies have evaluated touch-based and physical-based input methods across various interactive tasks and turbulence levels while considering task performance, display location and size, and workload and fatigue for pilots flying during simulations. Cockburn et al. [186] demonstrated that performance decrements were observed across touch-based and physical-based displays as turbulence increases. Touchscreen displays resulted in faster interactions than trackball interfaces when precision requirements were low, while diminished performance was observed with touchscreens when precision was required. Participants were shown to utilize strategies such as stabilizing their hand on the screen's edge to facilitate their ability to interact as vibrations increased. These findings are comparable with the current literature when considering the speed-accuracy tradeoff observed during user performance in AR compared to touchscreens with different indexes of difficulty. When target widths were larger, i.e. lower precision requirements, faster relative movement time was observed across both AR and touchscreen modalities. In AR, participants modulated speed and performed more deliberate aimed motions to facilitate target acquisitions without tactile feedback. The findings from the present work and research concerning interface design for flight decks suggest that although distinct user interface modalities can directly impact performance, other environmental constraints (such as the operator's sensorimotor ability and environmental vibration) also play a role in overall operational performance.

In addition to environmental factors like vibration and task goals such as precision restrictions, other task constraints including target location and size have also been associated with varying levels of task performance and user perceptions. Dodd et al. [187] identified that the size and location of touchscreen displays were important factors impacting task performance during different turbulence levels. Larger size displays (15 inches diagonal), compared to smaller displays (8 inches diagonal), improved map planning task time suggesting visible area as an effective design capability to improve touch-based flight displays. Higher perceptions of fatigue, more errors, and longer task completion times were identified with interactions on touch-based displays overhead and out-board (e.g., located at the side of the pilots). The higher perceptions of fatigue with overhead target acquisitions were also exhibited in the current research when comparing task performance in AR with a touchscreen, suggesting location-based recommendations for human-computer interaction displays across operational applications. Common user-required actions should be reserved for forward-located displays while overhead actions should be limited or assigned to more uncommon user actions, a recommendation also detailed in occupational ergonomics literature [188].

The physical characteristics of an input system also impact how users interface with a given human-computer interaction system. Dodd et al. [187] presented a disadvantage of touchscreens when compared to physical buttons and knobs, which is the lack of direct physical feedback after user input is received. When comparing touchscreens to a physical-based input such as a button, physical depression or tactile clicks are provided. With a pull switch, direct feedback is provided through the active or inactive states of the button based on the positional location of the switch. In contrast, touchscreens do not provide the pilots with critical physical feedback that the device has received the intended input actions. While additional assistive cues such as auditory tones can support one's knowledge of successful input actions, user performance may still be impacted without tactile feedback. The present research revealed performance decrements within AR compared to a touchscreen, which may partially be due to the complete absence of tactile feedback provided by hologram-based interactions. While the AR interface was equipped with proximity lighting for interface buttons, depression animations for click events, and auditory sounds during user input actions, performance was still constrained when compared to a touch-based interface. These findings suggest that physical-based, touch-based, and augmented reality-based interfaces provide potential advantages to various operational occupations, however, the design and implementation of these displays can influence operator performance, workload, and fatigue.

Operator performance for aimed movements in operational settings can be supported through design; however, each interface modality may require distinct multi-modal design recommendations due to the innate characteristics of the interface. While visual, auditory, and haptic feedback may support both touchscreen and AR interfaces, the implementation of these assistive cues may be influenced by the operational needs of the users and the task demands. NASA's Human Integration Design Handbook provides a standardized recommendation for various mission display interfaces within the context of operational tasks and astronaut's physical capabilities; however, guidance is currently lacking for AR-based displays. The HIDH [107] recommends touchscreen button target height and width range from 9.5 mm to 22 mm (page 995) while physical button dimensions range from 10 mm to 25 mm for barehanded interactions with the fingertip. The current recommendations from the present research lie outside these ranges and ultimately provide preliminary recommendations for augmented reality interfaces. Formal analysis of various holographic button dimensions may need to be performed to develop an acceptable range of dimension values.

7.3 Research Contributions

This thesis makes several contributions to the existing literature for augmented reality assessment tools, rehabilitation, and aerospace countermeasure development:

1. AURORA is the first AR application to incorporate earth-based standards for inflight assessments of vestibular and sensorimotor assessments.

2. AURORA provides custom embedded algorithms to extract user performance measures (eye tracking, hand tracking, head kinematics, and head positional data) from the AR headset's onboard sensors.

3. While adhering to human factors design principles, AURORA demonstrated that verbal and written task instructions combined with animations promoted task comprehension and confidence in self-administration of assessments.

4. Guidance on AR design and training for both young and older users is offered to address emergent usability concerns related to depth perception, user input, and training.

5. The research demonstrated a training protocol that sufficiently yielded asymptotic performance in participants for a multidirectional tapping task.

6. Four distinct acquisition error types, three of which are unique to augmented reality interactions, were characterized and quantified with respect to aimed target acquisitions.

7. User perceptions of measures of workload and fatigue were rated more harshly when users were able to compare to another modality, rather than evaluating AR use alone.

8. The work demonstrates that despite rapid aimed movements in AR yielding poorer performance compared to touchscreens, AR still provides useful differentiated performance across task difficulty.

9. User strategy as measured by step characteristics and head and torso kinematics were shown to be influenced by holographic content and field of view constraints, suggesting that modified thresholds of performance are needed to qualitatively score performance in dynamic and operational AR assessments.

10. The use of AR for goal-orientated tasks resulted in longer task completion times suggesting a new temporal threshold of performance is required to incorporate the longer baseline completion times observed in AR.

11. Perceptions of task performance were generally discordant with direct measures of performance such as error rate, suggesting less confidence within AR for completing task objectives due to lower perceptions of task success. The influence of prolonged augmented reality (AR) usage on perceptions of task performance remains uncertain; nonetheless, it appears that first-time users and those with limited experience tend to perceive poorer performance in AR.

12. For vestibular assessments, meaningful measures of task performance were obtained and the use of AR did not impact the successful completion of the functional tasks, indicating AR may be a useful space-conscious alternative to traditional assessments.

13. The research provides design guidance for AR mission display devices in NASA's Human Information Design Handbook (HIDH) [107], which currently lacks guidance for Extended Reality (XR) technologies.

14. The dissertation presented a new assessment device designed to assist astronauts in autonomously performing vestibular and sensorimotor evaluations during spaceflight, without reliance on ground support. This innovation directly informs the knowledge and technology gap (SM-204) identified in NASA's Human Research Program Roadmap relating to the development and testing of post-planetary-landing self-administered testing and rehabilitation tool.[105].

7.4 Contribution of Research to NASA Human Research Program Gaps

This thesis informs gaps identified in NASA's Human Research Program Roadmap relating to the development and testing of assessment tools and countermeasures for evaluating sensorimotor function and quantifying the impact of weightlessness on postural control, locomotion, and manual control. The development and testing of AURORA directly supported the SM-204 roadmap gap associated with the creation of a post-planetary-landing self-administered testing and rehabilitation tool [105]. The SM-204 research gap specifically requests "using ISS as a post-planetary-landing, verify a set of self-administered countermeasures to mitigate motion sickness and enhance functional recovery." The gap also seeks to "characterize how these prevention or treatments reduce the risk of injury or enable crewmembers to perform critical operational tasks, e.g., using sensorimotor spaceflight standard measure, relative to the outcomes using standard supervised reconditioning program" [105].

To address the self-administered aspect of the SM-204 research gap, the AURORA system was developed with prominent human factors principles to support administration without external assistance and usability for first-time users and users who may experience extended duration between uses of the system. The human-in-the-loop usability testing demonstrated positive perceptions of usability across two diverse age groups spanning current and future astronaut age demographics. The work identified aspects of interface design that supported task understanding such as written and verbal instructions accompanied by animations. Although usability concerns were identified with a subset of users, future design recommendations are provided to promote self-administration for all users. While healthy non-astronaut participants were used throughout the research, there may be opportunities to extend the research to ISS missions with astronauts.

The evaluation of postural control and task completion strategies for dynamic and operational balance assessments within AR compared to physical objects evaluated the viability of AURORA as a post-planetary-landing self-administered testing tool (SM-204) [105]. The research demonstrated that users were able to complete the functional task successfully within AR and meaningful measures of postural control and performance were obtained that can be extracted directly from the AR headset to promote a standalone assessment tool. Although task variability such as step height varied in AR there was no impact on task error for the healthy participants who volunteered in the research. AURORA provides a relatively unobtrusive monitoring tool for head movement to potentially provide guidance for the level of reconditioning needed by a given crew member. Performance measures derived from the research can provide quantification of a crew member's relative risk of injury such as an increased risk of falls due to greater postural sway thereby directly addressing SM-204.

The SM-101 gap identifies sensorimotor risk and the need to "define the magnitude of change and time course of recovery, and how these impact critical mission tasks" [102]. The gap also seeks to "define how the risk (likelihood and severity) varies as a function of microgravity transit time, magnitude of G transition, and other contributing factors (prior experience, biomarkers, space radiation)" [102]. The present research demonstrates a means to assess postflight recovery of postural control and locomotion which has implications for quantifying the effects of weightlessness on gross motor control following gravity transitions (SM-101) [102]. While the current research evaluated AR as a tool for assessing sensorimotor performance, future research is required to capture preflight, in-flight, and postflight data at distinct time points with the AR tool to define magnitudes of performance change due to microgravity exposure and the time course of recovery. Analog environments such as parabolic flight studies or task performance following exposure to a centrifuge, could provide additional knowledge of the impact of various gravity levels and vestibular disorientation on task performance before in-flight testing.

The SM-102 gap similarly identifies sensorimotor risk and the need to "define the magnitude of change and time course of recovery, and how these impact critical mission tasks (e.g., define the risk for asking crew to pilot a rover following landing). Define how the risk (likelihood and severity) varies as a function of microgravity transit time, magnitude of G transition, and other contributing factors (prior experience, biomarkers, space radiation)" The current research also exhibited a means to assess postflight recovery of hand-eye coordination performance with the benefit of embedded sensors and onboard computing which has implications for fine motor control and manual crew control identified in the SM-102 gap [99]. The viability of augmented reality as a post-planetary-landing testing tool was evaluated through the comparative analyses of gold-standard Earth-based systems. Although target acquisition performance in AR was decreased across all dependent measures compared to touchscreen use, the sensorimotor hand-eye coordination study demonstrated that AR could identify changes in performance in rapid aimed movements when the difficulty of the task was modulated using different target widths and distances. Additional research has highlighted the capability of AR for detecting changes in assessment performance for the same hand-eye coordination task during vestibular-induced disorientation using Galvanic Vestibular Stimulation [189]. The research supports the use of AR as a tool for assessing fine motor control and informing manual crew control to support the SM-102 gap. Additional assessments in AR can be developed to simulate a landing procedure to further evaluate sensorimotor performance as it relates to critical mission tasks such as manual landings.

The current research focused on the evaluation of AR as an assessment tool; however, the research may extend into evaluating the use of AR as a countermeasure by promoting gradual head movements during readaptation. Although the research did not aim to quantify the effect of sensorimotor decrements on spatial orientation and motion sickness, future capabilities could be integrated into the AURORA tool to capture perceptions of orientation with respect to gravitational vertical or help to document motion sickness through subjective ratings. Future work focused on AR as a countermeasure tool would support HRP gaps SM-201 [102], SM-202 [103], and SM-203 [104]. The SM-203 gaps relates to defining "the incidence of motion sickness following landing and how this impacts the crewmembers ability to accomplish critical tasks" [104]. The current work begins to inform aspects of HRP gaps with healthy users, however, additional research is required to understand the sensitivity of the tool for sensorimotor decrements preflight and postflight with crew members to inform gaps SM-101 [102], SM-102 [99], SM-103 [100].

7.5 Limitations, Future Work, and Recommendations

While the research explored several motion-based and time-based measures for rapid aimed-arm movements and functional balance with a sample of healthy participants, future studies should examine these findings across larger population samples and sensorimotor impaired populations such as recently returned astronauts. The primary objective of the research was to initially assess the feasibility of using augmented reality in conducting evaluations with healthy individuals, with the aim of gathering insights that would guide future assessments involving individuals with sensorimotor alterations. Ultimately, the performance of users in scenarios involving vestibular and sensorimotor disorientation is likely to differ in terms of functional task execution and hand-eye coordination. Subsequent research should focus on understanding how augmented reality influences these circumstances.

Although all users received a training period before participating in the experimental studies, and this training was considered adequate for the testing procedures, it is possible that extended use of augmented reality could lead to performance improvements, potentially affecting the outcomes. Initial findings from the research suggest that augmented reality could serve as a valuable instrument for conducting longitudinal studies on performance. Consequently, future investigations should contemplate the effect of training duration within AR, the retention of AR skills after training, and the implementation of extended augmented reality usage regimens to monitor changes in performance.

The AR system was comprised of a robust collection of assessments guided by clinical and operational insights. Although only a subset of the assessments within AURORA were formerly evaluated within the research, the remaining tasks within the static and operational balance and hand-eye assessment modules should be explored. Nonetheless, forthcoming research may also investigate alternative assessment tasks tailored to address precise research inquiries or facets of human performance, thereby contributing to a broader understanding of task preparedness and daily activity readiness informed by AR assessments. Extensions of the sensorimotor investigations could include the evaluations of random target acquisition and the extension of the interaction space into three dimensions. A study may examine task performance when the holographic target array is superimposed on a flat physical surface to assess the effects of tactile feedback. For functional balance and operational tasks, studies may explore alternative motion-based and time-based measures of performance. While visual cueing and leveraging AR capabilities to improve perceptual and interaction performance were used throughout usability development, studies may consider the role of assistive cues in improving AR task performance.

While the research captured eye-tracking data for the multidirectional tapping task, only the hand-tracking metrics were used to compare performance to a touchscreen device. Future research may consider traditional eye-tracking measures to inform user strategy within AR or to evaluate the vestibular-ocular reflex or nystagmus. Measures derived from the eye-tracking sensor stream may include fixation count, regressive fixation, fixation duration, saccade peak velocity, and time between fixations.

In order to assess the preparedness of the crew for EVA, further research is essential to establish performance benchmarks throughout the preflight and postflight recovery period from sensorimotor and vestibular disruptions. The research demonstrated that task execution using augmented reality in a healthy population took more time compared to conventional physical tasks. Consequently, conventional clinical time-based criteria for identifying vestibular impairments and elevated fall risk are not applicable. Additional research is needed to customize these thresholds specifically for astronauts, aiming to provide insights into inflight performance declines in comparison to preflight data. The research findings also highlighted decreases in head and torso angular velocities and accelerations, a common approach in astronauts to reduce postflight motion sickness symptoms. However, it's worth noting that the augmented reality environment introduced higher levels of head pitch, which could either aid in recovery or potentially worsen symptoms. Further research is necessary to evaluate the effects of augmented reality on this aspect of future use in spaceflight applications.

7.6 Concluding Remarks

This thesis initiated a novel evaluation of augmented reality for use in sensorimotor and neurovestibular assessments for spaceflight applications, aiming to assess astronaut performance to inform readiness for extravehicular activities. Despite the divergence in user performance when compared to physical object-based tasks, augmented reality remains a promising solution for inflight assessments, tailored to spaceflight constraints and enhanced by integrated sensors. Subsequent efforts are needed to establish performance thresholds and develop methodologies for interpreting and scoring user performance in specific assessment tasks, with the aim of providing valuable support for decision-making processes. As augmented reality technology advances and becomes more accessible, this tool holds the potential not only for spaceflight applications but also for in-home care, enabling self-administered assessments and training for individuals with vestibular impairments, aging populations, athletes, and military operations where the risk of traumatic brain injury results in similar impairments as crew members following spaceflight missions.

APPENDIX A

Human-in-the-loop Usability Questionnaire

The following presents the usability survey questionnaire utilized in Chapter 3. Responses were represented on a 7-point scale with anchors provided for each level. Free response questions were administered at the end of the survey to allow participants to comment on aspects of the application for future improvement.

A.0.1 Overall Reactions

- **Use of the AURORA application was...**
  (Extremely Easy, Moderately Easy, Slightly Easy, Neither Easy nor Difficult, Slightly Difficult, Moderately Difficult, Extremely Difficult)

- **I was ___ with the AURORA application as a whole.**
  (Extremely Satisfied, Moderately Satisfied, Slightly Satisfied, Neither Satisfied nor Dissatisfied, Slightly Dissatisfied, Moderately Dissatisfied, Extremely Dissatisfied)

- **My overall experience with the AURORA application was...**
  (Extremely Positive, Moderately Positive, Slightly Positive, Neither Positive nor Negative, Slightly Negative, Moderately Negative, Extremely Negative)

A.0.2 Learning/Perceived Ease of Use

- **Learning to operate the application was...**
  (Extremely Easy, Moderately Easy, Slightly Easy, Neither Easy nor Difficult, Slightly Difficult, Moderately Difficult, Extremely Difficult)

- **Learning the hand interaction gestures was...**
  (Extremely Easy, Moderately Easy, Slightly Easy, Neither Easy nor Difficult, Slightly Difficult, Moderately Difficult, Extremely Difficult)

- **Learning the voice commands was...**
  (Extremely Easy, Moderately Easy, Slightly Easy, Neither Easy nor Difficult, Slightly Difficult, Moderately Difficult, Extremely Difficult)

A.0.3 System Capabilities

- **The system speed was...**
  (Extremely Fast, Moderately Fast, Slightly Fast, Neither Fast nor Slow, Slightly Slow, Moderately Slow, Extremely Slow)
- **The system's navigation between assessments was...**
  (Extremely Reliable, Moderately Reliable, Slightly Reliable, Neither Reliable nor Unreliable, Slightly Unreliable, Moderately Unreliable, Extremely Unreliable)
- **The system's ability to recognize hand gestures was...**
  (Extremely Reliable, Moderately Reliable, Slightly Reliable, Neither Reliable nor Unreliable, Slightly Unreliable, Moderately Unreliable, Extremely Unreliable)
- **The system's ability to recognize voice commands was...**
  (Extremely Reliable, Moderately Reliable, Slightly Reliable, Neither Reliable nor Unreliable, Slightly Unreliable, Moderately Unreliable, Extremely Unreliable)
- **The system's ability to help correct mistakes or undo actions was...**
  (Extremely Easy, Moderately Easy, Slightly Easy, Neither Easy nor Difficult, Slightly Difficult, Moderately Difficult, Extremely Difficult)

A.0.4 Screen/UI/Terminology

- **The visibility of the colors chosen within the application were...**
  (Extremely Easy, Moderately Easy, Slightly Easy, Neither Easy nor Difficult, Slightly Difficult, Moderately Difficult, Extremely Difficult)
- **The audio sounds and instructions were...**
  (Extremely Supportive, Moderately Supportive, Slightly Supportive, Neither Supportive nor Distracting, Slightly Distracting, Moderately Distracting, Extremely Distracting)
- **The text within the application was generally ___ to read.**
  (Extremely Easy, Moderately Easy, Slightly Easy, Neither Easy nor Difficult, Slightly Difficult, Moderately Difficult, Extremely Difficult)

- **The terms used throughout the system were...**
  (Extremely Straightforward, Moderately Straightforward, Slightly Straightforward, Neither Straightforward nor Confusing, Slightly Confusing, Moderately Confusing, Extremely Confusing)

- **The presentation of task instructions were...**
  (Extremely Straightforward, Moderately Straightforward, Slightly Straightforward, Neither Straightforward nor Confusing, Slightly Confusing, Moderately Confusing, Extremely Confusing)

A.0.5 Comfort

- **I experienced uncomfortable sensations (nausea, dizziness, etc...) while using the AR headset.**
  (Strongly Agree, Moderately Agree, Slightly Agree, Neither Agree nor Disagree, Slightly Disagree, Moderately Disagree, Strongly Disagree)

- **I experienced ___ neck/shoulder fatigue after wearing the headset for the duration of the study.**
  (No pain, Mild pain, Moderate pain, Severe pain, Very Severe pain)

- **Please indicate your level of mental effort to use the application.**
  (Extremely Low, Moderately Low, Slightly Low, Neither Low nor High, Slightly High, Moderately High, Extremely High)

- **Please indicate your level of physical effort to use the application.**
  (Extremely Low, Moderately Low, Slightly Low, Neither Low nor High, Slightly High, Moderately High, Extremely High)

A.0.6 Intentions to Use

- **In the future, I would ___ the system for assessing my physical capabilities in a clinical environment.**
  (Never use, Almost never use, Occasionally use, Sometimes use, Almost every time use, Always use)

- **The information provided was ___ if I were to self-administered the assessments in the future.**

(Extremely Sufficient, Moderately Sufficient, Slightly Sufficient, Neither Sufficient nor Insufficient, Slightly Insufficient, Moderately Insufficient, Extremely Insufficient)

- **I would ___ others to use the AR application.**
  (Strongly Recommend, Moderately Recommend, Slightly Recommend, Neither Recommend nor Discourage, Slightly Discourage, Moderately Discourage, Strongly Discourage)

