51st International Conference on Environmental Systems
10-14 July 2022, Saint Paul, Minnesota

ICES-2022-367

Experimental Investigation of Minimum Cabin Sizes at Varying Gravity Levels

Zachary Lachance[1], David Akin[2], Charles Hanner[1], and Nicolas Bolatto[1]
University Of Maryland, College Park, MD, 20742, USA

The return to planning and development of near-term human exploration missions beyond low Earth orbit has driven renewed investigation of low size, low mass, and cost-effective human spacecraft. However, very little experimental data on the effects of smaller cabin sizes on crew performance exists; that which does is mainly focused on microgravity habitation in low Earth orbit, and thus is not directly extensible to the Moon or Mars. The focus of this research is to experimentally investigate the impact of reducing habitat size on crew performance to determine the minimum effective habitat volume for future manned spacecraft. This paper summarizes ongoing research being conducted by the University of Maryland Space Systems Laboratory with support from the NASA Moon-to-Mars Exploration Systems and Habitation (X-Hab) program to investigate minimum effective habitat and spacecraft sizing, as well as results and conclusions to date for crew effectiveness within restricted cabin volumes under short-term, high-workload testing conditions. Utilizing modular resizable habitat mockups, tests in habitats ranging from 5 to 25 m^3 were conducted in simulated micro- and lunar gravities through underwater testing with body segment parametric ballasting, as well as surface testing in Earth gravity. The impact of size and configuration on crew effectiveness was measured by timed habitat translations, which are compared along with qualitative data to arrive at spacecraft sizing conclusions. While the underwater simulation environment prevents long-duration studies, thus not allowing for analysis of the psychological impacts of smaller habitat sizes, the short-term, high-workload human effectiveness in varying gravity environments can provide new insights into the sizing of future manned spacecraft designs.

I. Introduction

Historically, habitat sizing has been accomplished through the utilization of predictive methods such as the "Celentano curves"[1,2] or regression methods based on heritage designs such as those in the latest version of NASA's Human Integration and Design Handbook[3] (HIDH). While work, such as that by Rudisill et al.,[4] has sought to combine these and other methods with some limited experimental investigation, to date no method of habitat sizing rigorously based on experimental evidence has been devised, and none of the available methods directly take into consideration the impact of gravity on habitat sizing. As human spaceflight interest increases with the planned Artemis missions, eventual missions to Mars and (hopefully) beyond, and the rise of commercial human spaceflight, the issue of required habitat sizing remains ambiguous at best. Furthermore, the data which we do have is restricted mainly to low Earth orbit, where the cost of additional habitat volume is minimal compared to other solar system destinations. In the near-term future, essential design decisions regarding the required habitable volume for lunar and martian ascent/decent vehicles, pressurized rovers, and long-term space and surface habitats will need to be made, and investigation of these considerations is critical to future mission success.

Current guidelines predicated on regression fits of low Earth orbit data, where the cost of additional habitat volume is significantly lower than elsewhere in the solar system, are likely to drive systems to be oversized relative to the actual need, which risks mission viability. Oversized habitats also result in higher costs, have longer development times, and require larger launch vehicles. However, utilizing vehicles and habitats that are too small for crew health and functionality risk the mission itself. Spaces that are overly confining can negatively impact crew performance, mental health, and even crew safety. This study therefore aims to establish an experimental basis upon which these design decisions for future space vehicles/habitats and the first gravitational habitats can be made, especially outside of low Earth orbit where the higher cost of additional habitable volume may be the limiting factor for feasibility and affordability. Utilizing underwater testing with ballasted test subjects, testing was conducted in simulated micro

---

[1] Graduate Research Assistant, Space Systems Laboratory
[2] Director, Space Systems Laboratory. Professor, Department of Aerospace Engineering

and lunar gravity to analyze the impact of habitat size on test subject performance. While only short-term in nature, insights can still be garnered to influence future habitat sizing and also can form the basis for future, more involved testing programs.

This study is the continuation of research detailed in "Experimental Investigation of Minimum Required Cabin Sizing in Varying Gravity Levels".[5] As such, additional information, especially regarding the background of this research and details of previous work, can be found there.