A.0.7 Free Response

- **Were there any instances where the application did not perform as intended?**
- **If you could recommend any design changes what would they be?**
- **Do you have any comments on the selected gestures and voice activation?**
- **Please provide any other general comments or suggestions.**

APPENDIX B

Human-in-the-loop Usability Questionnaire Supplemental Results

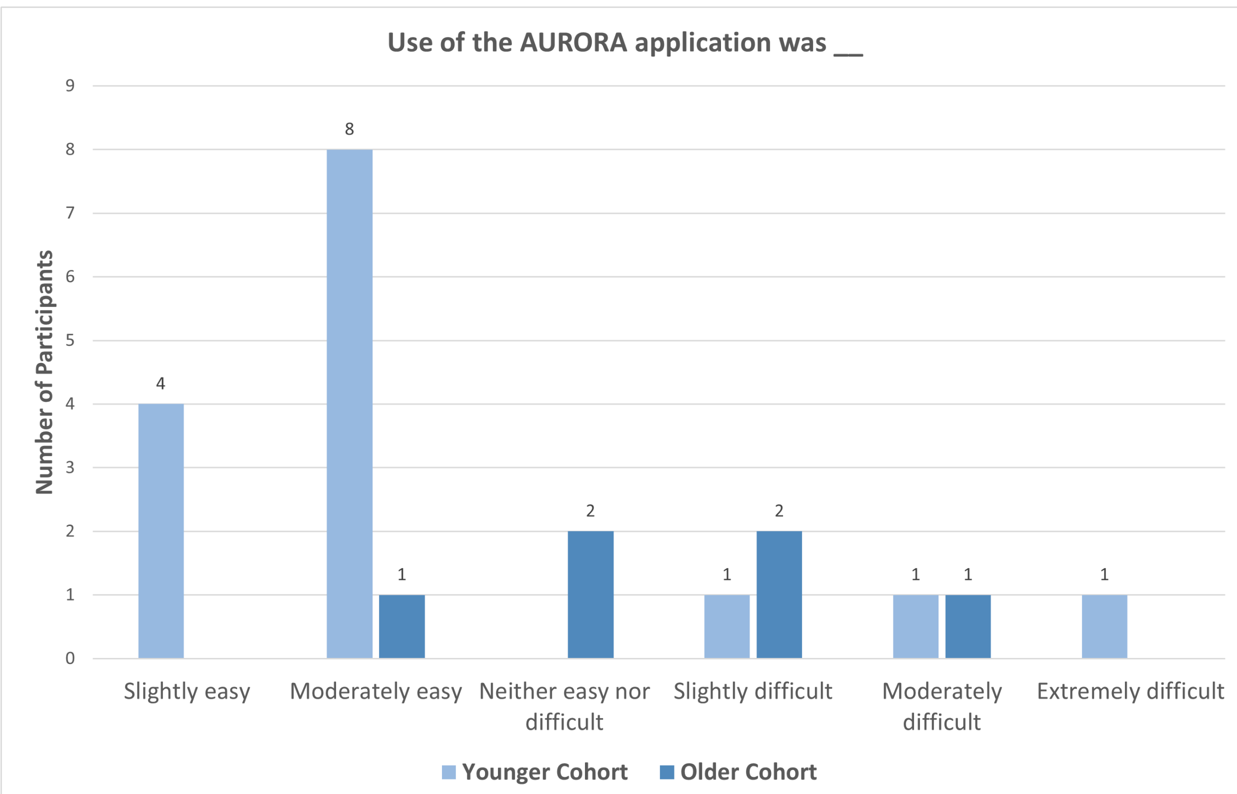

Figure B.1: Perceptions of the ease of use of the AURORA application.

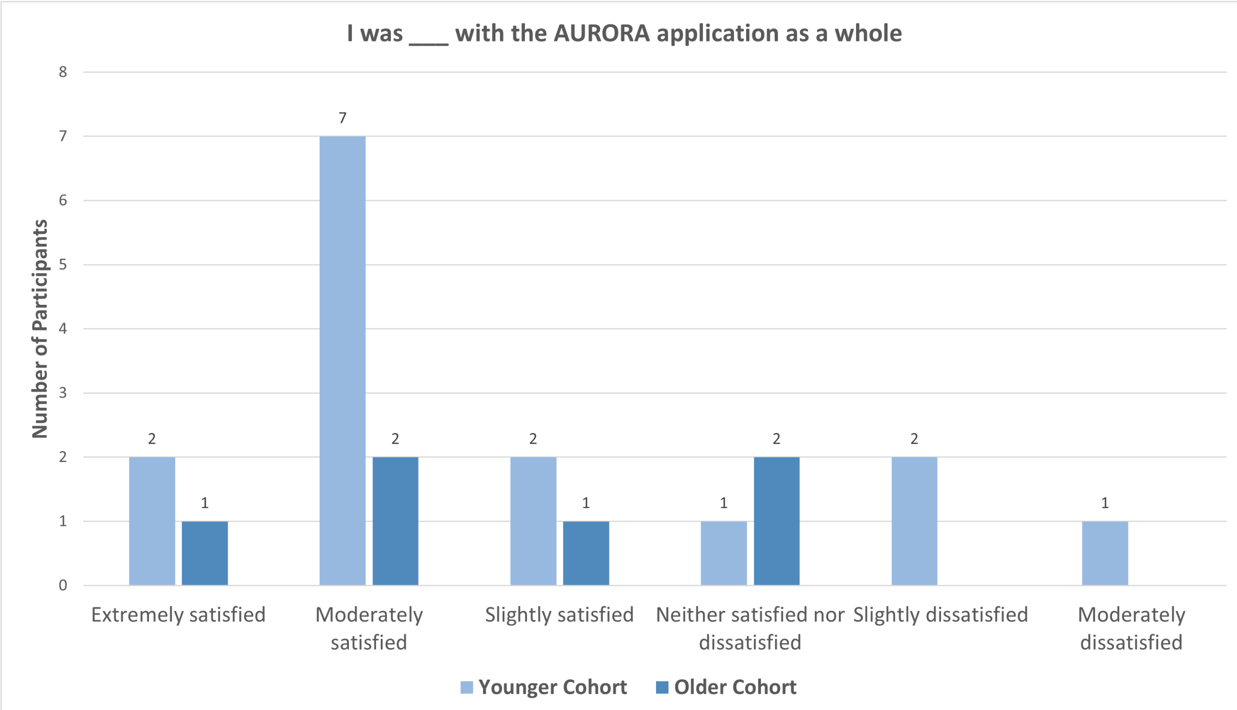

Figure B.2: Perceptions of user satisfaction with the AURORA application.

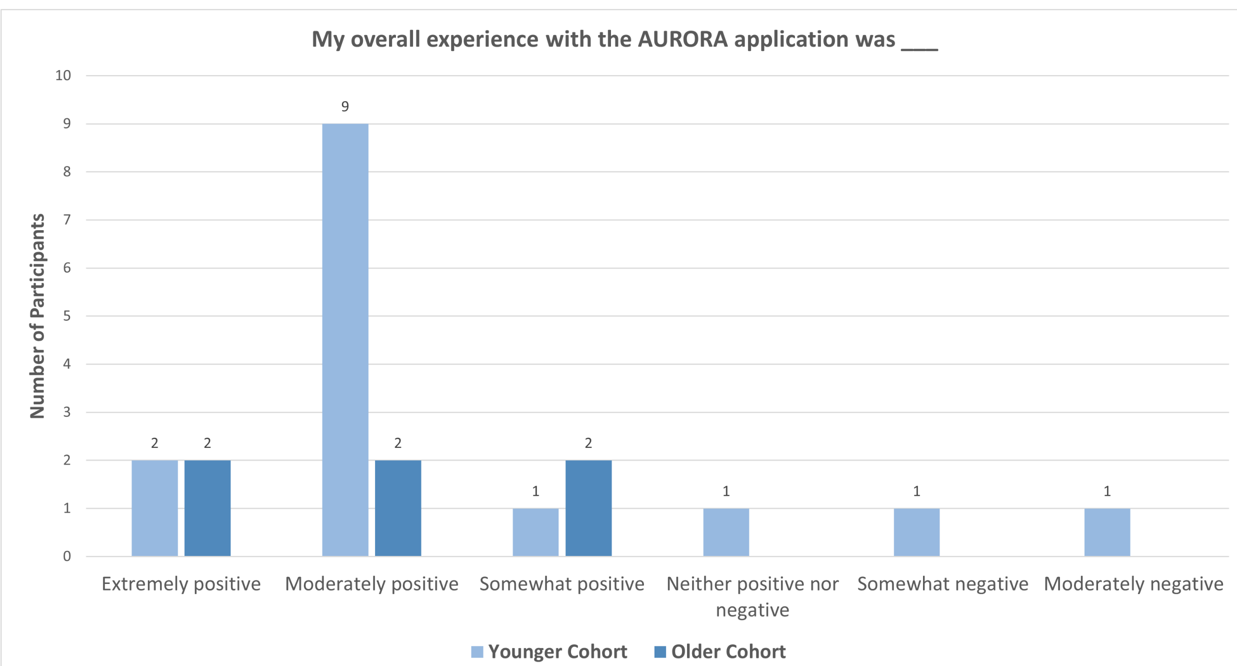

Figure B.3: Perceptions of the overall experience using the AURORA application.

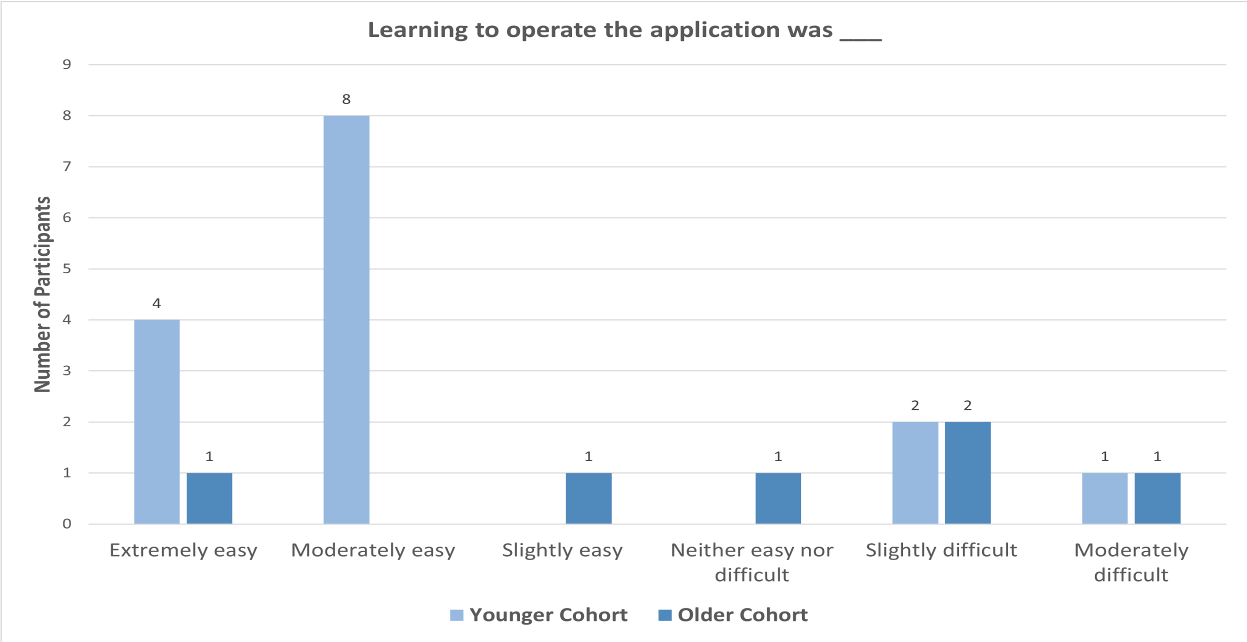

Figure B.4: Perceptions of learning to operate the AURORA application.

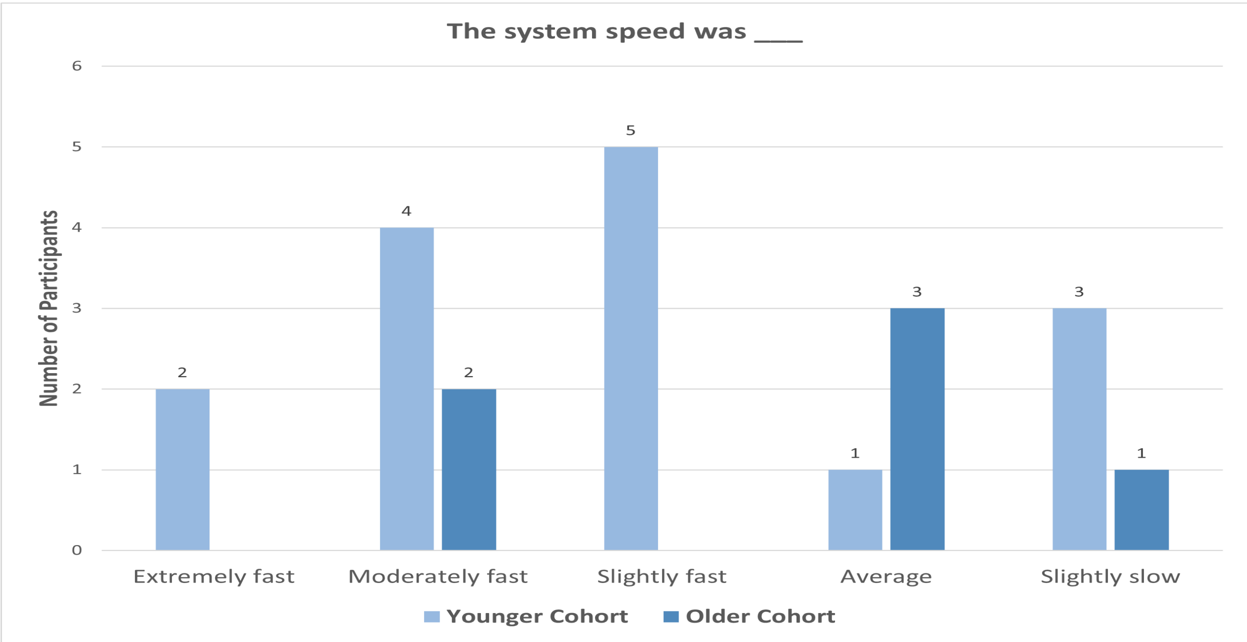

Figure B.5: Perceptions of system speed.

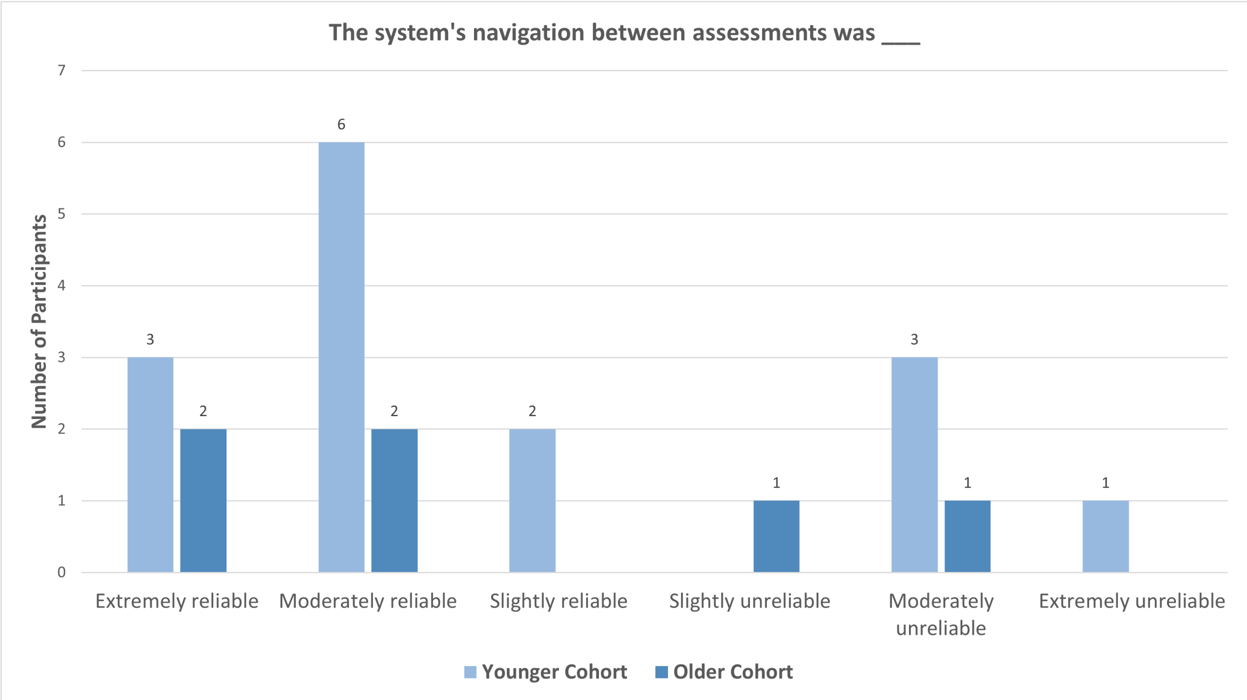

Figure B.6: Perceptions of the reliability of system navigation between assessments.

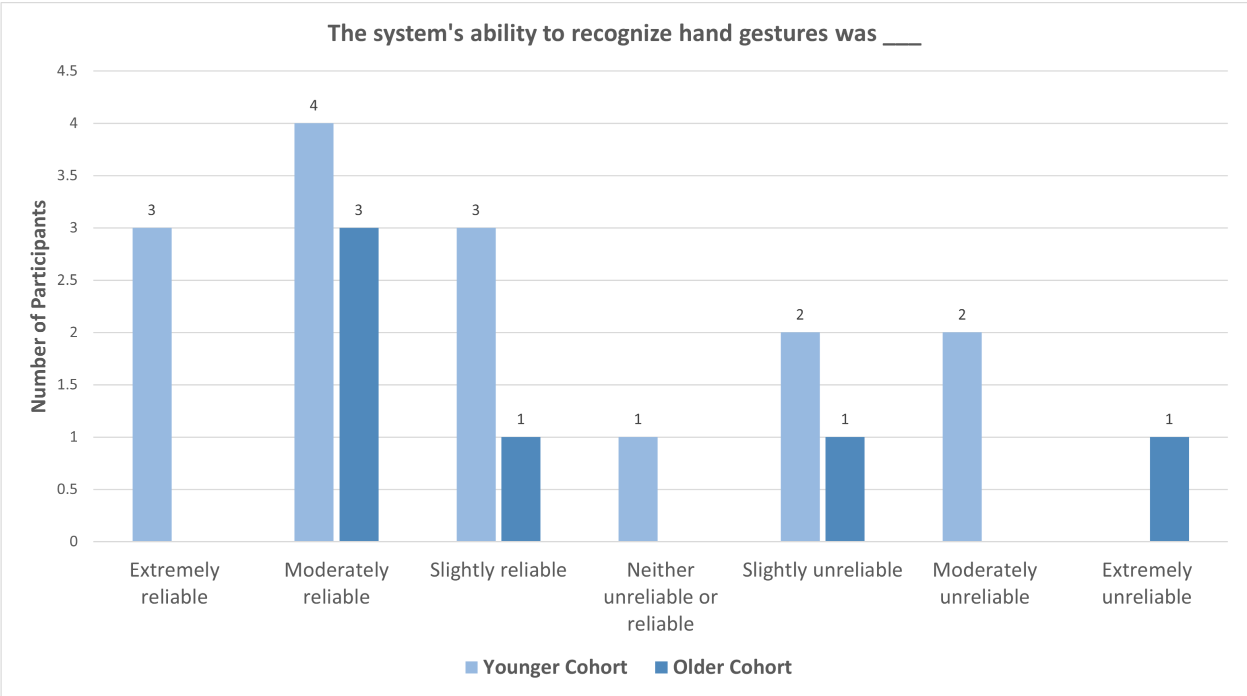

Figure B.7: Perceptions of the system's ability to recognize user input via the hand gestures.

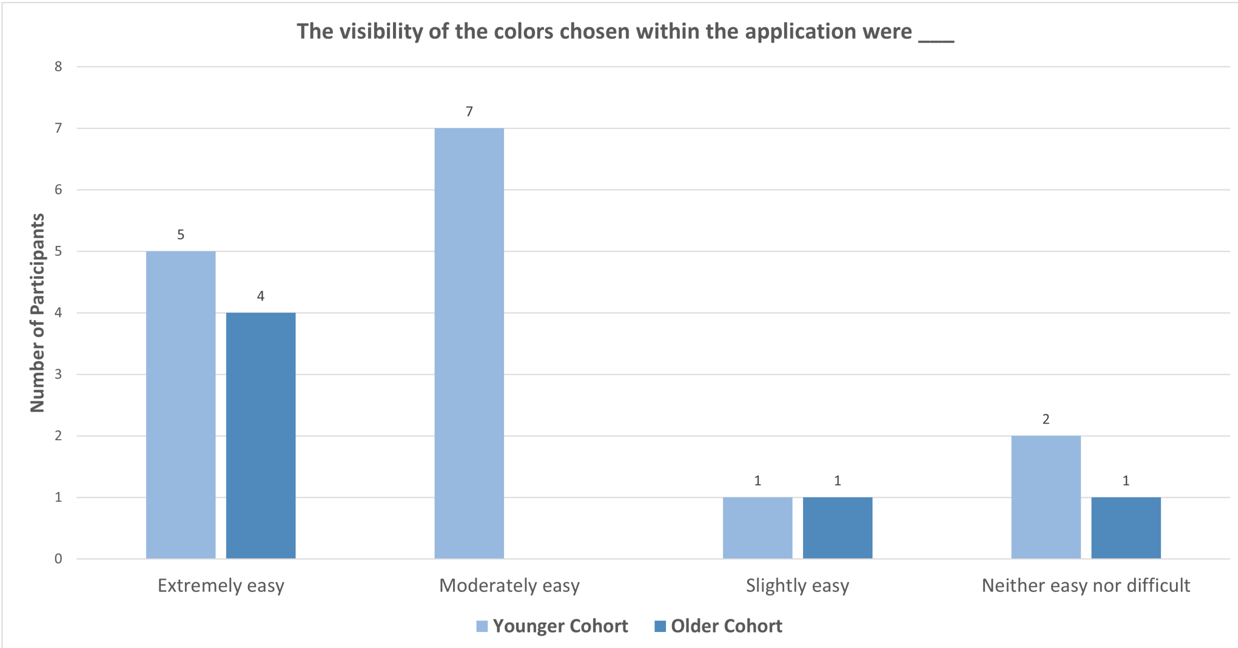

Figure B.8: Perceptions of the visibility of holograms with the chosen application color scheme.

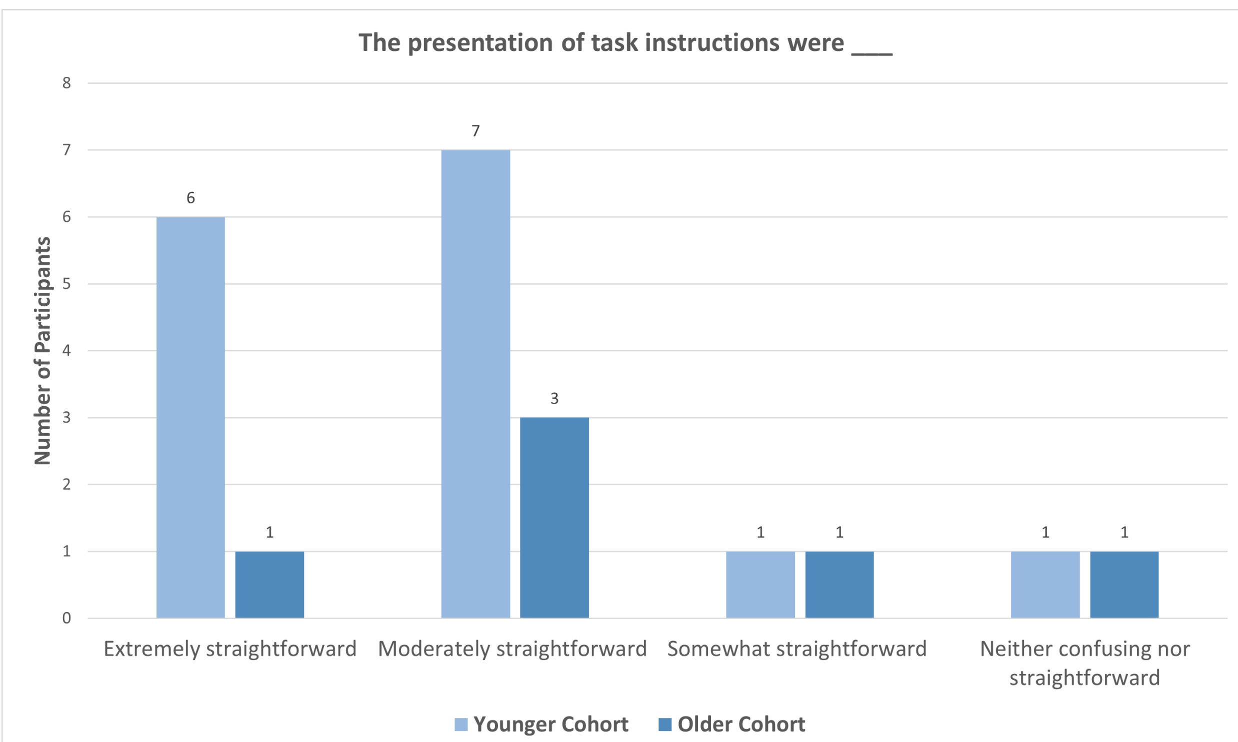

Figure B.9: Perceptions of the supportiveness of the audio sounds and instructions.

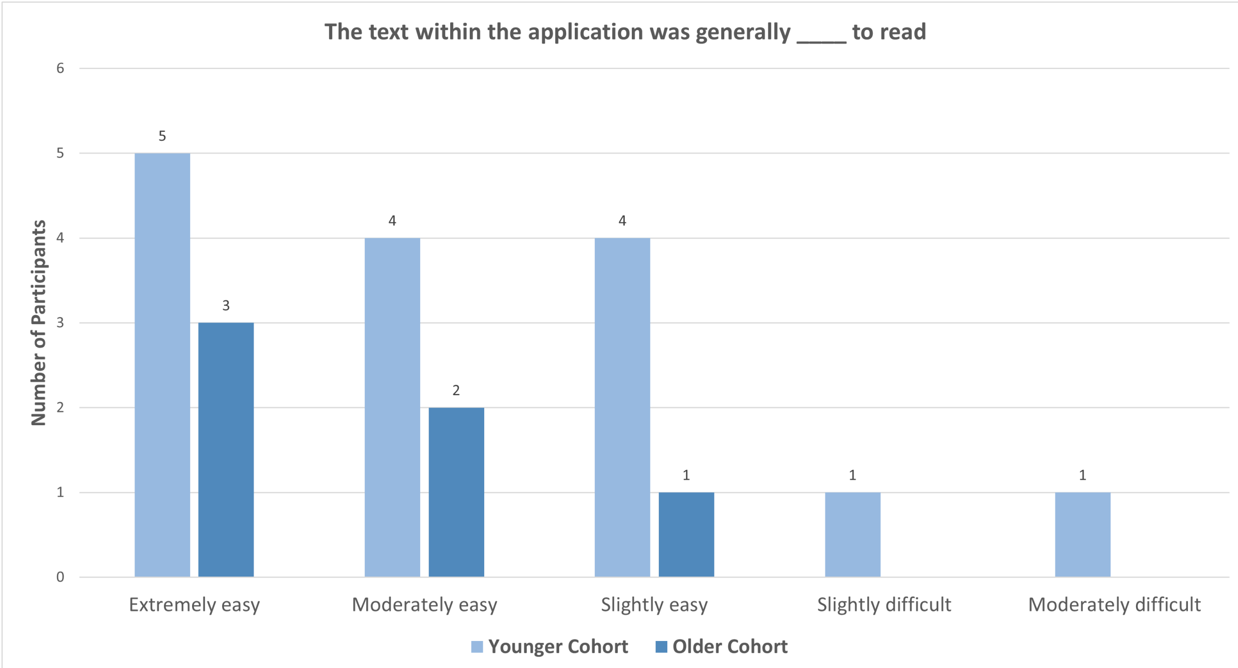

Figure B.10: Perceptions of the readability of written content in the AURORA application.

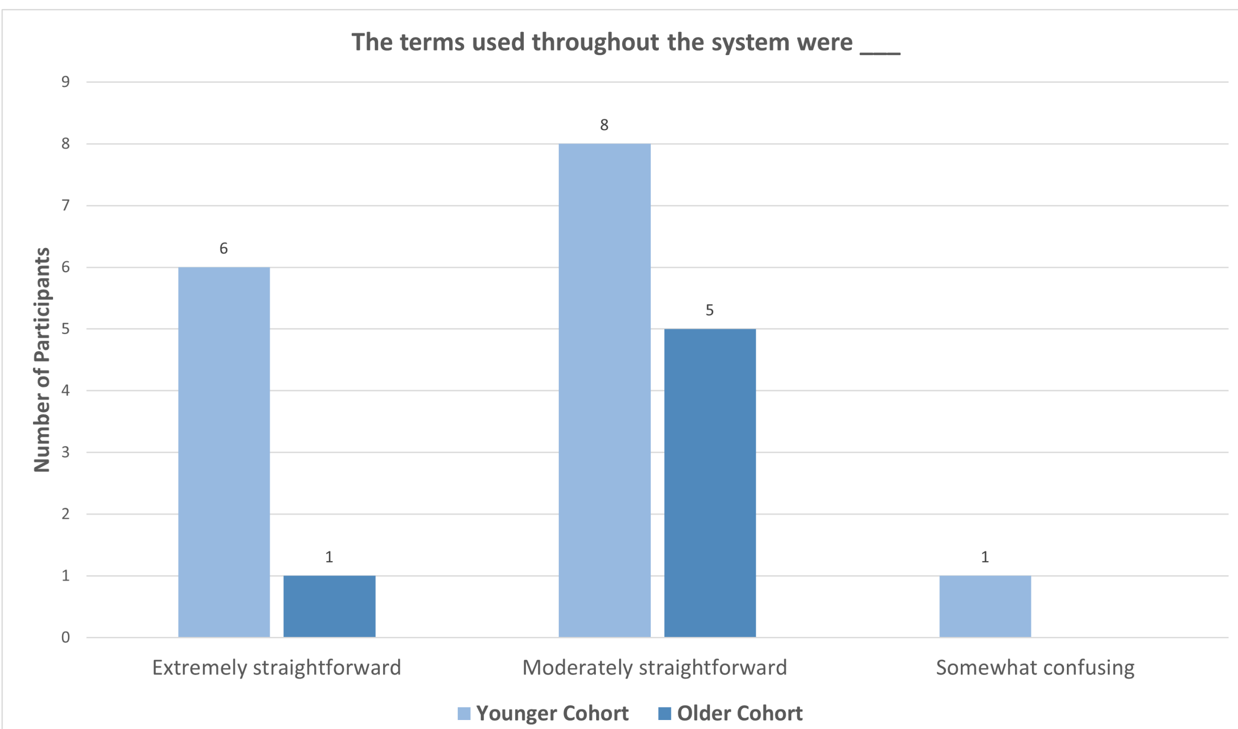

Figure B.11: Perceptions of the terminology used throughout the AURORA application.

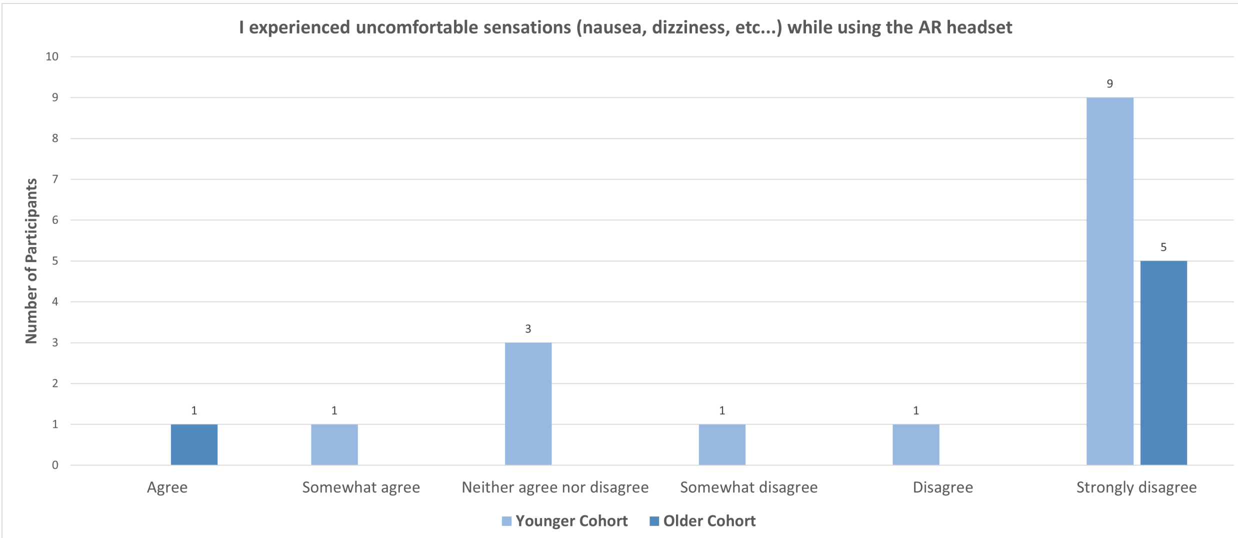

Figure B.12: Perceptions of discomfort while using the AURORA application.

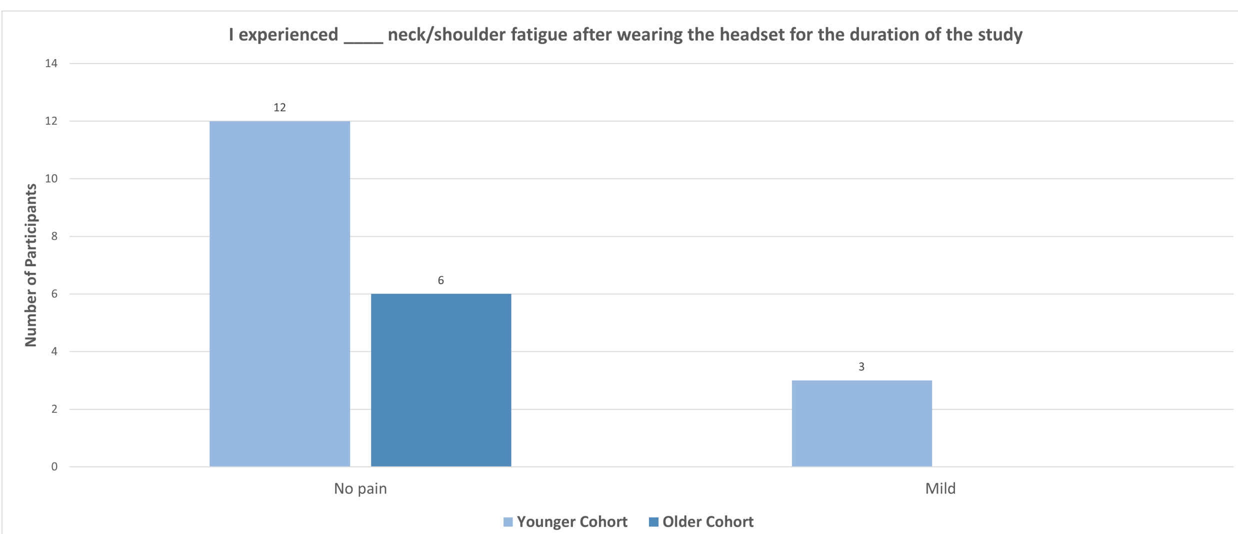

Figure B.13: Perceptions of neck fatigue while using the AURORA application.

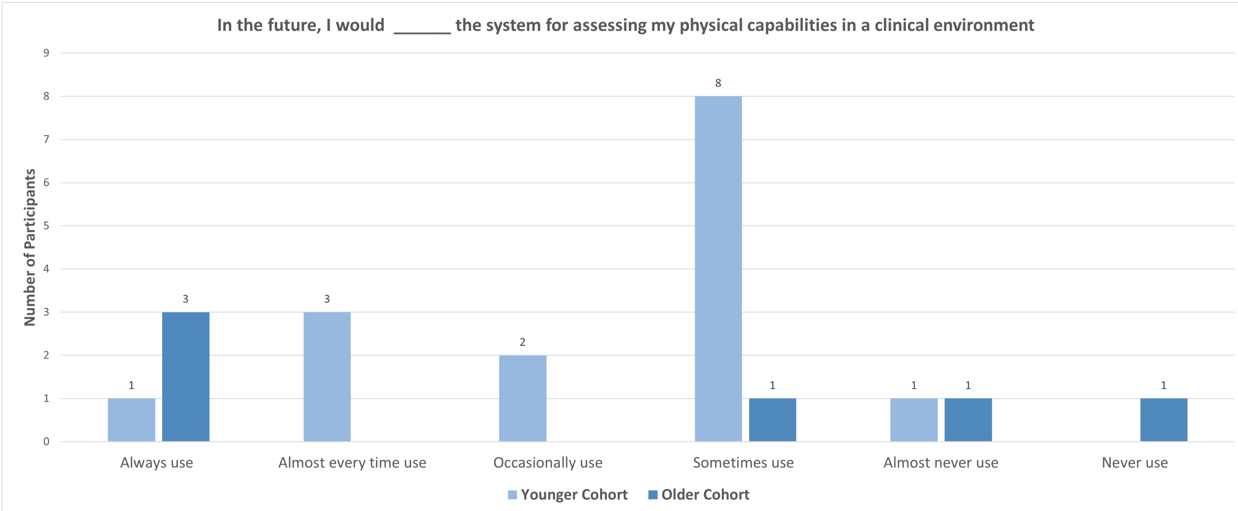

Figure B.14: Participant intention of future use if the system were available.

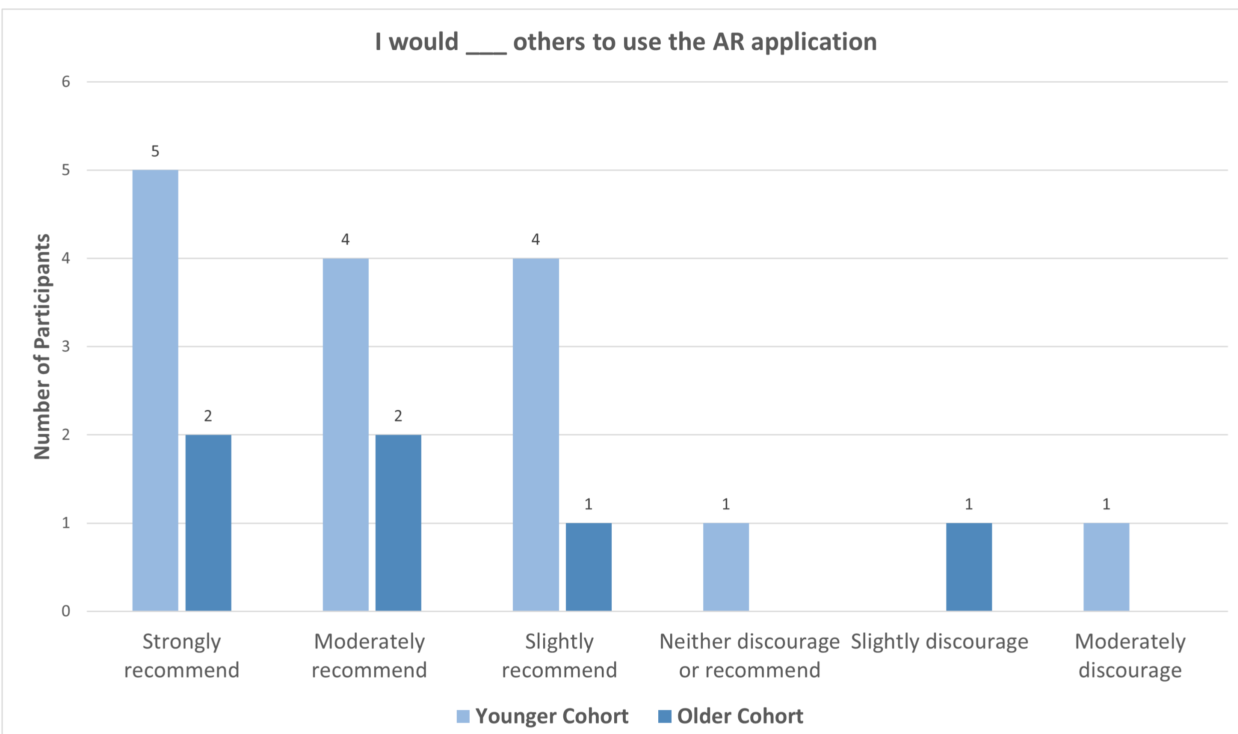

Figure B.15: Perceptions of the degree to which participants would recommend the AURORA system to others.

APPENDIX C

Sensorimotor Validation Study ANOVA Results

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Index of Difficulty	2652.2	1	2652.2	159.35	< 0.0001
Modality	22027.1	1	22027.1	1323.48	< 0.0001
Participant(Task Order)	2462.4	27	91.2	5.48	< 0.0001
Task Order	157.2	3	52.4	0.57	0.6366
Index of Difficulty * Modality	50.3	1	50.3	3.02	0.0859
Index of Difficulty * Task Order	14	3	4.7	0.28	0.8392
Modality * Task Order	151.5	3	50.5	3.03	0.0339
Index of Difficulty * Modality * Task Order	43.5	3	14.5	0.87	0.4597
Error	1348.1	81	16.6		
Total	29134.4	123			

Table C.1: The ANOVA model fit for the dependent variable of accuracy.