II. Hardware Design

The foundation of this research is the creation of various sized habitats through the utilization of a set of twelve modular rack structures that are similar in concept to the International Standard Payload Rack (ISPR) modules used on the International Space Station (ISS). The exterior dimensions of the racks are approximately one meter in width, two meters in height, and a half meter in depth, and they are constructed from 1515-Lite series 80/20™ aluminum extrusion. Of these racks, three are outfitted with shelving for cargo transfer bag (CTB) simulators or other experimental components, one contains a deployable table, one contains a removable mass simulator similar in nature to an ISS Environmental Control and Life Support System (ECLSS) air filter, and the remainder are covered on the interior side by a plastic sheet to better simulate the psychological effects of being inside a small habitat. The mass simulator is capable of simulating a variety of masses by controlling the ratio of water to air within a contained pressure vessel.

The design of these racks focuses on modularity, ease of reproduction, and ease of modification to allow for rapid changes to the testing configuration. Through the modular rack structure, it is possible to simulate cylindrical, single-floor, vertically- and horizontally-oriented habitats as shown in Figure 1, with 6-12 racks allowing for testing internal volumes ranging from approximately 5-25m^3 as shown. The simple rack configuration allows for additional racks to be rapidly constructed, thus enabling the creation of larger sizes of habitats should they be desired in future testing. Additionally, the ease of modification allows for the expansion of the test matrix to include different testing methodologies through the utilization of current or future integrated simulation tools, such as the mass simulator and CTB shelves.

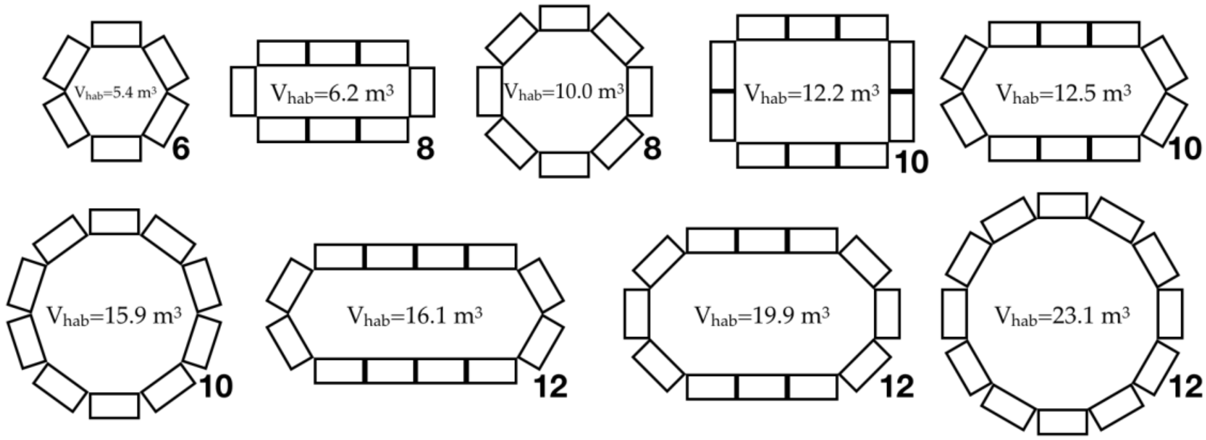


Figure 1. Variation of internal volume based on rack module configuration, shown as top views. Numbers in bold refer to number of racks in each configuration. Circular sections represent vertically-oriented habitats.[5]

Racks remain independent until the time of testing, where they are attached to each other via both internal angle and external angle locking bars that attach to posts on the top corners of the rack structures. The internal locking bars hold the interior corners of adjacent racks together, while the exterior bars complete the triangle formed by the gap between adjacent racks and constrain the angle required to achieve the required polygon. Once all bars are in place, the result is a highly rigid and stationary overall structure that can serve as both gripping and push-off surfaces for translations during microgravity simulations.