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Index of Difficulty	0.00001	1	1.19285e-05	8.91	0.0037
Modality	0.00007	1	7.2438e-05	54.12	< 0.0001
Participant(Task Order)	0.00005	27	1.95432e-06	1.46	0.0991
Task Order	0.00001	3	1.95246e-06	1	0.4083
Index of Difficulty * Modality	0.00001	1	1.25819e-05	9.4	0.0029
Index of Difficulty * Task Order	0	3	2.82585e-07	0.21	0.8884
Modality * Task Order	0.00001	3	1.94894e-06	1.46	0.2327
Index of Difficulty * Modality * Task Order	0	3	3.75706e-07	0.28	0.8392
Error	0.00011	81	1.33841e-06		
Total	0.00027	123			

Table C.2: The ANOVA model fit for the dependent variable of the overall convex area.

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Index of Difficulty	10.6009	1	10.6009	148.61	< 0.0001
Modality	57.4257	1	57.4257	805.04	< 0.0001
Participant(Task Order)	2.7653	27	0.1024	1.44	0.1093
Task Order	0.5234	3	0.1745	1.7	0.1899
Index of Difficulty * Modality	6.1534	1	6.1534	86.26	< 0.0001
Index of Difficulty * Task Order	0.4276	3	0.1425	2.16	0.1208
Modality * Task Order	0.3028	3	0.1009	1.41	0.2445
Index of Difficulty * Modality * Task Order	0.3128	3	0.1043	1.46	0.2312
Error	5.7779	81	0.0713		
Total	85.0908	123			

Table C.3: The ANOVA model fit for the dependent variable of Average movement time.

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Index of Difficulty	0.26016	1	0.26016	73.6	< 0.0001
Modality	2.59506	1	2.59506	734.14	< 0.0001
Participant(Task Order)	0.26416	27	0.00978	2.77	0.0002
Task Order	0.01044	3	0.00348	0.36	0.7853
Index of Difficulty * Modality	0.13215	1	0.13215	37.39	< 0.0001
Index of Difficulty * Task Order	0.01021	3	0.0034	0.96	0.4146
Modality * Task Order	0.01209	3	0.00403	1.14	0.3378
Index of Difficulty * Modality * Task Order	0.0165	3	0.0055	1.56	0.2065
Error	0.28632	81	0.00353		
Total	3.62301	123			

Table C.4: The ANOVA model fit for the dependent variable of the error rates.

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Index of Difficulty	26.1	1	26.104	55.69	< 0.0001
Modality	1356	1	13!55.996	2892.82	< 0.0001
Participant(Task Order)	34.04	27	1.261	2.69	0.0003
Task Order	15.82	3	5.273	4.18	0.0148
Index of Difficulty * Modality	2.39	1	2.39	5.1	0.0266
Index of Difficulty * Task Order	4.95	3	1.648	3.52	0.0188
Modality * Task Order	7.54	3	2.514	5.36	0.002
Index of Difficulty * Modality * Task Order	2.02	3	0.672	1.43	0.2389
Error	37.97	81	0.469		
Total	1486.99	123			

Table C.5: The ANOVA model fit for the dependent variable of throughput of tasks.

APPENDIX D

Dynamic Balance Assessments Comparative Analysis Supplemental Results

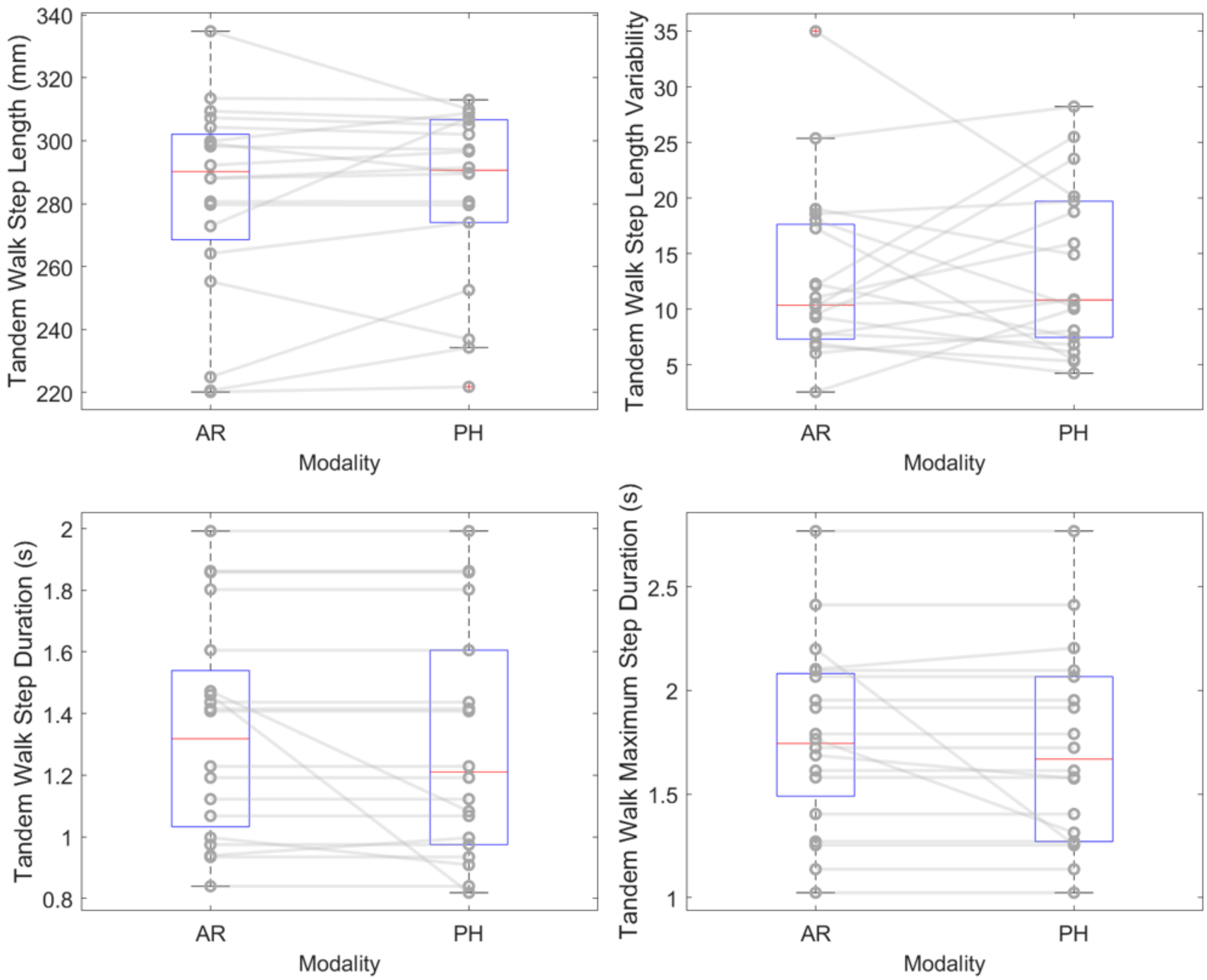

Figure D.1: Tandem Walk task-specific dependent measures.

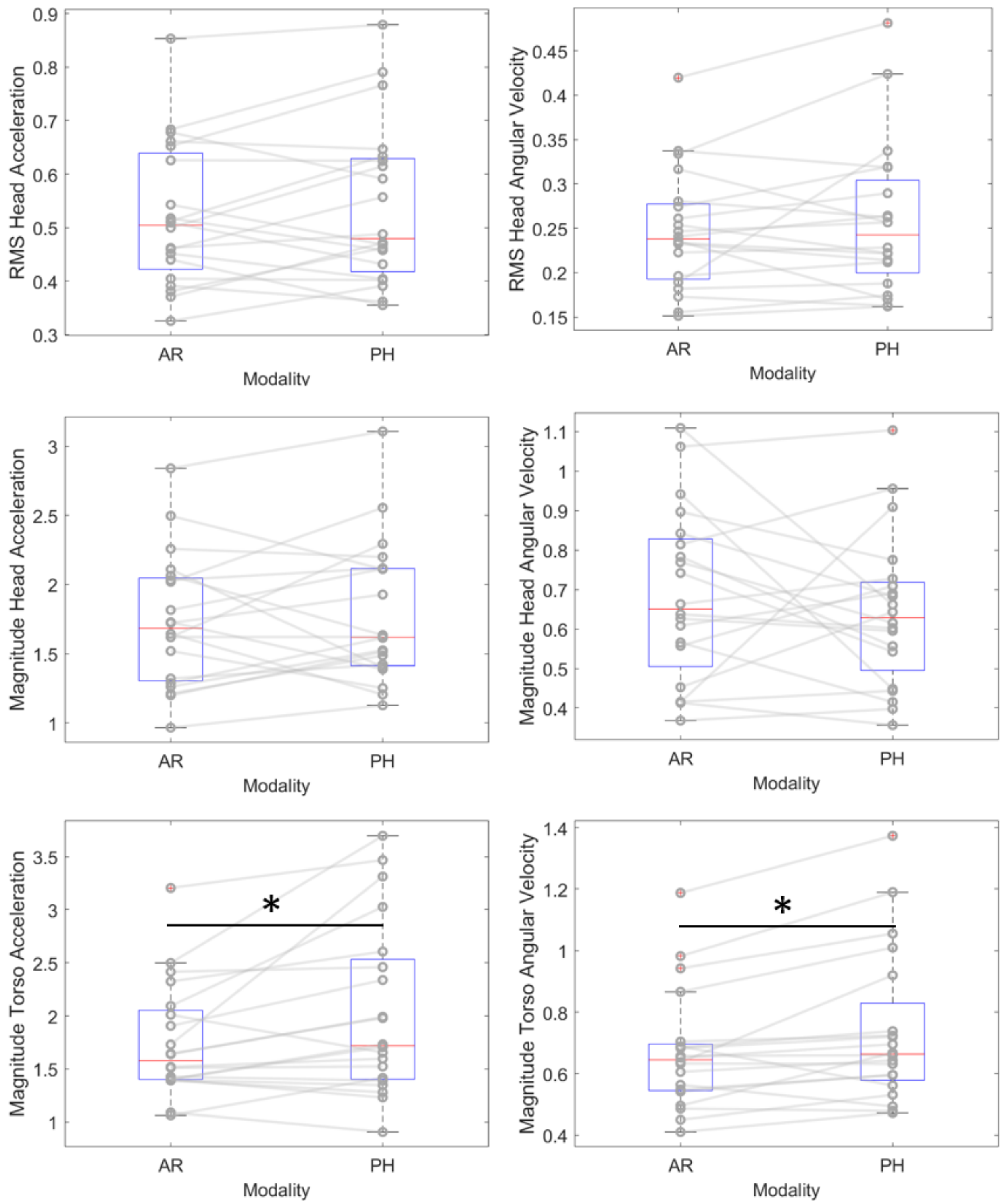

Figure D.2: Tandem walk inertial measurement unit dependent measures. Magnitude torso acceleration ($p = 0.017$). Magnitude torso angular velocity ($p = 0.006$).

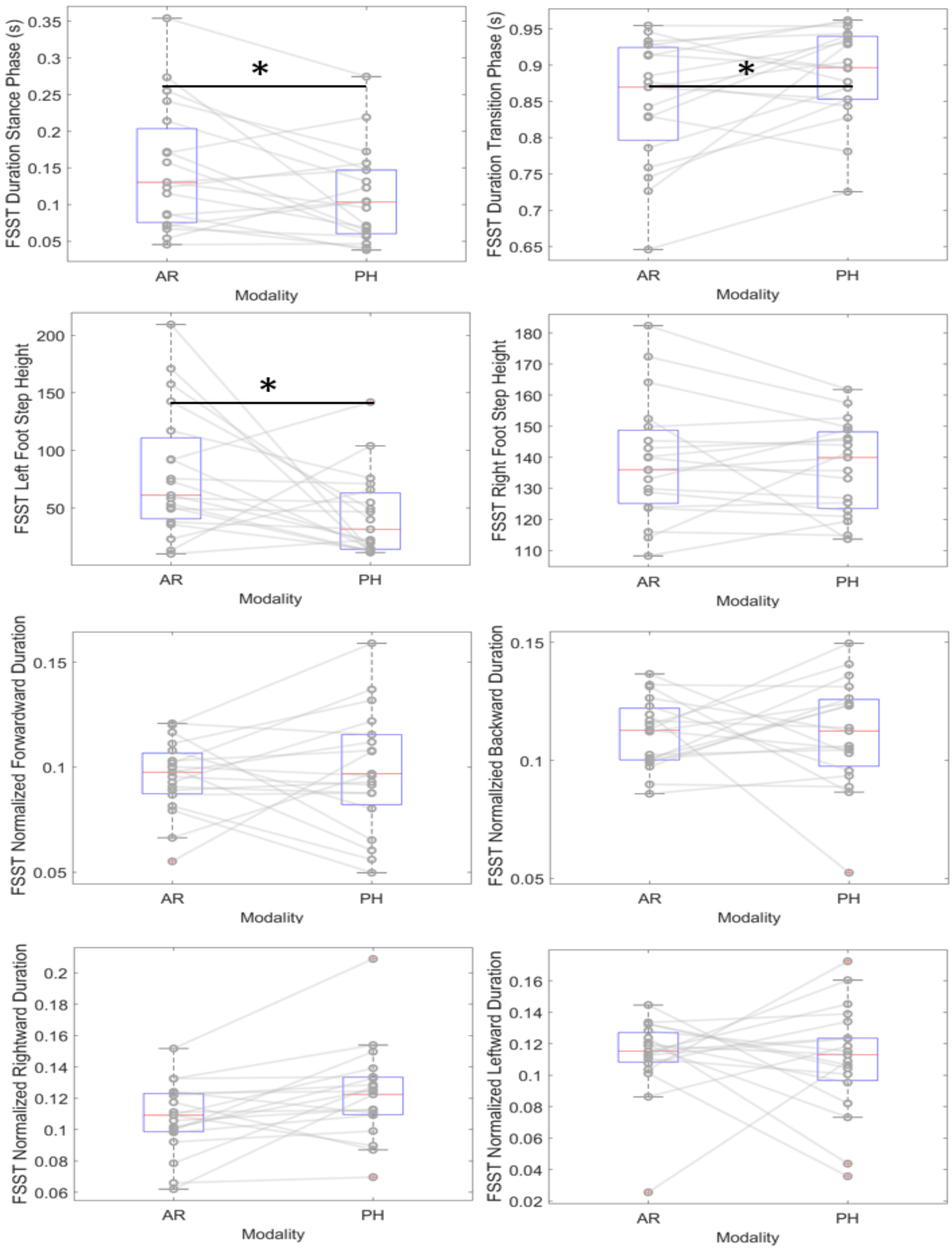

Figure D.3: Four Square Step Test task-specific dependent measures. Duration stance phase (p = 0.029). Duration transition phase (p = 0.029). Left foot step height (p = 0.035).

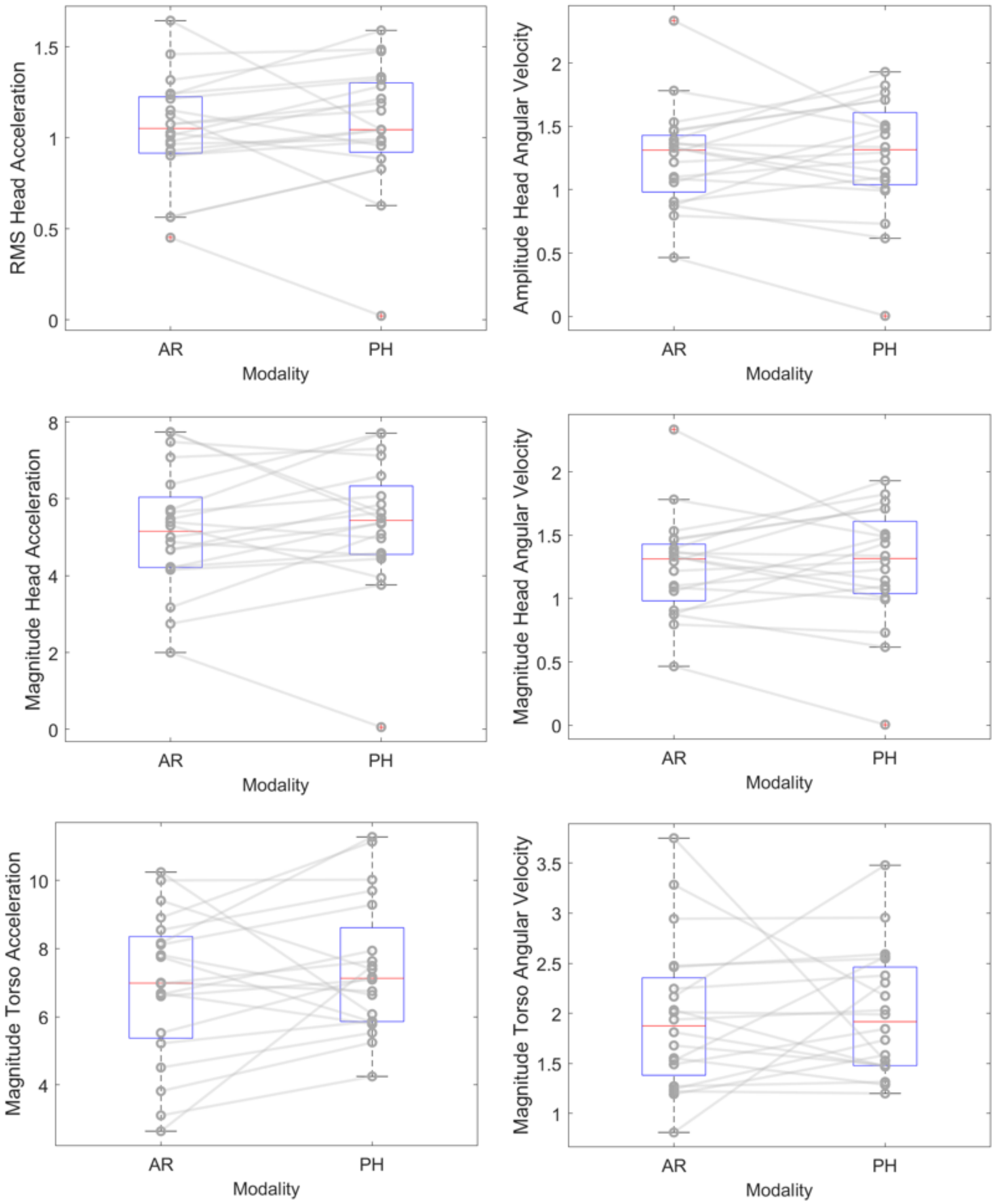

Figure D.4: Four Square Step Test inertial measurement unit dependent measures.

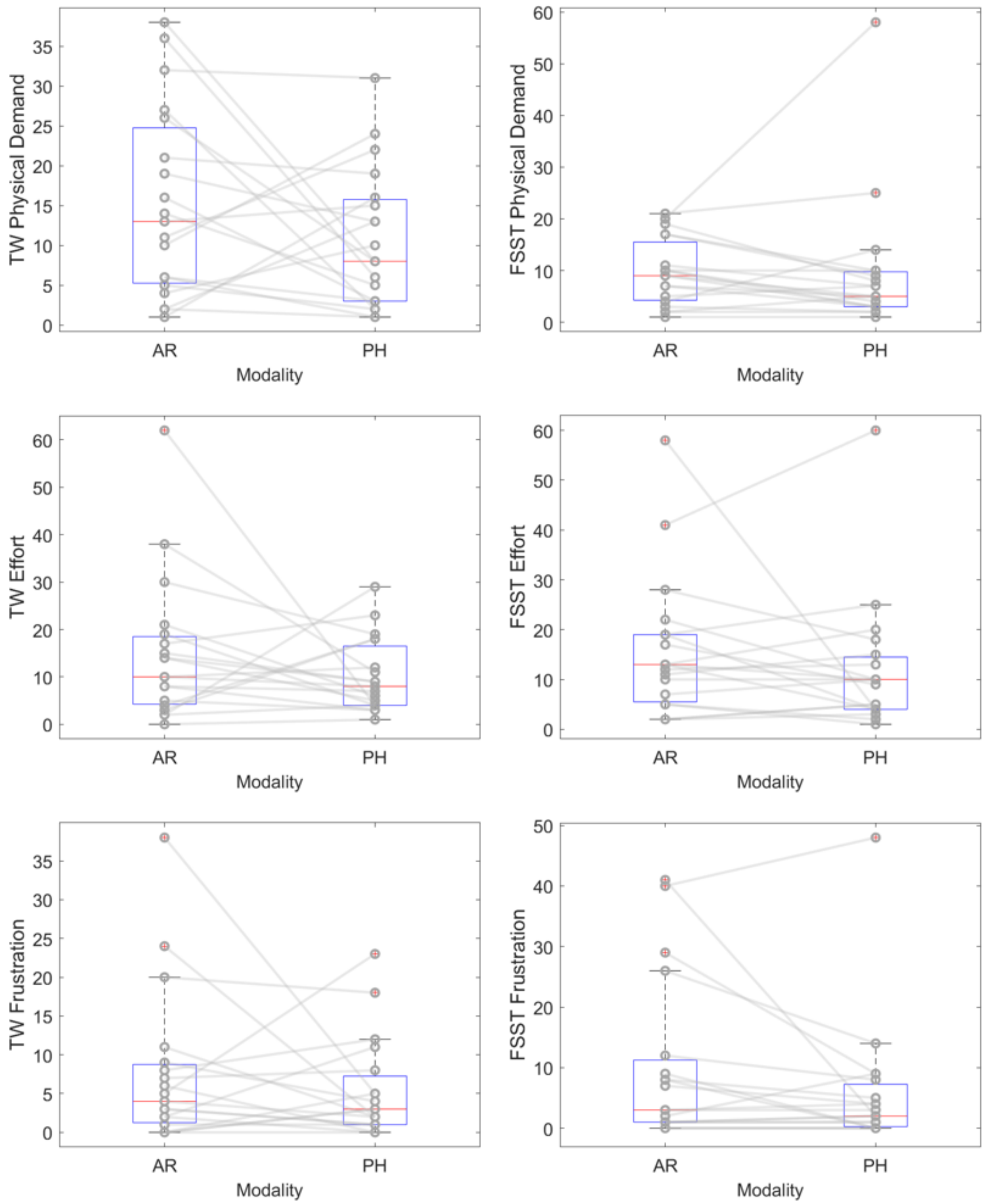

Figure D.5: NASA Task Load Index measures for the Tandem Walk and Four Square Step Test.

APPENDIX E

Operational Assessments Comparative Analysis Supplemental Results

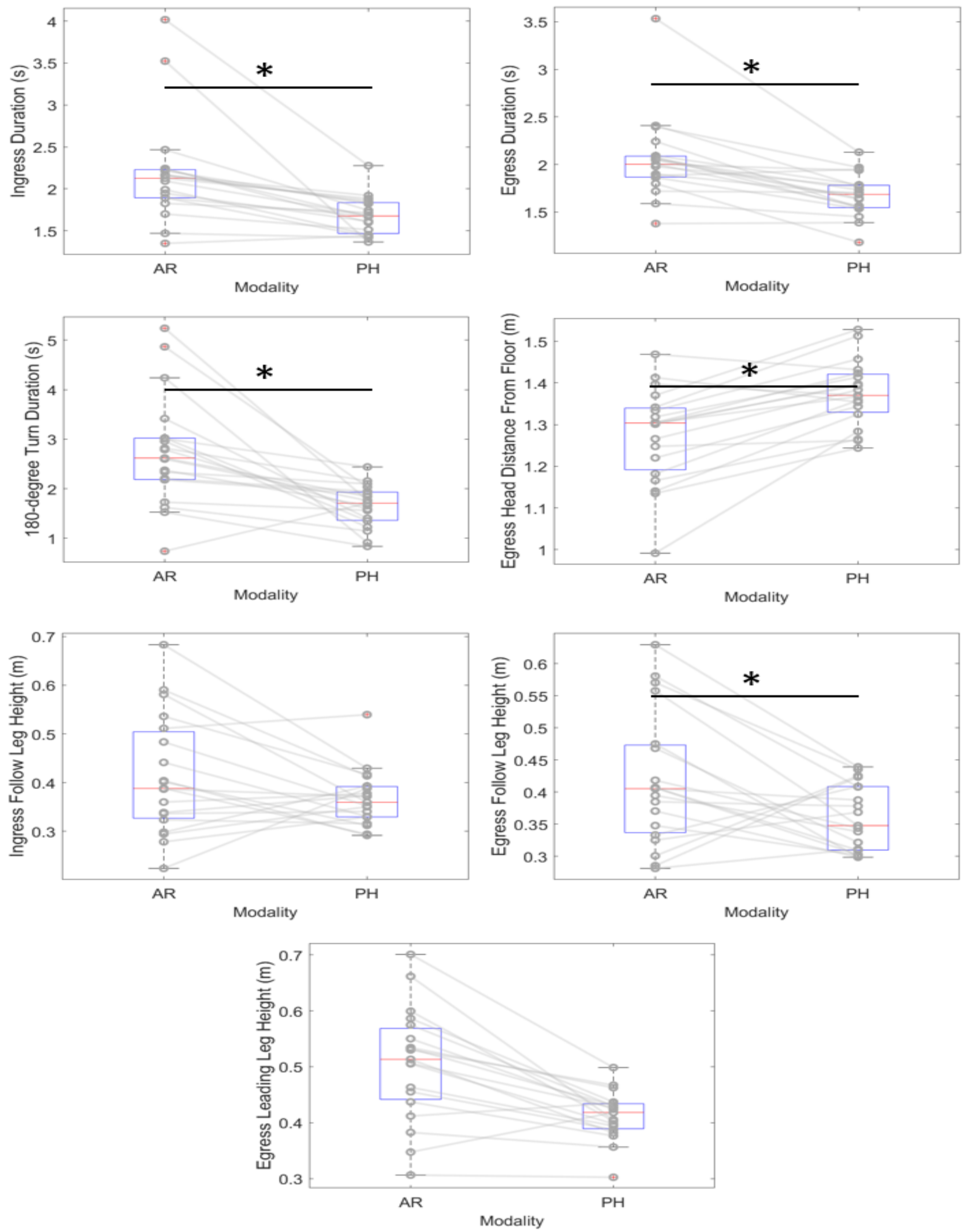

Figure E.1: Ingress and Egress Task dependent measures. Ingress Duration ($p < 0.001$). Egress duration ($p < 0.001$). 180-turn duration ($p < 0.001$). Egress head distance from floor ($p < 0.001$). Egress follow leg height (p= 0.032).

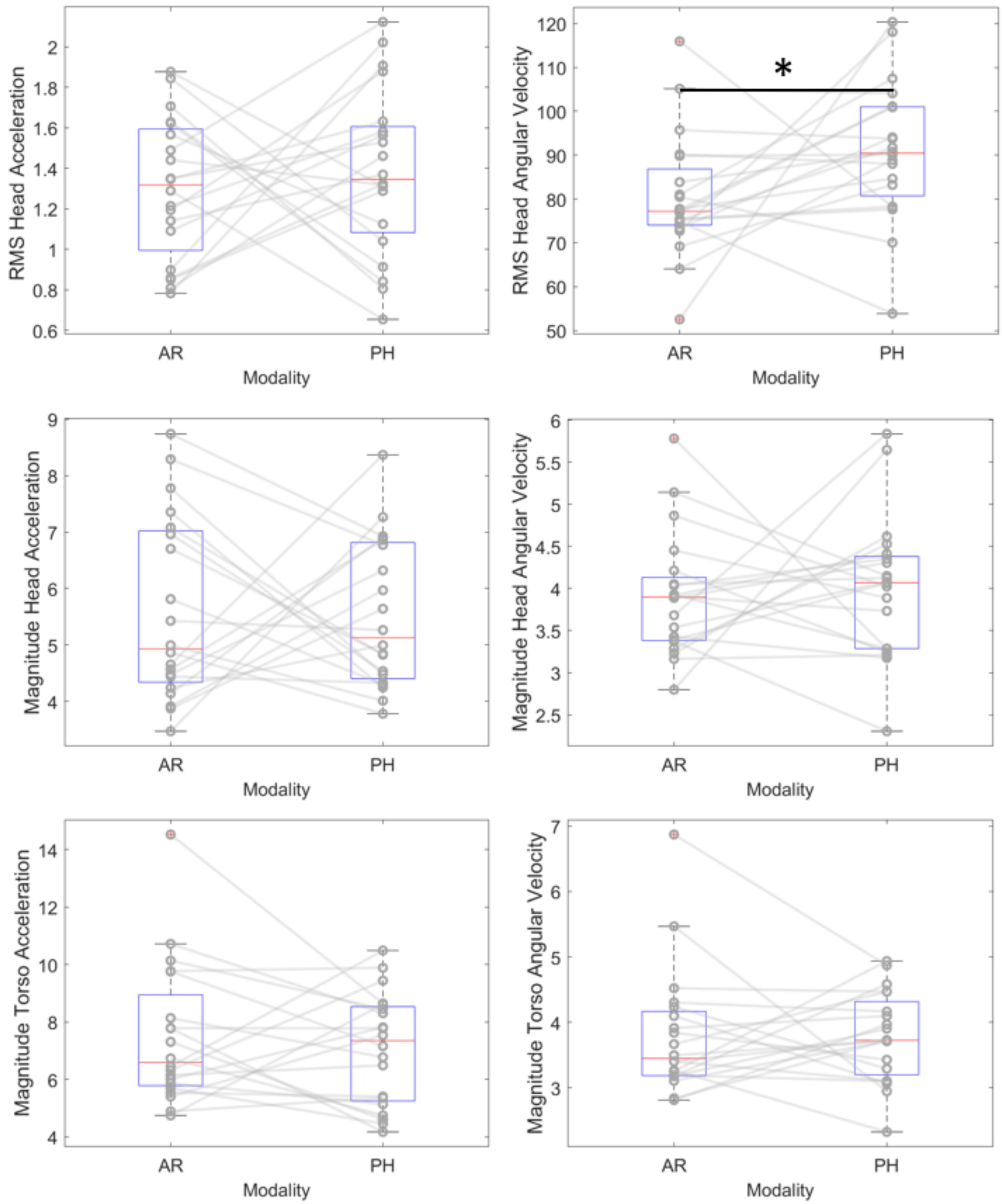

Figure E.2: Ingress and Egress Task inertial measurement unit dependent measures. RMS head angular velocity ($p = 0.011$).

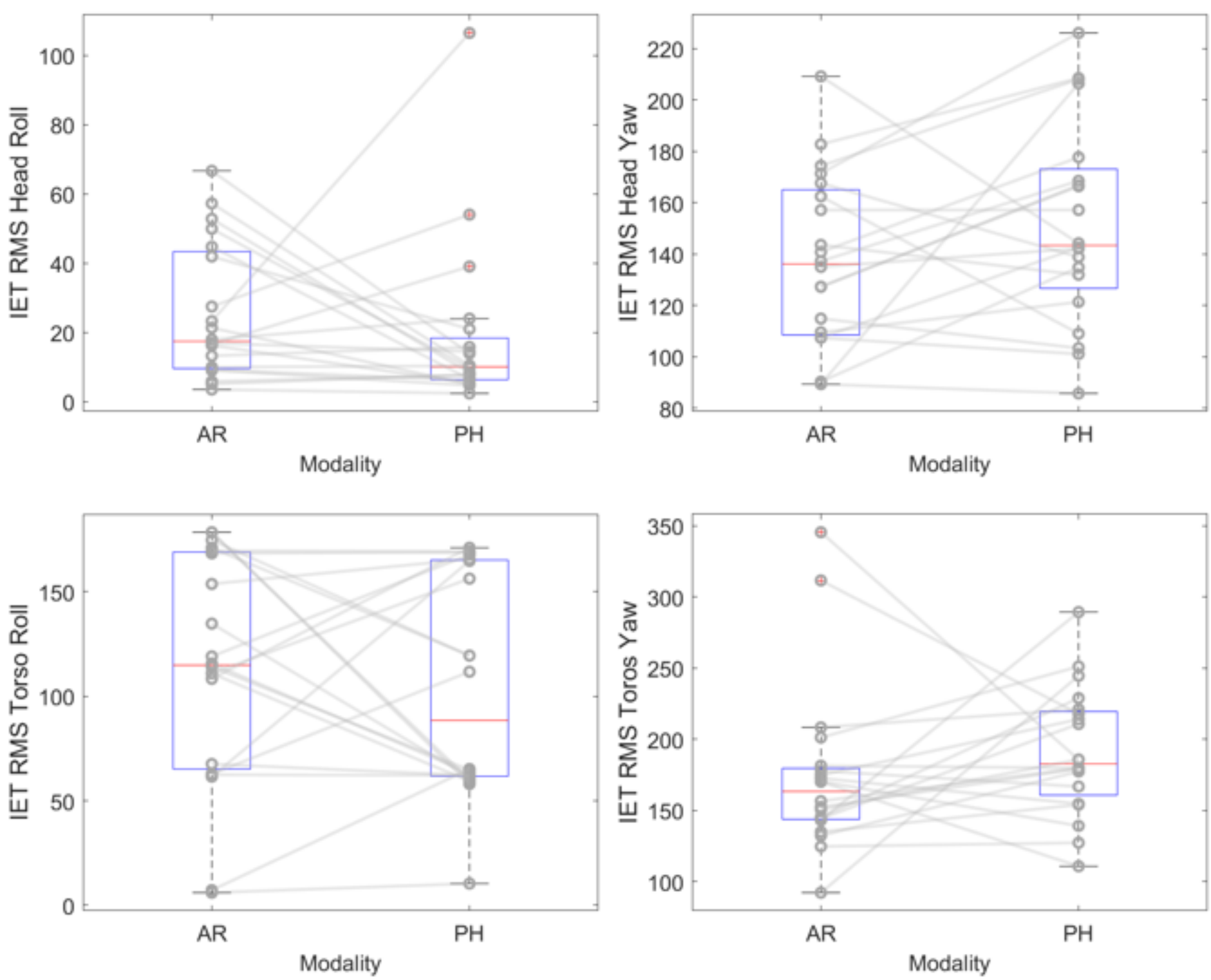

Figure E.3: Ingress and Egress Task orientation-based dependent measures.

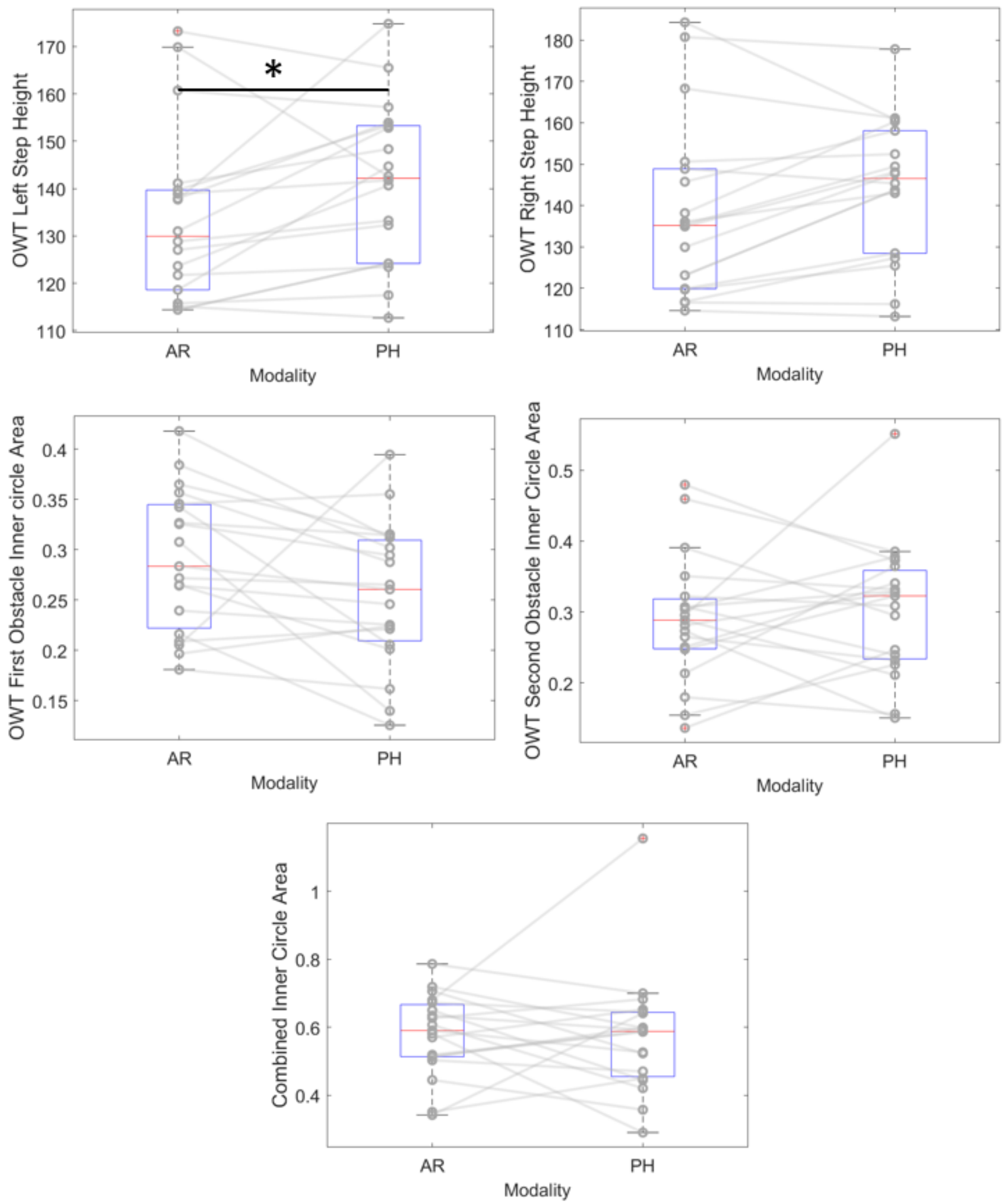

Figure E.4: Obstacle Weave task-specific dependent measures. Left step height ($p = 0.008$).

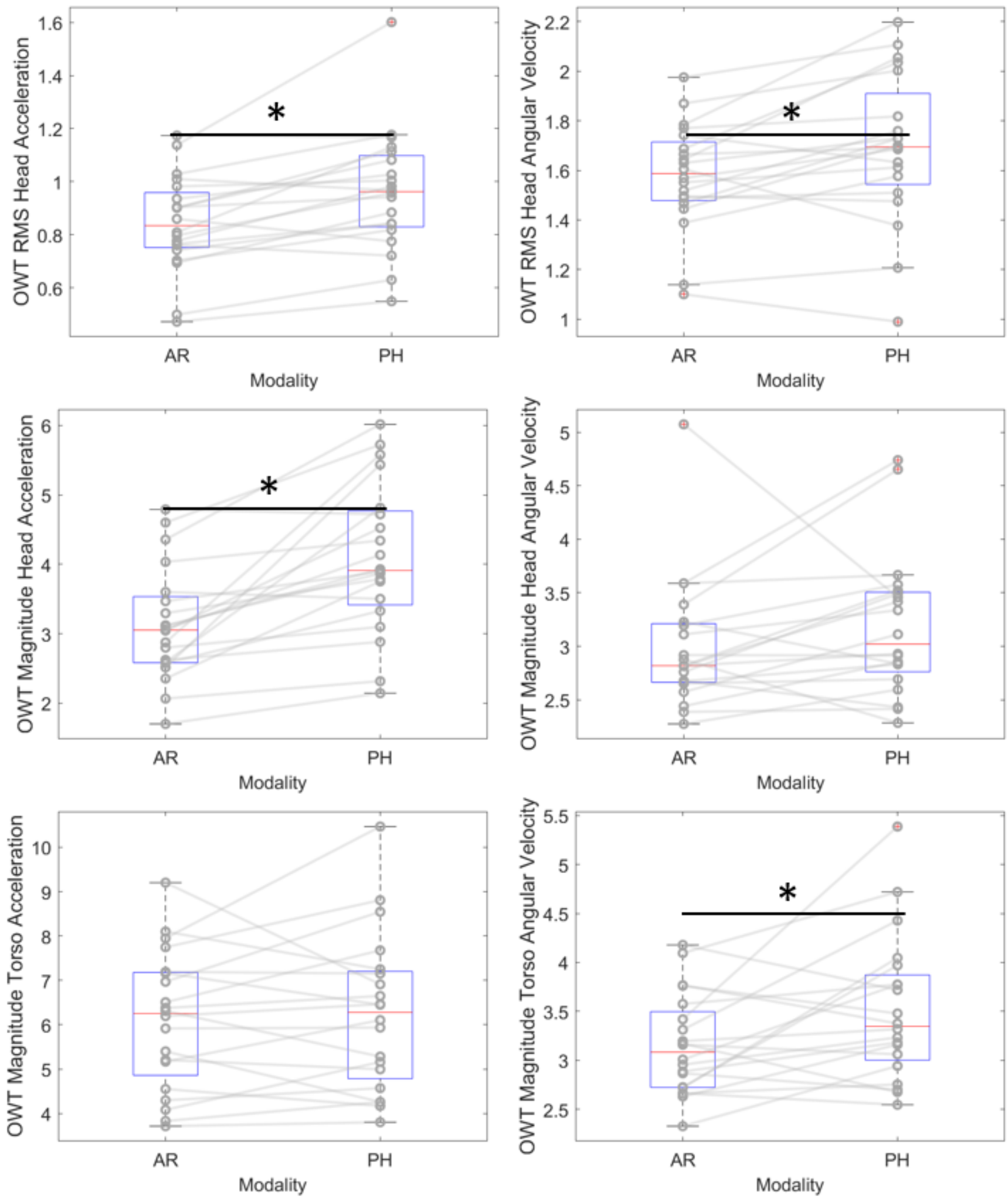

Figure E.5: Obstacle Weave inertial measurement unit dependent measures. RMS head acceleration ($p < 0.001$). RMS head angular velocity ($p = 0.006$). Magnitude head acceleration ($p < 0.001$). Magnitude torso angular velocity ($p = 0.027$).