For the purposes of the testing detailed in this paper, only vertically-oriented (cylindrical) habitats with six, eight, ten, and twelve racks were tested due to limitations in conducting underwater testing (described in Section III-B). Vertical habitat orientations were chosen after determination between the team and NASA reviewers during the first phase of this research, which was conducted as part of the NASA 2020 Moon to Mars Exploration Systems and

Habitation (X-Hab) Academic Innovation Challenge program,[5] that these habitats are indicative of the current design focus of NASA efforts, and are thus more representative of future habitat designs. While the habitat size range can be as small as three modules and theoretically as large as desired, the number of racks was chosen to be within the range of 6-12 racks. This was the result of preliminary testing which determined that configurations smaller than six racks were too difficult to test in safely, and that configurations larger than twelve racks (approximately 5m in outer diameter) were oversized relative to current habitat design efforts and the current testing methodology. While several habitat designs are larger in volume than the 12 rack configuration, the majority consist of several modules as opposed to a single large module in order to conform with current launch vehicle payload size limitations; the analog for these habitats in this testing methodology would be analysis on the individual module sizing as opposed to the overall habitat itself. The remaining configurations, consisting of eight and ten racks, were chosen to span the volume range within the number of tests that were predicted to be achievable in the current testing phase.

III. Underwater Testing Capabilities and Considerations

A. Body Segment Parametric Ballasting

By conducting underwater testing, it is possible to simulate different gravity environments by controlling the force balance between buoyancy and (Earth) gravity. This is accomplished by adding additional weights to subjects through body segment parametric ballasting to achieve the desired gravity effect. Ideally, each body segment would be negatively ballasted according to anthropometric proportions; however, this would result in large, cumbersome weight schemes restricting subject mobility and comfort. A simpler scheme focusing on torso, thigh, and ankle distributions of 62%, 13%, and 6% of total body mass respectively was found to be an appropriate substitution with little kinematic loss.[6] The torso distribution was split evenly between the chest and upper back. Figure 2 shows a pair of ballasted test subjects. This system allows the test subjects to retain a similar center of gravity and provides minimal adjustments to nominal leg walking dynamics beyond hydrodynamic forces,[5,7] which is further explained in section III-B-4.

Figure 2. Test subjects wearing body segment parametric ballast to simulate lunar weight. Weights are located on the chest, upper back, thighs, and ankles.

B. Limitations of the Underwater Environment

While the utilization of the underwater environment enables the simulation of various gravity levels that would otherwise be impossible for this testing, there are several limitations that must be discussed due to their significant impact on the data and conclusions that can be drawn from it.

1. Limitations to Potential Subject Pool

While a large number of subjects would have been desirable so as to eliminate several biological factors such as physical fitness, age, gender, etc. from the data, the requirement that subjects be divers certified by both a recognized diving certification agency and the UMd Diving Control Board drastically curtails the number of subjects that can be utilized. Furthermore, any testing utilizing body segment parametric ballasting requires that divers be full face mask certified to enable verbal communication both with other divers and with surface personnel. This is done for safety reasons, as the drastic reduction in underwater mobility from being weighted necessitates faster response from safety divers and the deck chief in the event of an emergency situation. The ability to communicate between test subjects also results in a more realistic test, as a space habitat crew would be able to communicate and coordinate verbally with each other. The COVID-19 pandemic resulted in a two-year suspension of the UMD diving certification process, so for the testing period covered in this paper only three people had the proper certifications to qualify as test subjects for this research. The certification process has only recently resumed, so it is hoped that in the future it will be possible to conduct additional testing. However, at present, only one subject group was available for the purposes of this testing.

2. *Required Testing Overhead*

Additionally, unlike testing in a standard laboratory environment, the overhead for underwater testing is highly constraining. For an underwater test with two weighted subjects, at least ten total personnel are required: two subjects, two safety divers per subject, a photographer, a deck chief, a test conductor, and a data collector. Additionally, dives necessitate a significant increase in setup and breakdown time, as all divers must check, prepare, and put on their equipment, the habitat racks must be lowered into the tank and then assembled underwater, and divers must put on their testing weights (which can only be done at the bottom of the tank); this must all then be repeated in reverse during breakdown. Overall, the setup and breakdown process adds approximately 3-4 hours to a test, while comparable surface testing only requires approximately 1 hour. This also does not include the need for changing air tanks for seven divers, two of whom are utilizing full face masks which consume air at a greater rate due to increased leakage, which adds additional time overhead. As such, the person-hour cost of a single dive is, at minimum, 50 hours for a two hour test. While several tests are conducted during one dive when possible to maximize the testing to setup/breakdown time ratio, it was found that (after the setup and breakdown time) two hours of testing was the limit of diver endurance. This is due to the highly fatiguing nature of diving, especially when subjects were weighted; thus, it is often only possible to test two habitat configurations per dive. The high time cost combined with the low number of certified divers available, and the resulting need to coordinate very specific schedules, means that it was only possible to test once a week, and more often limited to twice a month.