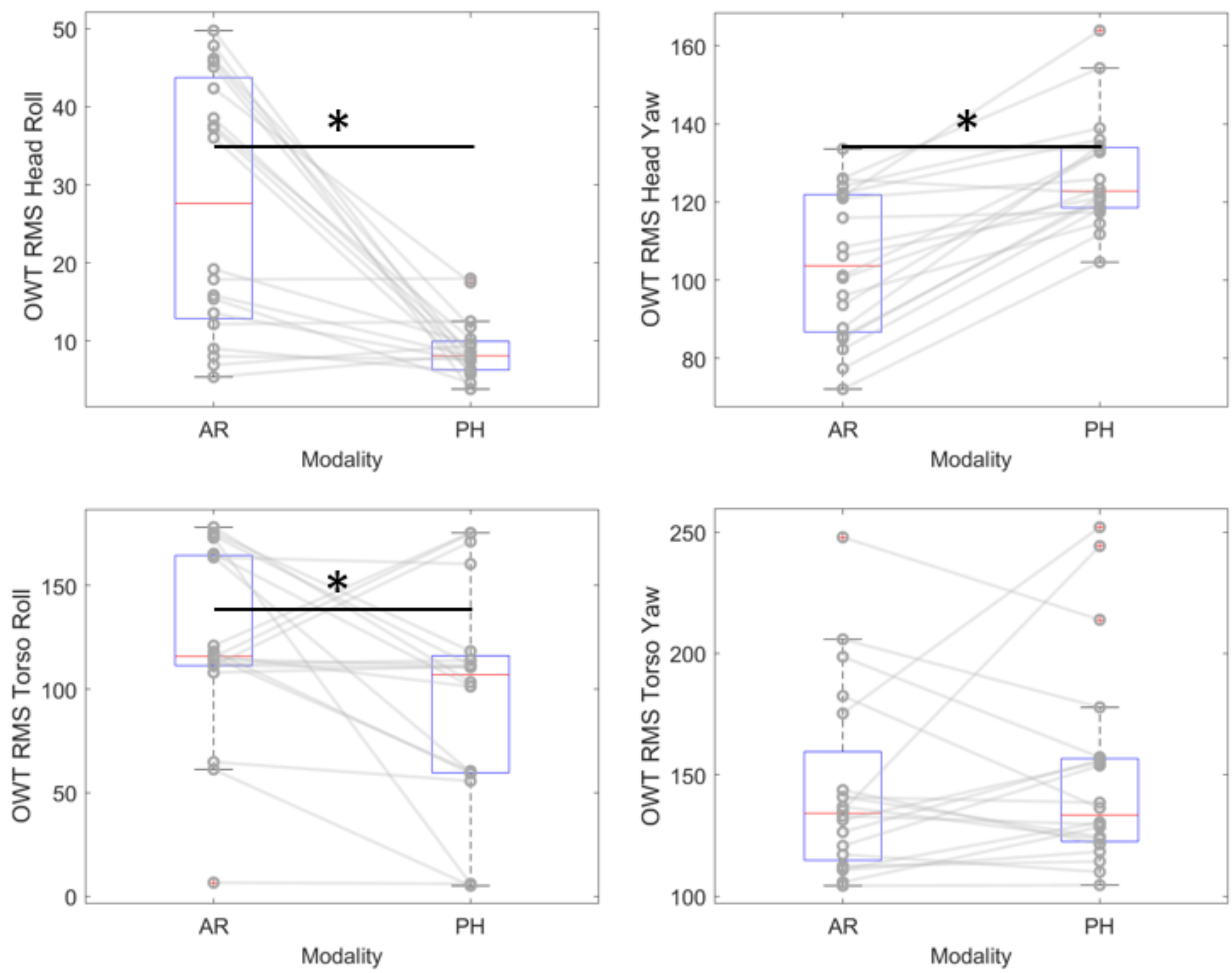

Figure E.6: Obstacle Weave orientation-based dependent measures. RMS head roll ($p < 0.001$). RMS head yaw ($p < 0.001$). RMS torso roll ($p = 0.041$).

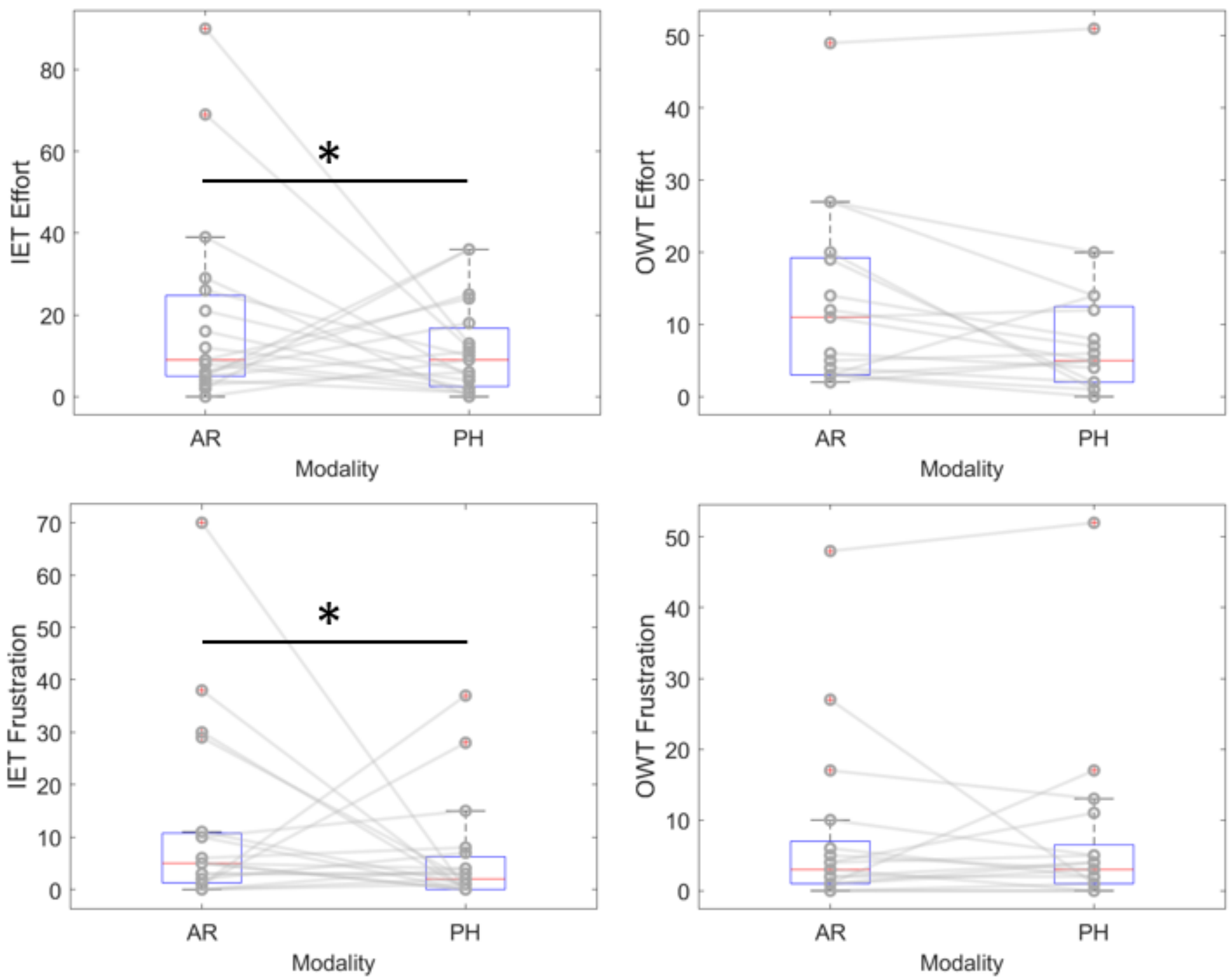

Figure E.7: NASA TLX measures for the Ingress and Egress Task and the Obstacle Weave.

APPENDIX F

Neurovestibular Assessments: Relative Angular Velocity of the Head and Torso

The following appendix presents a supplemental analysis of the neurovestibular assessment tasks comprising the tandem walk, four square step test, obstacle weave task, and the ingress and egress task. The relative angular velocity of the head and torso was evaluated to provide additional understanding of the head stabilization strategies utilized by participants. An average relative angular velocity value of zero suggests that the head and torso rotational motion are coupled. This particular strategy has been described as en bloc movement of the head and torso where the two body segments move as a single unit [190]. A positive value suggests that the head's rotational motion is greater than that of the torso, while a negative value indicates that the torso's rotational motion is greater than that of the head. In these instances, the relative head and trunk movement indicates that the segments move more independently from each other. The quantification of the relative movement between the head and torso can provide additional knowledge of user strategies in the augmented reality environment when compared to the physical implementations of the neurovestibular tasks in Chapters 5 and 6.

F.1 Methods

The raw angular velocity signals were provided from the inertial measurement unit sensors located on the head and the torso. An extended Kalman filter was used as a sensor fusion algorithm to estimate the orientation of the IMU, an analogous approach used in Chapters 5 and 6 Sections 5.1.4 and 6.1.4. The angular velocity signals were transformed from the local IMU frame of reference to the global axis, i.e., North (positive Y-axis), West (negative X-axis), Up (positive Z-axis), using the quaternions obtained from the extended Kalman filter. A conversion factor (108/pi) was applied to the angular velocity signals of the head and torso to represent the motion from radians per second to degrees per second. The L2-norm was obtained from the angular velocity along all three axes using equation 5.1.4. The relative angular velocity (ω_{relative}) of the head and torso using the RMS of the difference in magnitude angular velocity can be expressed as:

$$\omega_{\text{relative}_g} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\|\omega_{\text{head}_{ig}}\| - \|\omega_{\text{torso}_{ig}}\|)^2} \tag{F.1}$$

where:

ω_{relative} is the relative angular velocity,

N is the number of data points,

$\|\omega_{\text{head}_{ig}}\|$ is the L2-norm of the angular velocity of the head at time i in the global reference frame (g),

$\|\omega_{\text{torso}_{ig}}\|$ is the L2-norm of the angular velocity of the torso at time i in the global reference frame (g).

F.2 Results

A paired t-test analysis, similar to the statistical analyses applied in Chapters 5 and 6, was conducted to determine differences in head-trunk coordination strategies. A level of significance of 0.05 was applied and effect sizes were calculated using Cohen's method.

Significant differences between the augmented reality and physical modalities were identified for the two dynamic balance assessments (i.e., tandem walk and four square step test) and the two operational assessment tasks (i.e., the ingress and egress task and the obstacle weave) (Figure F.1. In AR, stabilization of the head with respect to the torso was demonstrated by a median near zero relative angular velocity value for the head and torso during the four square step test and obstacle weave task. A similar trend towards an en bloc stabilization strategy was found for the tandem walk and ingress and egress task in the physical environment.

During the tandem walk task in the physical environment, the relative angular velocity of the head and torso was closer to zero ($M = 15.10$ deg/s) than the AR-task performance suggesting more coupling of the head and torso, while the head moved more independently of the torso in the AR environment ($M = 28.91$ deg/s; $t(19) = -3.45, p = 0.003, d = 0.87$). The ingress and egress task resulted in head-to-trunk stabilization in the physical environment ($M = 9.12$ deg/s; $t(19) = -2.29, p = 0.033, d = 0.48$) and more uncoupled head movement from the torso in the AR environment ($M = 19.24$ deg/s).

Performance during the four square step test exhibited a greater magnitude of head-to-trunk coupling in the AR environment ($M = 4.06$ deg/s; $t(19) = -3.04, p = 0.007, d = 0.74$) while the torso movement was higher than that of the head movement in the physical environment ($M = -15.10$ deg/s). Similar head stabilization to the torso was identified for the obstacle weave task in AR ($M = -0.09$ deg/s; $t(19) = -6.15, p < 0.001, d = 1.06$) while the torso moved more independently of the head in the physical condition ($M = -14.66$ deg/s).

F.3 Discussion

The results from the present appendix evaluating the relative angular velocity of the head and torso provide a supplemental understanding of user performance from the findings of Chapters 5 and 6. In Chapter 5, the AR environment for the four square step test did not significantly alter the average magnitudes of head and torso angular velocity when compared to the physical environment. However, the present analysis suggests a significant difference in the relative motion of the head and torso within each modality where a head-to-trunk stabilization strategy was more apparent in the AR environment for the FSST task. Based on these results, the relative angular velocity was a more sensitive performance metric than the evaluation of the head and trunk angular velocity individually. The difference in the relative head and torso angular motion supplement findings from Chapter 5 for the tandem walk task where decreases in the magnitude of the torso angular velocity were identified as a compensatory strategy to promote the visibility of the holograms with the head. Higher angular head movement relative to the torso was shown in the AR environment while Chapter 5 identified a reduction in the absolute motion of the torso for the tandem walk task. The results from the present section coupled with Chapter 5 provide a broader understanding of user motion concerning relative inter-segmental changes and absolute head and torso motion while performing vestibular tasks in AR.

In Chapter 6, the operational assessment tasks demonstrated a consistent reduction in the head and torso angular velocities when considered across modalities. The present results suggest an increase in angular motion of the head relative to the torso in the ingress and egress task which is consistent with the increased magnitude of the downward head pitch in AR while no differences were identified for the torso pitch across both environments. The uncoupling of head motion from the torso supported more visual scanning of the task environment with the restricted field of view of the headset. The increased head movement also supported the clearance of the top of the task object. A divergent strategy was identified for the obstacle weave task in the AR environment where the head's angular motion was held more consistent with the torso during navigation around the two floor-based obstacles. These results coupled with the results in Chapter 6 suggest that the head and torso angular motion was reduced for the obstacle weave when comparing AR to the physical environment and the relative head angular motion was also reduced when comparing directly to the torso motion within the modalities. The coupling of the head and torso for the obstacle weave task in AR supported the successful circumnavigation of the obstacles and may be attributed to the reduced field of view of the headset. The average angular motion of the torso was larger when compared to head angular motion in the physical environment for the obstacle weave task where peripheral vision was supported.

The present results indicate that the use of augmented reality influenced diverse head-trunk coordination strategies during dynamic and operational balance assessments when compared to the completion of these tasks in a physical environment. Head-to-trunk stabilization and alterations in head-trunk coordination during locomotion have been identified postflight in crew members flowing both short and long-duration missions predominantly driven by motion sickness symptoms [11]. Head-to-trunk stabilization, or en bloc movement, has been demonstrated across various populations such as vestibular [191] patients and aging populations [192]. Future research should consider head-trunk coordination as a dependent measure for quantifying the crew's sensorimotor impairment and informing crew readiness and rehabilitation needs inflight following gravity transitions.

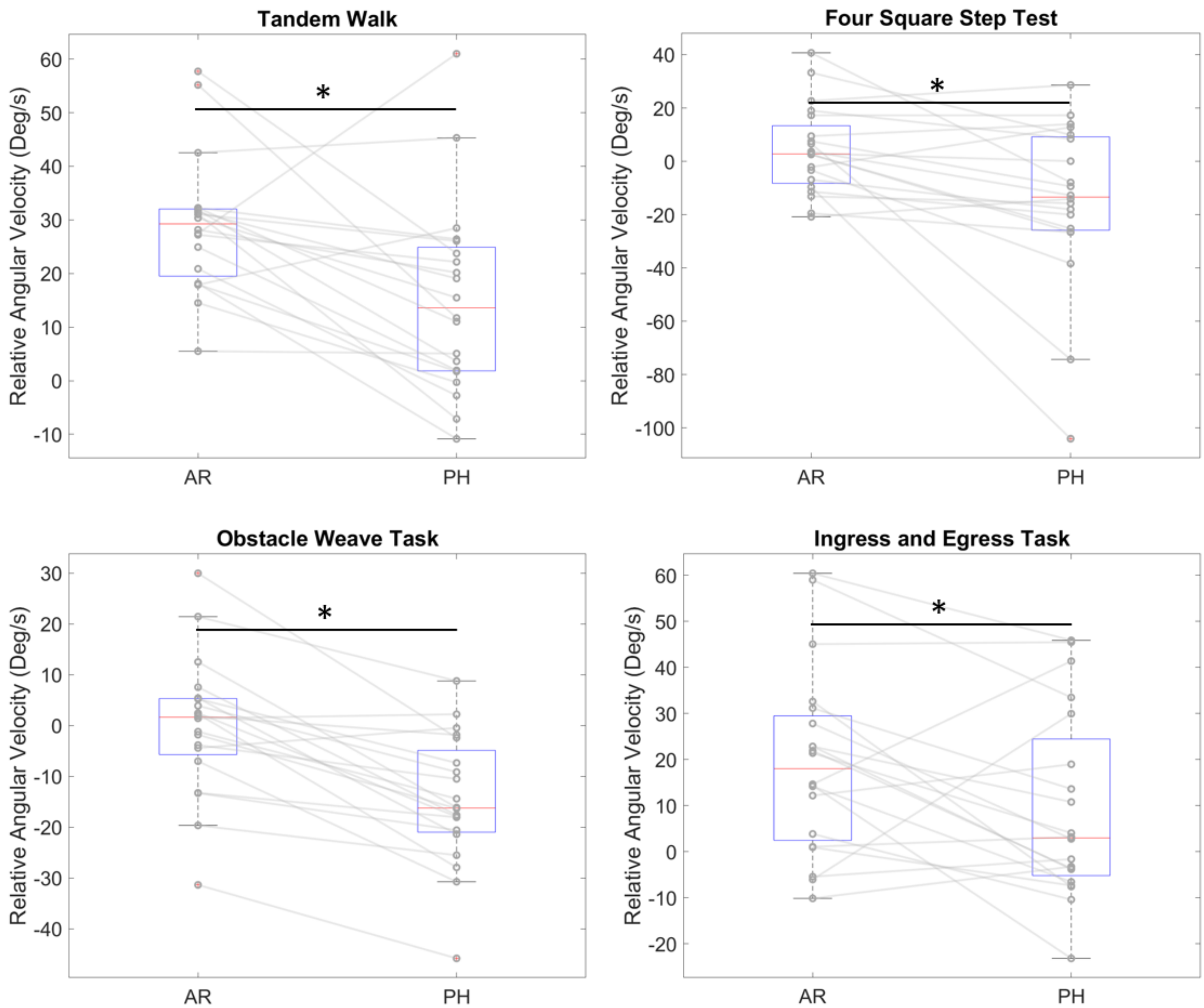

Figure F.1: The relative angular velocity of the head with respect to the torso is presented for the augmented reality and physical modalities for the dynamic and operational balance assessment tasks.

BIBLIOGRAPHY

[1] G. Clément, S. C. Moudy, T. R. Macaulay, M. O. Bishop, and S. J. Wood, “Mission-critical tasks for assessing risks from vestibular and sensorimotor adaptation during space exploration,” *Frontiers in Physiology*, vol. 13, Nov. 2022.

[2] S. Eggers and D. Zee, “Evaluating the dizzy patient: Bedside examination and laboratory assessment of the vestibular system,” *Seminars in Neurology*, vol. 23, no. 1, pp. 047–058, 2003.

[3] B. D. Lawson, A. H. Rupert, and B. J. McGrath, “The neurovestibular challenges of astronauts and balance patients: Some past countermeasures and two alternative approaches to elicitation, assessment and mitigation,” *Frontiers in Systems Neuroscience*, vol. 10, Nov 2016.

[4] H. S. Cohen, K. T. Kimball, A. P. Mulavara, J. J. Bloomberg, and W. H. Paloski, “Posturography and locomotor tests of dynamic balance after long-duration spaceflight,” *Journal of Vestibular Research*, vol. 22, pp. 191–196, Nov 2012.

[5] S. Scooley, S. Paniagua, J. Jaz, J. Mathew, B. Hanson, and Q. Wen, “Hololens 2 hardware.”

[6] R. W. Soukoreff and I. S. MacKenzie, “Towards a standard for pointing device evaluation, perspectives on 27 years of fitts’ law research in hci,” *Int. J. Hum.-Comput. Stud.*, vol. 61, p. 751–789, Dec 2004.

[7] “National Aeronautics and Space Administration Human Research Roadmap,” *NASA*.

[8] G. Clement and M. F. Reschke, *Neuroscience in Space*. New York, NY: Springer, 2008 ed., Aug 2008.

[9] M. F. Reschke, E. F. Good, and G. R. Clément, “Neurovestibular symptoms in astronauts immediately after space shuttle and international space station missions,” *OTO Open*, vol. 1, Oct 2017.

[10] W. H. Paloski, M. F. Reschke, F. O. Black, D. D. Doxey, and D. L. Harm, “Recovery of postural equilibrium control following spaceflight,” *Annals of the New York Academy of Sciences*, vol. 656, pp. 747–754, May 1992.

[11] J. J. Bloomberg, B. T. Peters, S. L. Smith, W. P. Huebner, M. F. Reschke, J. J. Bloomberg, and M. F. Reschke, “Locomotor head-trunk coordination strategies following space flight,” *J Vestib Res*, vol. 7, no. 2-3, pp. 161–177, 1997.

[12] C. A. Miller, B. T. Peters, R. R. Brady, J. R. Richards, R. J. Ploutz-Snyder, A. P. Mulavara, and J. J. Bloomberg, "Changes in toe clearance during treadmill walking after long-duration spaceflight," *Aviation, Space, and Environmental Medicine*, vol. 81, pp. 919–928, Oct 2010.

[13] M. Heer and W. H. Paloski, "Space motion sickness: Incidence, etiology, and countermeasures," *Autonomic Neuroscience*, vol. 129, pp. 77–79, Oct 2006.

[14] C. Layne, A. Mulavara, P. McDonald, C. Pruett, I. Kozlovskaya, and J. Bloomberg, "Alterations in human neuromuscular activation during over ground locomotion after long-duration spaceflight," *J Gravit Physiol*, vol. 11, pp. 1–16, 01 2004.

[15] K. Holden, "Effects of Lunar Mission Gravitational Transitions on Fine Motor Task Performance," in *Lunar Surface Science Workshop*, vol. 2241 of *LPI Contributions*, p. 5160, May 2020.

[16] K. Holden, M. Greene, E. Vincent, A. Sándor, S. Thompson, A. Feiveson, and B. Munson, "Effects of long-duration microgravity and gravitational transitions on fine motor skills," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, May 2022.

[17] S. Wood, J. Loehr, and M. Guilliams, "Sensorimotor reconditioning during and after spaceflight," *NeuroRehabilitation*, vol. 29, pp. 185–195, Oct 2011.

[18] NASA, "Risk of impaired control of spacecraft/associated systems and decreased mobility due to vestibular/sensorimotor alterations associated with space flight," evidence report, National Aeronautics and Space Administration Human Research Program, Lyndon B. Johnson Space Center, Houston, Texas, June 2016.

[19] S. J. Wood, W. H. Paloski, and J. B. Clark, "Assessing sensorimotor function following ISS with computerized dynamic posturography," *Aerospace Medicine and Human Performance*, vol. 86, pp. 45–53, Dec 2015.

[20] A. P. Mulavara, A. H. Feiveson, J. Fiedler, H. Cohen, B. T. Peters, C. Miller, R. Brady, and J. J. Bloomberg, "Locomotor function after long-duration space flight: effects and motor learning during recovery," *Experimental Brain Research*, vol. 202, pp. 649–659, Feb 2010.

[21] H. S. Cohen, J. Stitz, H. Sangi-Haghpeykar, S. P. Williams, A. P. Mulavara, B. T. Peters, and J. J. Bloomberg, "Tandem walking as a quick screening test for vestibular disorders," *The Laryngoscope*, vol. 128, pp. 1687–1691, Dec 2017.

[22] R. Newton, "Review of tests of standing balance abilities," *Brain Injury*, vol. 3, pp. 335–343, Jan 1989.

[23] P. A. Gribble, J. Hertel, and P. Plisky, "Using the star excursion balance test to assess dynamic postural-control deficits and outcomes in lower extremity injury: A literature and systematic review," *Journal of Athletic Training*, vol. 47, pp. 339–357, May 2012.

[24] A. Panjan and N. Sarabon, "Review of methods for the evaluation of human body balance," *Sport Science Review*, vol. 19, Jan 2010.

[25] A. M. Braly, B. Nuernberger, and S. Y. Kim, “Augmented reality improves procedural work on an international space station science instrument,” *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 61, pp. 866–878, Jan 2019.

[26] K. Marshall-Bowman, M. R. Barratt, and C. R. Gibson, “Ophthalmic changes and increased intracranial pressure associated with long duration spaceflight: An emerging understanding,” *Acta Astronautica*, vol. 87, pp. 77–87, June 2013.

[27] G. C. Demontis, M. M. Germani, E. G. Caiani, I. Barravecchia, C. Passino, and D. Angeloni, “Human pathophysiological adaptations to the space environment,” *Frontiers in Physiology*, vol. 8, Aug 2017.

[28] K. Marshall-Goebel, R. Damani, and E. M. Bershad, “Brain physiological response and adaptation during spaceflight,” *Neurosurgery*, vol. 85, pp. E815–E821, Jun 2019.

[29] A. V. Ombergen, S. Jillings, B. Jeurissen, E. Tomilovskaya, A. Rumshiskaya, L. Litvinova, I. Nosikova, E. Pechenkova, I. Rukavishnikov, O. Manko, S. Danylichev, R. M. Ruhl, I. B. Kozlovskaya, S. Sunaert, P. M. Parizel, V. Sinitsyn, S. Laureys, J. Sijbers, P. zu Eulenburg, and F. L. Wuyts, “Brain ventricular volume changes induced by long-duration spaceflight,” *Proceedings of the National Academy of Sciences*, vol. 116, no. 21, pp. 10531–10536, 2019.

[30] T. K. Clark, “Effects of spaceflight on the vestibular system,” in *Handbook of Space Pharmaceuticals*, pp. 273–311, Springer International Publishing, 2022.

[31] J. Carriot, I. Mackrous, and K. E. Cullen, “Challenges to the vestibular system in space: How the brain responds and adapts to microgravity,” *Frontiers in Neural Circuits*, vol. 15, Nov 2021.

[32] M. Shen and W. H. Frishman, “Effects of spaceflight on cardiovascular physiology and health,” *Cardiology in Review*, vol. 27, pp. 122–126, May 2019.

[33] L. C. Shackelford, “Musculoskeletal response to space flight,” in *Principles of Clinical Medicine for Space Flight*, pp. 581–607, Springer New York, 2019.

[34] B. Crucian, R. J. Simpson, S. Mehta, R. Stowe, A. Chouker, S.-A. Hwang, J. K. Actor, A. P. Salam, D. Pierson, and C. Sams, “Terrestrial stress analogs for spaceflight associated immune system dysregulation,” *Brain, Behavior, and Immunity*, vol. 39, pp. 23–32, Jul 2014.

[35] G. R. Taylor, I. Konstantinova, G. Sonnenfeld, and R. Jennings, “Chapter 1 changes in the immune system during and after spaceflight,” in *Advances in Space Biology and Medicine*, pp. 1–32, Elsevier, 1997.

[36] M. F. Reschke, J. J. Bloomberg, D. L. Harm, W. H. Paloski, C. Layne, and V. McDonald, “Posture, locomotion, spatial orientation, and motion sickness as a function of space flight,” *Brain Research Reviews*, vol. 28, pp. 102–117, Nov 1998.

[37] Z. S. Patel, T. J. Brunstetter, W. J. Tarver, A. M. Whitmire, S. R. Zwart, S. M. Smith, and J. L. Huff, “Red risks for a journey to the red planet: The highest priority human health risks for a mission to mars,” *npj Microgravity*, vol. 6, Nov 2020.

[38] B. L. Riemann and S. M. Lephart, "The sensorimotor system, part i: the physiologic basis of functional joint stability," *J. Athl. Train.*, vol. 37, pp. 71–79, Jan 2002.

[39] G. J. Augustine, D. Fitzpatrick, and D. Purves, *Neuroscience Including Sylvius CD-ROM*. Sunderland, MA: Sinauer Associates, 3 ed., Jul 2004.

[40] L. Minor, "Physiological principles of vestibular function on earth and in space," *Otolaryngology - Head and Neck Surgery*, vol. 118, pp. S5–S15, Mar 1998.

[41] A. E. Nicogossian, R. S. Williams, C. L. Huntoon, C. R. Doarn, J. D. Polk, and V. S. Schneider, eds., *Space physiology and medicine*. New York, NY: Springer, 4 ed., Dec 2016.

[42] M. F. Reschke, J. J. Bloomberg, D. L. Harm, and W. H. Paloski, "Space flight and neurovestibular adaptation," *The Journal of Clinical Pharmacology*, vol. 34, pp. 609–617, Jun 1994.

[43] G. Clément, "The neuro-sensory system in space," in *Fundamentals of Space Medicine*, pp. 95–142, Springer New York, 2011.

[44] L. Young and P. Sinha, "Spaceflight influences on ocular counterrolling and other neurovestibular reactions," *Otolaryngology - Head and Neck Surgery*, vol. 118, pp. S31–S34, Mar. 1998.

[45] C. M. Oman and K. E. Cullen, "Brainstem processing of vestibular sensory exafference: implications for motion sickness etiology," *Experimental Brain Research*, vol. 232, pp. 2483–2492, May 2014.

[46] C. M. Oman, "Spacelab experiments on space motion sickness," *Acta Astronautica*, vol. 15, pp. 55–66, Jan 1987.

[47] K. E. Hupfeld, H. R. McGregor, V. Koppelmans, N. E. Beltran, I. S. Kofman, Y. E. D. Dios, R. F. Riascos, P. A. Reuter-Lorenz, S. J. Wood, J. J. Bloomberg, A. P. Mulavara, and R. D. Seidler, "Brain and behavioral evidence for reweighting of vestibular inputs with long-duration spaceflight," *Cerebral Cortex*, vol. 32, pp. 755–769, Aug. 2021.

[48] J. J. Bloomberg, M. F. Reschke, G. Clément, A. P. Mulavara, and L. C. Taylor, "Risk of impaired control of spacecraft/associated systems and decreased mobility due to vestibular/sensorimotor alterations associated with space flight," 2015.

[49] H. S. Song and H. J. Lee, "Fear of falling and associated factors among patients with peripheral vestibular hypofunction," *J Exerc Rehabil*, vol. 16, pp. 162–167, Apr 2020.

[50] S. Genah, M. Monici, and L. Morbidelli, "The effect of space travel on bone metabolism: Considerations on today's major challenges and advances in pharmacology," *International Journal of Molecular Sciences*, vol. 22, p. 4585, Apr 2021.

[51] Y. M. Paez, L. I. Mudie, and P. S. Subramanian, "Spaceflight associated neuro-ocular syndrome (SANS): A systematic review and future directions," *Eye and Brain*, vol. Volume 12, pp. 105–117, Oct 2020.

[52] T. H. Mader, C. R. Gibson, C. A. Otto, A. E. Sargsyan, N. R. Miller, P. S. Subramanian, S. F. Hart, W. Lipsky, N. B. Patel, and A. G. Lee, "Persistent asymmetric optic disc swelling after long-duration space flight: Implications for pathogenesis," *Journal of Neuro-Ophthalmology*, vol. 37, pp. 133–139, Jun 2017.

[53] A. G. Lee, T. H. Mader, C. R. Gibson, W. Tarver, P. Rabiei, R. F. Riascos, L. A. Galdamez, and T. Brunstetter, "Spaceflight associated neuro-ocular syndrome (SANS) and the neuro-ophthalmologic effects of microgravity: a review and an update," *npj Microgravity*, vol. 6, Feb 2020.

[54] M. B. Stenger, W. J. Tarver, T. Brunstetter, C. R. Gibson, S. S. Laurie, S. Lee, B. R. Macias, T. H. Mader, C. Otto, S. M. Smith, *et al.*, "Evidence report: risk of spaceflight associated neuro-ocular syndrome (SANS)," tech. rep., 2017.

[55] T. H. Mader, C. R. Gibson, M. R. Barratt, N. R. Miller, P. S. Subramanian, H. E. Killer, W. J. Tarver, A. E. Sargsyan, K. Garcia, S. F. Hart, L. A. Kramer, R. Riascos, T. J. Brunstetter, W. Lipsky, P. Wostyn, and A. G. Lee, "Persistent globe flattening in astronauts following long-duration spaceflight," *Neuro-Ophthalmology*, vol. 45, pp. 29–35, Sept. 2020.

[56] C. R. Ferguson, L. P. Pardon, S. S. Laurie, M. H. Young, C. R. Gibson, T. J. Brunstetter, W. J. Tarver, S. S. Mason, P. A. Sibony, and B. R. Macias, "Incidence and progression of chorioretinal folds during long-duration spaceflight," *JAMA Ophthalmology*, vol. 141, p. 168, Feb 2023.

[57] T. H. Mader, C. R. Gibson, A. F. Pass, L. A. Kramer, A. G. Lee, J. Fogarty, W. J. Tarver, J. P. Dervay, D. R. Hamilton, A. Sargsyan, J. L. Phillips, D. Tran, W. Lipsky, J. Choi, C. Stern, R. Kuyumjian, and J. D. Polk, "Optic disc edema, globe flattening, choroidal folds, and hyperopic shifts observed in astronauts after long-duration space flight," *Ophthalmology*, vol. 118, pp. 2058–2069, Oct 2011.

[58] G. Clément, S. J. Wood, W. H. Paloski, and M. F. Reschke, "Changes in gain of horizontal vestibulo-ocular reflex during spaceflight," *Journal of Vestibular Research*, vol. 29, pp. 241–251, Nov 2019.

[59] L. N. Kornilova, V. Grigorova, and G. Bodo, "Vestibular function and sensory interaction in space flight," *J. Vestib. Res.*, vol. 3, no. 3, pp. 219–230, 1993.

[60] L. N. Kornilova, I. A. Naumov, K. A. Azarov, and V. N. Sagalovitch, "Gaze control and vestibular-cervical-ocular responses after prolonged exposure to microgravity," *Aviation, Space, and Environmental Medicine*, vol. 83, pp. 1123–1134, Dec 2012.

[61] L. N. Kornilova, S. V. Sagalovitch, V. V. Temnikova, and A. G. Yakushev, "Static and dynamic vestibulo-cervico-ocular responses after prolonged exposure to microgravity," *J. Vestib. Res.*, vol. 17, no. 5-6, pp. 217–226, 2007.

[62] J. J. Uri, W. E. Thornton, T. P. Moore, and S. L. Pool, "Visual suppression of the vestibulo-ocular reflex during space flight," 1989.

[63] M. F. Reschke, O. I. Kolev, and G. Clément, "Eye-head coordination in 31 space shuttle astronauts during visual target acquisition," *Scientific Reports*, vol. 7, Oct 2017.

[64] O. Bock, S. Abeele, and U. Eversheim, "Sensorimotor performance and computational demand during short-term exposure to microgravity," *Aviat. Space Environ. Med.*, vol. 74, pp. 1256–1262, Dec 2003.

[65] D. M. Wolpert, "Computational approaches to motor control," *Trends in Cognitive Sciences*, vol. 1, no. 6, pp. 209–216, 1997.

[66] M. I. Jordan and D. E. Rumelhart, "Forward models: Supervised learning with a distal teacher," *Cognitive Science*, vol. 16, no. 3, pp. 307–354, 1992.

[67] D. Williams, A. Kuipers, C. Mukai, and R. Thirsk, "Acclimation during space flight: effects on human physiology," *CMAJ*, vol. 180, no. 13, pp. 1317–1323, 2009.

[68] G. Clément, "The musculo-skeletal system in space," in *Fundamentals of Space Medicine*, pp. 181–216, Springer New York, 2011.

[69] R. H. Fitts, S. W. Trappe, D. L. Costill, P. M. Gallagher, A. C. Creer, P. A. Colloton, J. R. Peters, J. G. Romatowski, J. L. Bain, and D. A. Riley, "Prolonged space flight-induced alterations in the structure and function of human skeletal muscle fibres," *The Journal of Physiology*, vol. 588, pp. 3567–3592, Sep 2010.

[70] H. Vandenburgh, J. Chromiak, J. Shansky, M. D. Tatto, and J. Lemaire, "Space travel directly induces skeletal muscle atrophy," *The FASEB Journal*, vol. 13, pp. 1031–1038, June 1999.

[71] D. Moosavi, D. Wolovsky, A. Depompeis, D. Uher, D. Lennington, R. Bodden, and C. Garber, "The effects of spaceflight microgravity on the musculoskeletal system of humans and animals, with an emphasis on exercise as a countermeasure: a systematic scoping review," *Physiological Research*, pp. 119–151, Apr 2021.

[72] M. Berger, S. Mescheriakov, E. Molokanova, S. Lechner-Steinleitner, N. Seguer, and I. Kozlovskaya, "Pointing arm movements in short- and long-term spaceflights," *Aviat. Space Environ. Med.*, vol. 68, pp. 781–787, Sep 1997.

[73] O. Bock, B. Fowler, and D. Comfort, "Human sensorimotor coordination during spaceflight: an analysis of pointing and tracking responses during the "neurolab" space shuttle mission," *Aviat. Space Environ. Med.*, vol. 72, pp. 877–883, Oct 2001.

[74] D. Manzey, B. Lorenz, H. Heuer, and J. Sangals, "Impairments of manual tracking performance during spaceflight: more converging evidence from a 20-day space mission," *Ergonomics*, vol. 43, no. 5, pp. 589–609, 2000. PMID: 10877478.

[75] A. P. Mulavara, B. T. Peters, C. A. Miller, I. S. Kofman, M. F. Reschke, L. C. Taylor, E. L. Lawrence, S. J. Wood, S. S. Laurie, S. M. C. Lee, R. E. Buxton, T. R. May-Phillips, M. B. Stenger, L. L. Ploutz-Snyder, J. W. Ryder, A. H. Feiveson, and J. J. Bloomberg, "Physiological and functional alterations after spaceflight and bed rest," *Medicine & Science in Sports & Exercise*, vol. 50, pp. 1961–1980, Sep 2018.

[76] S. T. Moore, V. Dilda, T. R. Morris, D. A. Yungher, H. G. MacDougall, and S. J. Wood, "Long-duration spaceflight adversely affects post-landing operator proficiency," *Scientific Reports*, vol. 9, Feb 2019.

[77] F. O. Black and W. H. Paloski, "Computerized dynamic posturography: What have we learned from space?," *Otolaryngology–Head and Neck Surgery*, vol. 118, Mar 1998.

[78] F. O. BLACK, "What can posturography tell us about vestibular function?," *Annals of the New York Academy of Sciences*, vol. 942, pp. 446–464, Jan 2006.

[79] H.-C. Chen, J. A. Ashton-Miller, N. B. Alexander, and A. B. Schultz, "Stepping over obstacles: Gait patterns of healthy young and old adults," *Journal of Gerontology*, vol. 46, pp. M196–M203, Nov 1991.

[80] H.-C. Chen, A. B. Schultz, J. A. Ashton-Miller, B. Giordani, N. B. Alexander, and K. E. Guire, "Stepping over obstacles: Dividing attention impairs performance of old more than young adults," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 51A, pp. M116–M122, May 1996.

[81] H. S. Cohen and K. T. Kimball, "Usefulness of some current balance tests for identifying individuals with disequilibrium due to vestibular impairments," *J. Vestib. Res.*, vol. 18, no. 5-6, pp. 295–303, 2008.

[82] B. Merker and K. Podell, "Grooved pegboard test," in *Encyclopedia of Clinical Neuropsychology*, pp. 1176–1178, Springer New York, 2011.

[83] E.-Y. Lee, V. T. Tran, and D. Kim, "A novel head mounted display based methodology for balance evaluation and rehabilitation," *Sustainability*, vol. 11, p. 6453, Nov. 2019.

[84] F. Mostajeran, F. Steinicke, O. J. Ariza Nunez, D. Gatsios, and D. Fotiadis, "Augmented reality for older adults: Exploring acceptability of virtual coaches for home-based balance training in an aging population," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, (New York, NY, USA), p. 1–12, Association for Computing Machinery, 2020.

[85] I. Kouris, M. Sarafidis, T. Androutsou, and D. Koutsouris, "Holobalance: An augmented reality virtual trainer solution forbalance training and fall prevention," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4233–4236, 2018.

[86] M. Liston, G. Genna, C. Maurer, D. Kikidis, D. Gatsios, D. Fotiadis, D.-E. Bamiou, and M. Pavlou, "Investigating the feasibility and acceptability of the holobalance system compared with standard care in older adults at risk for falls: study protocol for an assessor blinded pilot randomized controlled study," *BMJ Open*, vol. 11, no. 2, 2021.

[87] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement.," *Journal of Experimental Psychology: General*, vol. 121, no. 3, pp. 262–269, 1992.

[88] D. M. Mifsud, A. S. Williams, F. Ortega, and R. J. Teather, "Augmented reality fitts' law input comparison between touchpad, pointing gesture, and raycast," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 590–591, 2022.

[89] "Ergonomic requirements for office work with visual display terminals (vdts)—part 9—requirements for non-keyboard input devices iso 9241-9: 2000(e)," tech. rep., February 2002.

[90] I. S. MacKenzie, "Fitts' throughput and the remarkable case of touch-based target selection," in *Interacción*, 2015.

[91] B. Shneiderman and C. Plaisant, *Designing the User Interface: Strategies for Effective Human-computer Interaction*. Addison Wesley, 08 2010.