3. *Limitations to Length of Testing*

As mentioned briefly, underwater testing is substantially limited by diver endurance. In the course of this testing, dives were often limited to two hours worth of testing before divers grew tired enough to noticably affect performance, though it should be noted subject endurance would likely improve in a more free-form activity than the highly repetitive directed motions of the current test protocol. Additionally, while the Neutral Buoyancy Research Facility at UMd has sufficient air processing capability to sustain dive operations indefinitely through tank changes, underwater testing prevents test subjects from eating, drinking, taking extended breaks, or any other functions that are usually required for longer testing. As such, long-term testing to analyze psychological effects of subjects across internal habitat volumes becomes highly impractical without the inclusion of appropriate break procedures.

4. *Hydrodynamic Drag*

One of the largest impacts to the fidelity of the testing as compared to the real gravitational environments that are being simulated is the impact of hydrodynamic drag on subject motion. Not only does the presence of this drag decrease the translation rates of the subjects, but in many cases it fundamentally changes motion around the habitat itself. In microgravity, translations linearly across a habitat are possible by pushing off of a surface due to the lack of any appreciable resistive forces to slow the motion of the astronaut. However, such motions are nearly impossible in the underwater environment without the addition of a swimming motion on the part of the subjects (which would be a significant departure from true microgravity motions). As a result, the majority of all microgravity translations occurred circumferentially with hand-over-hand propulsion around the racks forming the habitat structure.

Similarly, hydrodynamic drag likely reduces the impacts of weight differences between lunar and martian gravity. Since the drag term is so large, the majority of the effort on the part of the subject is dedicated to overcoming this resistance. As a result, subjects are unable to take full advantage of the additional traction that martian gravity provides nor the reduced effort required in lunar gravity. Thus, lunar and martian gravity motions are likely more similar in the underwater environment than they otherwise would be. It is also noteworthy that the gait of subjects in the underwater testing was extremely similar to walking in Earth gravity and lacked the skipping/bouncing motion that was preferred during the Apollo missions. While the testing subjects tried to simulate these or similar gaits at times, it was found that the drag made these highly impractical and greatly increased the physical and cognitive exertion; thus the subjects were allowed to adopt whichever gait suited their preference during testing. It is still unclear if these preferred lunar gaits were a function of the environment, the Apollo spacesuits, or both,[8] but they were nonetheless not possible to replicate in these tests due to the extreme effort required to overcome drag.

5. *Umbilical Management*

In order to achieve higher fidelity testing, subjects were attached to their air supplies via a “hookah” system, supplying breathing air from a remote tank via a 30-ft hose. This was done so that their bulky air tanks could be stationed outside of the habitat, and an air supply umbilical would be long enough to freely accommodate subject motion inside the habitat. In this manner, the typical buoyancy control device and 80 ft^3 air tank would not affect the subject's center of gravity or gait. While it was to a lesser effect than the alternative, it was still observed that the umbilicals were affecting the subject's path planning: subjects would occasionally take non-ideal paths simply to avoid the risk of tangling lines, especially in the consecutive motion tasks (described in Section IV). Although a limited amount of tangling was permissible due to the amount of slack in the umbilicals, significant tangling could pull on the face masks and cause water entry which created a significant safety concern. Because of this, subjects would start searching for a path while directions were still being given, or they would sometimes delay beginning their motion until the other subject made it clear in which direction they planned on moving. The effects of path planning prior to motions were magnified for the subject who was first to receive directions, as they had additional time to visualize their path before they were allowed to begin moving.

The cognitive effort required for maintaining umbilical awareness that would not otherwise be present in a space-based habitat may have an unknown impact to the test results. However, the use of an auxiliary diver managing the subject's umbilicals from above the habitat appeared to lower this effort during practice runs. With the hovering diver taking up the slack in the umbilicals and keeping them out of the habitat volume, they were less prone to tangling. Additionally, any tangles created were kept from pulling on the subjects' face masks. These actions, or perhaps the knowledge that someone else was monitoring/preventing tangling, appeared to reduce the subjects' delays and planning prior to motion, especially during consecutive motion tasks. This improved umbilical handling strategy was not used in tests following its discovery so as to not affect any trends in the data, but it will be used in the future to re-run tests, compare results, and determine its efficacy.

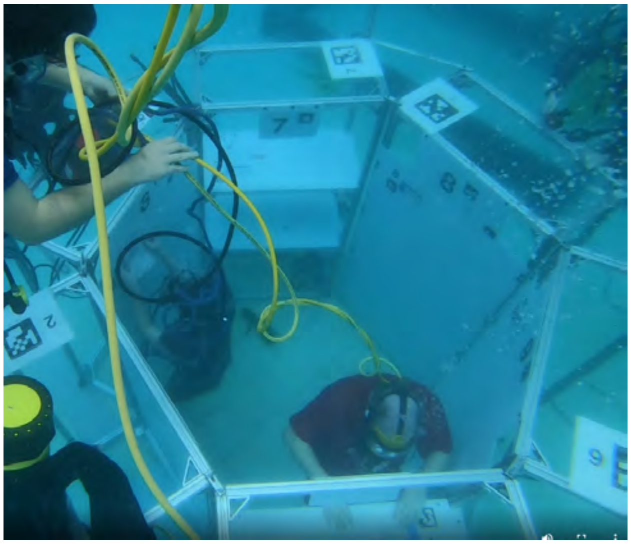


Figure 3. Untangling of the test subject umbilicals.