[92] J. Nielsen, "Enhancing the explanatory power of usability heuristics," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '94, (New York, NY, USA), p. 152–158, Association for Computing Machinery, 1994.

[93] C. D. Wickens and C. M. Carswell, "The proximity compatibility principle: Its psychological foundation and relevance to display design," *Human Factors*, vol. 37, no. 3, pp. 473–494, 1995.

[94] R. A. Grier, A. Bangor, P. Kortum, and S. C. Peres, "The system usability scale: Beyond standard usability testing," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 57, no. 1, pp. 187–191, 2013.

[95] P. Harper and K. Norman, "Improving user satisfaction: The questionnaire for user interaction satisfaction version 5.5," 01 1993.

[96] A. D. Andre and C. D. Wickens, "When users want what's not best for them," *Ergonomics in Design*, vol. 3, no. 4, pp. 10–14, 1995.

[97] M. R. Drew, B. Falcone, and W. L. Baccus, "What does the system usability scale (sus) measure? validation using think aloud verbalization and behavioral metrics," (Berlin, Heidelberg), Springer-Verlag, 2018.

[98] J. Sauro and J. R. Lewis, "Standardized usability questionnaires," in *Quantifying the User Experience*, pp. 185–248, Elsevier, 2016.

[99] "Characterize the effects of short and long-duration weightlessness, with and without deep-space radiation, on manual control (fine motor control) after g transitions.," *National Aeronautics and Space Administration Human Research Project*.

[100] "Characterize the effects of short and long-duration weightlessness, with and without deep-space radiation, on spatial orientation and motion sickness after g transitions," *National Aeronautics and Space Administration Human Research Project*.

[101] "Evaluate how weightlessness-induced changes in sensorimotor/vestibular function relate to and/or interact with changes in other brain functions (sleep, cognition, attention)," *National Aeronautics and Space Administration Human Research Project*.

[102] "Characterize the effects of short and long-duration weightlessness, with and without deep-space radiation, on postural control and locomotion (gross motor control) after g transitions," *National Aeronautics and Space Administration Human Research Project*.

[103] "Develop and test manual control countermeasures, such as vibrotactile assistance vest, and other human factors aids," *National Aeronautics and Space Administration Human Research Project*.

[104] "Develop and test SMS countermeasures.," *National Aeronautics and Space Administration Human Research Project*.

[105] "Develop and test post-planetary-landing self-administered testing and rehab tool.," *National Aeronautics and Space Administration Human Research Project*.

[106] "2020 nasa technology taxonomy," *National Aeronautics and Space Administration*.

[107] "Nasa human integration design handbook (hidh)," *National Aeronautics and Space Administration*.

[108] M. Mancini and F. B. Horak, "The relevance of clinical balance assessment tools to differentiate balance deficits," *Eur. J. Phys. Rehabil. Med.*, vol. 46, pp. 239–248, June 2010.

[109] "Space flight human-system standard volume 1: Crew health," tech. rep., National Aeronautics and Space Administration (NASA), 2022.

[110] "Human integration design handbook (hidh)," Tech. Rep. Revision 1, NASA, Houston, TX, 2014.

[111] J. Bienz, B. Hanson, and J. Jaz, "Hololens 2 battery and charging," 2021.

[112] M. Ghislieri, L. Gastaldi, S. Pastorelli, S. Tadano, and V. Agostini, "Wearable inertial sensors to assess standing balance: A systematic review," *Sensors*, vol. 19, p. 4075, 09 2019.

[113] G. Zari, S. Condino, F. Cutolo, and V. Ferrari, "Magic leap 1 versus microsoft HoloLens 2 for the visualization of 3d content obtained from radiological images," *Sensors*, vol. 23, p. 3040, Mar. 2023.

[114] W. Emery and A. Camps, "Chapter 3 - optical imaging systems," in *Introduction to Satellite Remote Sensing* (W. Emery and A. Camps, eds.), pp. 85–130, Elsevier, 2017.

[115] E. Miller, B. Hanson, S. Swaminathan, S. Paniagua, and J. Jaz, "Improve visual quality and comfort," Sep 2022.

[116] A. J. Marcori, P. H. M. Monteiro, J. A. Oliveira, M. Doumas, and L. A. Teixeira, "Single leg balance training: A systematic review," *Perceptual and Motor Skills*, vol. 129, pp. 232–252, Jan 2022.

[117] B. Joo, J. L. Marquez, and P. G. Osmotherly, "Ten-second tandem stance test: A potential tool to assist walking aid prescription and falls risk in balance impaired individuals," *Archives of Rehabilitation Research and Clinical Translation*, vol. 4, p. 100173, Mar. 2022.

[118] D. R. Bell, K. M. Guskiewicz, M. A. Clark, and D. A. Padua, “Systematic review of the balance error scoring system,” *Sports Health: A Multidisciplinary Approach*, vol. 3, pp. 287–295, Apr 2011.

[119] G. Clément, S. J. Wood, W. H. Paloski, and M. F. Reschke, “Changes in gain of horizontal vestibulo-ocular reflex during spaceflight,” *Journal of Vestibular Research*, vol. 29, pp. 241–251, Nov 2019.

[120] S. L. Whitney, G. F. Marchetti, L. O. Morris, and P. J. Sparto, “The reliability and validity of the four square step test for people with balance deficits secondary to a vestibular disorder,” *Arch. Phys. Med. Rehabil.*, vol. 88, pp. 99–104, Jan 2007.

[121] W. Dite and V. A. Temple, “A clinical test of stepping and change of direction to identify multiple falling older adults,” *Archives of Physical Medicine and Rehabilitation*, vol. 83, pp. 1566–1571, Nov 2002.

[122] G. Gray, *Lower Extremity Functional Profile*. Wynn Marketing, Inc, 1995.

[123] L. Olmsted, C. Carcia, J. Hertel, and S. Shultz, “Efficacy of the star excursion balance tests in detecting reach deficits in subjects with chronic ankle instability,” *Journal of athletic training*, vol. 37, pp. 501–506, 01 2003.

[124] I. S. MacKenzie, T. Kauppinen, and M. Silfverberg, “Accuracy measures for evaluating computer pointing devices,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’01, (New York, NY, USA), p. 9–16, Association for Computing Machinery, 2001.

[125] K. L. Norman and K. D. Norman, “Comparison of relative versus absolute pointing devices,” tech. rep., 2010.

[126] R. Blodgett, “Frequently asked questions,” *National Aeronautics and Space Administration*, Feb 2022.

[127] H. Weiss and L. Stirling, “Usability evaluation of an augmented reality sensorimotor assessment tool for astronauts,” in *2023 IEEE Aerospace Conference*, IEEE, Mar 2023.

[128] V. Venkatesh and F. D. Davis, “A theoretical extension of the technology acceptance model: Four longitudinal field studies,” *Management Science*, vol. 46, no. 2, pp. 186–204, 2000.

[129] J. P. Chin, V. A. Diehl, and K. L. Norman, “Development of an instrument measuring user satisfaction of the human-computer interface,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’88, (New York, NY, USA), p. 213–218, Association for Computing Machinery, 1988.

[130] T. Jung, M. C. Tom Dieck, and D. tom Dieck, “A theoretical model of augmented reality acceptance,” *e-Review of Tourism Research*, vol. 5, 03 2014.

[131] M. E. C. Santos, T. Taketomi, C. Sandor, J. Polvi, G. Yamamoto, and H. Kato, “A usability scale for handheld augmented reality,” in *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology - VRST '14*, ACM Press, 2014.

[132] R. Johns, "Likert items and scales, survey question bank: Mathods fact sheet 1," *March, available at: www.surveynet.ac.uk/sqb/datacollection/likertfactsheet.pdf*, 2010.

[133] F. Rometsch, A. Casini, A. Drepper, A. Cowley, J. de Winter, and J. Guo, "Design and evaluation of an augmented reality tool for future human space exploration aided by an internet of things architecture," *Advances in Space Research*, 07 2022.

[134] T. Piumsomboon, A. Clark, M. Billinghurst, and A. Cockburn, "User-defined gestures for augmented reality," in *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, (New York, NY, USA), p. 955–960, Association for Computing Machinery, 2013.

[135] N. F. S. Jeffri and D. R. A. Rambli, "A review of augmented reality systems and their effects on mental workload and task performance," *Heliyon*, vol. 7, p. e06277, Mar. 2021.

[136] G. Ren, S. Wei, E. O'Neill, and F. Chen, "Towards the design of effective haptic and audio displays for augmented reality and mixed reality applications," *Advances in Multimedia*, vol. 2018, pp. 1–11, July 2018.

[137] F. Cutolo and V. Ferrari, "The role of camera convergence in stereoscopic video see-through augmented reality displays," *International Journal of Advanced Computer Science and Applications*, vol. 9, 2018.

[138] G. Clément, A. Skinner, and C. Lathan, "Distance and size perception in astronauts during long-duration spaceflight," *Life*, vol. 3, pp. 524–537, Dec 2013.

[139] G. Clément, H. C. M. Allaway, M. Demel, A. Golemis, A. N. Kindrat, A. N. Melinyshyn, T. Merali, and R. Thirsk, "Long-duration spaceflight increases depth ambiguity of reversible perspective figures," *PLOS ONE*, vol. 10, pp. 1–16, 07 2015.

[140] B. Beard, "The cognition and fine motor skills test batteries: Normative data and interdependencies," technical memorandum, National Aeronautics and Space Administration, Ames Research Center Mountain View, California, June 2019.

[141] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, Jul 1948.

[142] "Man-systems integration standards: Workstations," *National Aeronautics and Space Administration*, vol. Volume 1, Section 9, 1995.

[143] S. G. Hart and L. E. Staveland, "Development of NASA TLX (task load index): Results of empirical and theoretical research," pp. 139–183, 1988.

[144] MATLAB, *Convhull Documentation*. 2022.

[145] N. Chernov, "Circle fit (pratt method)," 2022.

[146] Y. Zhang, "Combining absolute and relative pointing for fast and accurate distant interaction," 2017.

[147] J.-Y. Oh and W. Stuerzlinger, “Laser pointers as collaborative pointing devices,” *Proceedings - Graphics Interface*, vol. 2002, 10 2002.

[148] H. Weiss and L. Stirling, “Usability evaluation of an augmented reality sensorimotor assessment tool for astronauts,” in *2023 IEEE Aerospace Conference*, IEEE, Mar. 2023.

[149] D. Schneider, V. Biener, A. Otte, T. Gesslein, P. Gagel, C. Campos, K. Č. Pucihar, M. Kljun, E. Ofek, M. Pahud, P. O. Kristensson, and J. Grubert, “Accuracy evaluation of touch tasks in commodity virtual and augmented reality head-mounted displays,” in *Symposium on Spatial User Interaction*, ACM, Nov 2021.

[150] E. Hoggan, S. A. Brewster, and J. Johnston, “Investigating the effectiveness of tactile feedback for mobile touchscreens,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’08, (New York, NY, USA), p. 1573–1582, Association for Computing Machinery, 2008.

[151] M. Frutos-Pascual, J. M. Harrison, C. Creed, and I. Williams, “Evaluation of ultrasound haptics as a supplementary feedback cue for grasping in virtual environments,” in *2019 International Conference on Multimodal Interaction*, ICMI ’19, (New York, NY, USA), p. 310–318, Association for Computing Machinery, 2019.

[152] D. Ungureanu, F. Bogo, S. Galliani, P. Sama, X. Duan, C. Meekhof, J. St{u}hmer, T. J. Cashman, B. Tekin, J. L. Sch{o}nberger, P. Olszta, and M. Pollefeys, “Hololens 2 research mode as a tool for computer vision research,” 2020.

[153] H. Ro, J.-H. Byun, Y. J. Park, N. K. Lee, and T.-D. Han, “Ar pointer: Advanced ray-casting interface using laser pointer metaphor for object manipulation in 3d augmented reality environment,” *Applied Sciences*, vol. 9, no. 15, 2019.

[154] J. H. Kim, H. Ari, C. Madasu, and J. Hwang, “Evaluation of the biomechanical stress in the neck and shoulders during augmented reality interactions,” *Applied Ergonomics*, vol. 88, p. 103175, Oct 2020.

[155] V. Reynaert, Y. Rekik, F. Berthaut, and L. Grisoni, “The effect of hands synchronicity on users perceived arms fatigue in virtual reality environment,” *International Journal of Human-Computer Studies*, vol. 178, p. 103092, 2023.

[156] X. Lou, Q. Zhao, Y. Shi, and P. Hansen, “Arm posture changes and influences on hand controller interaction evaluation in virtual reality,” *Applied Sciences*, vol. 12, no. 5, 2022.

[157] L. D. Clark, A. B. Bhagat, and S. L. Riggs, “Extending fitts’ law in three-dimensional virtual environments with current low-cost virtual reality technology,” *International Journal of Human-Computer Studies*, vol. 139, p. 102413, July 2020.

[158] A. Murata and H. Iwase, “Extending fitts' law to a three-dimensional pointing task,” *Human Movement Science*, vol. 20, pp. 791–805, Dec 2001.

[159] Y. Cha and R. Myung, “Extended fitts' law for 3d pointing tasks using 3d target arrangements,” *International Journal of Industrial Ergonomics*, vol. 43, pp. 350–355, July 2013.

[160] A. P.-D. de-la Lastra, R. Moreta-Martinez, M. García-Sevilla, D. García-Mato, J. A. Calvo-Haro, L. Mediavilla-Santos, R. Pérez-Mañanes, F. von Haxthausen, and J. Pascau, "HoloLens 1 vs. HoloLens 2: Improvements in the new model for orthopedic oncological interventions," *Sensors*, vol. 22, p. 4915, Jun 2022.

[161] G. Ligorio and A. Sabatini, "Extended kalman filter-based methods for pose estimation using visual, inertial and magnetic sensors: Comparative analysis and performance evaluation," *Sensors*, vol. 13, pp. 1919–1941, Feb 2013.

[162] A. M. Sabatini, "Estimating three-dimensional orientation of human body parts by inertial/magnetic sensing," *Sensors*, vol. 11, pp. 1489–1525, Jan 2011.

[163] G. Nguyen, J. Maclean, and L. Stirling, "Quantification of compensatory torso motion in post-stroke patients using wearable inertial measurement units," *IEEE Sensors Journal*, vol. 21, pp. 15349–15360, Jul 2021.

[164] G. Wu and P. R. Cavanagh, "ISB recommendations for standardization in the reporting of kinematic data," *Journal of Biomechanics*, vol. 28, pp. 1257–1261, Oct 1995.

[165] A. Gouelle and M. J. Highsmith, "Instrumented four square step test in adults with transfemoral amputation: Test-retest reliability and discriminant validity between two types of microprocessor knees," *Sensors*, vol. 20, p. 4782, Aug 2020.

[166] S. M. Shearin, K. J. McCain, and R. Querry, "Description of novel instrumented analysis of the four square step test with clinical application: A pilot study," *Gait & Posture*, vol. 82, pp. 14–19, 2020.

[167] K. Kim, Y. Gimmon, J. Millar, and M. Schubert, "Using inertial sensors to quantify postural sway and gait performance during the tandem walking test," *Sensors*, vol. 19, p. 751, Feb 2019.

[168] F. Torlak and M. Moffat, "PP17 four square step test normative data for healthy young adults," *British Journal of Sports Medicine*, vol. 48, pp. A11.3–A11, Jun 2014.

[169] W. Dite, H. J. Connor, and H. C. Curtis, "Clinical identification of multiple fall risk early after unilateral transtibial amputation," *Archives of Physical Medicine and Rehabilitation*, vol. 88, pp. 109–114, Jan 2007.

[170] J. M. Blennerhassett and V. M. Jayalath, "The four square step test is a feasible and valid clinical test of dynamic standing balance for use in ambulant people poststroke," *Archives of Physical Medicine and Rehabilitation*, vol. 89, pp. 2156–2161, Nov 2008.

[171] H. Shimada, M. Suzukawa, A. Tiedemann, K. Kobayashi, H. Yoshida, and T. Suzuki, "Which neuromuscular or cognitive test is the optimal screening tool to predict falls in frail community-dwelling older people?," *Gerontology*, vol. 55, no. 5, pp. 532–538, 2009.

[172] M. Kim, S. Seino, M. Kim, N. Yabushita, T. Okura, J. Okuno, and K. Tanaka, "Validation of lower extremity performance tests for determining the mobility limitation levels in community-dwelling older women," *Aging Clinical and Experimental Research*, vol. 21, pp. 437–444, Dec 2009.

[173] C. Hennah and M. Doumas, "Dual-task walking on real-world surfaces: Adaptive changes in walking speed, step width and step height in young and older adults," *Experimental Gerontology*, vol. 177, p. 112200, Jun 2023.

[174] H. S. Cohen, A. P. Mulavara, B. T. Peters, H. Sangi-Haghpeykar, D. H. Kung, D. R. Mosier, and J. J. Bloomberg, "Sharpening the tandem walking test for screening peripheral neuropathy," *Southern Medical Journal*, vol. 106, pp. 565–569, Oct 2013.

[175] S. Gonzalez, P. Stegall, S. M. Cain, H. C. Siu, and L. Stirling, "Assessment of a powered ankle exoskeleton on human stability and balance," *Applied Ergonomics*, vol. 103, p. 103768, Sept. 2022.

[176] H. Weiss, J. Tang, and L. Stirling, "Performance on a target acquisition task differs between augmented reality and touch screen displays," *Applied Ergonomics*, 2023. Submitted for publication.

[177] M. Ghislieri, L. Gastaldi, S. Pastorelli, S. Tadano, and V. Agostini, "Wearable inertial sensors to assess standing balance: A systematic review," *Sensors*, vol. 19, no. 19, 2019.

[178] J. Bloomberg and A. Mulavara, "Changes in walking strategies after spaceflight," *IEEE Engineering in Medicine and Biology Magazine*, vol. 22, pp. 58–62, Mar 2003.

[179] J. Golding, "Motion sickness," in *Handbook of Clinical Neurology*, pp. 371–390, Elsevier, 2016.

[180] G. Clément, S. C. Moudy, T. R. Macaulay, M. O. Bishop, and S. J. Wood, "Mission-critical tasks for assessing risks from vestibular and sensorimotor adaptation during space exploration," *Frontiers in Physiology*, vol. 13, Nov 2022.

[181] "ANSI/HFES 100-2007 Human factors engineering of computer workstations," *Human Factors and Ergonomics Society*, 2007.

[182] *Mixed Reality Microsoft Learn*, Oct 2021.

[183] J. L. Gabbard, J. E. Swan, and D. Hix, "The effects of text drawing styles, background textures, and natural lighting on text legibility in outdoor augmented reality," *Presence: Teleoper. Virtual Environ.*, vol. 15, p. 16–32, Feb 2006.

[184] S. Debernardis, M. Fiorentino, M. Gattullo, G. Monno, and A. E. Uva, "Text readability in head-worn displays: Color and style optimization in video versus optical see-through devices," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 1, pp. 125–139, 2014.

[185] R. A. Wynne, K. J. Parnell, M. A. Smith, K. L. Plant, and N. A. Stanton, "Can't touch this: Hammer time on touchscreen task performance variability under simulated turbulent flight conditions," *International Journal of Human–Computer Interaction*, vol. 37, p. 666–679, Mar 2021.

[186] A. Cockburn, C. Gutwin, P. Palanque, Y. Deleris, C. Trask, A. Coveney, M. Yung, and K. MacLean, “Turbulent touch: Touchscreen input for cockpit flight displays,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI ’17, (New York, NY, USA), p. 6742–6753, Association for Computing Machinery, 2017.

[187] S. R. Dodd, J. Lancaster, S. Grothe, B. DeMers, B. Rogers, and A. Miranda, “Touch on the flight deck: The impact of display location, size, touch technology & turbulence on pilot performance,” in *2014 IEEE/AIAA 33rd Digital Avionics Systems Conference (DASC)*, pp. 2C3–1–2C3–13, 2014.

[188] J. Lee, C. Wickens, Y. Liu, and L. Boyle, *Designing for People: An introduction to human factors engineering*. 08 2017.

[189] A. Allred, H. Weiss, T. Clark, and L. Stirling, “An augmented reality hand-eye sensorimotor impairment assessment for spaceflight operations,” *Aerospace Medicine and Human Performance*, 2024. In-press article.

[190] C. Assaiante and B. Amblard, “An ontogenetic model for the sensorimotor organization of balance control in humans,” *Human Movement Science*, vol. 14, no. 1, pp. 13–43, 1995.

[191] S. S. Paul, L. E. Dibble, R. G. Walther, C. Shelton, R. K. Gurgel, and M. E. Lester, “Characterization of head-trunk coordination deficits after unilateral vestibular hypofunction using wearable sensors,” *JAMA Otolaryngology–Head & Neck Surgery*, vol. 143, p. 1008, Oct 2017.

[192] A. J. Cocks, W. R. Young, T. J. Ellmers, R. C. Jackson, and A. M. Williams, “Concern about falling is associated with segmental control when turning in older adults,” *Gait & Posture*, vol. 88, pp. 105–108, 2021.