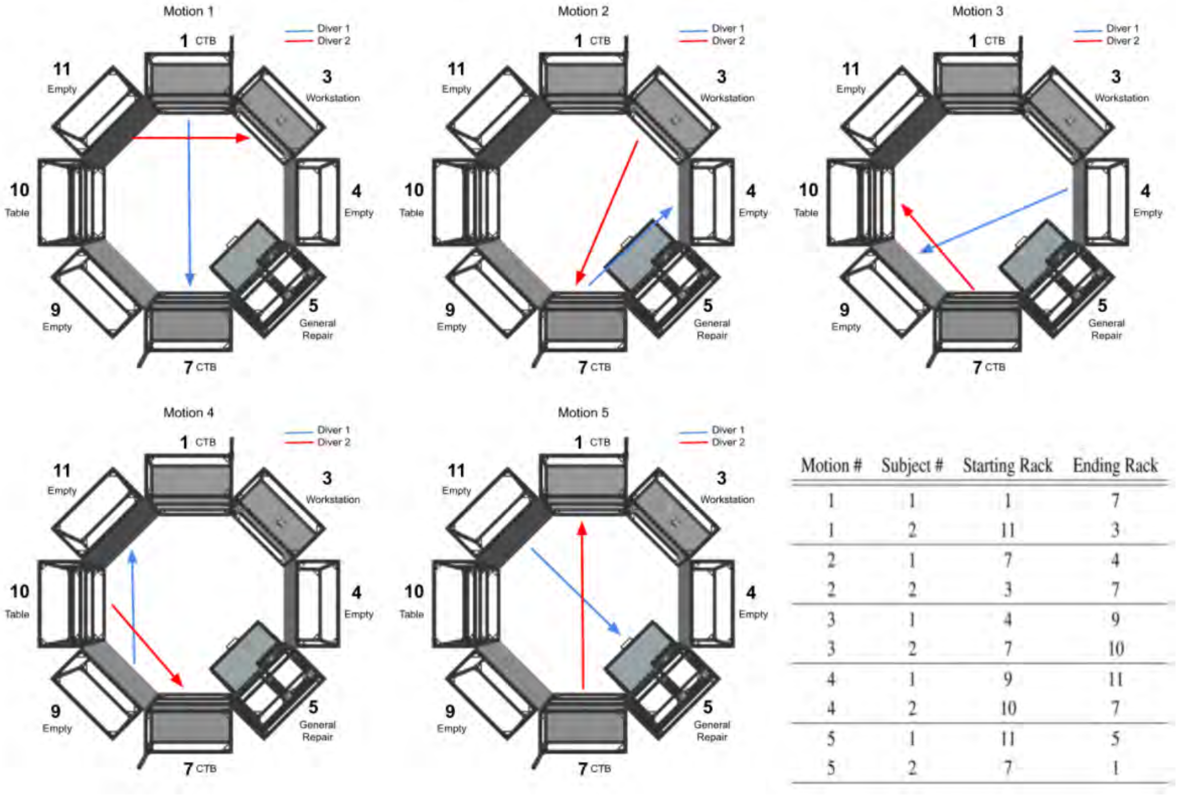
IV. Experimental Design

A. Task Description

The tasks consist of a series of predetermined habitat translations between rack modules. There are seven standard sequences that are completed per habitat size and gravity configuration within the test matrix, each of which consists of either five or ten motions per test subject. The sequences are conducted as follows:

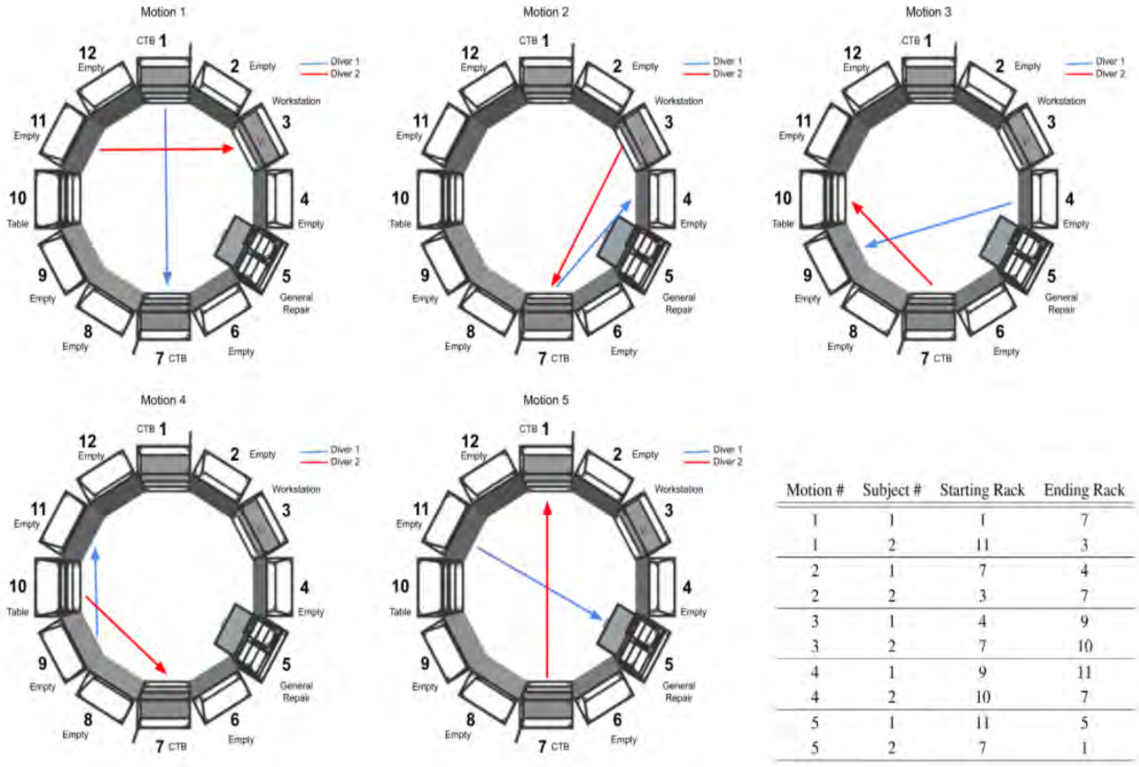
1. Both subjects, 5 standardized scripted motions each
2. Subject 1 only, 5 standardized scripted motions
3. Subject 2 only, 5 standardized scripted motions
4. Both subjects, 5 standardized scripted motions each
5. Both subjects, 10 randomized scripted motions each
6. Both subjects, 5 standardized consecutive motions each
7. Both subjects, 10 randomized consecutive motions each

These sequences are broken into two categories: scripted motions and consecutive motions. In the scripted motion testing, the destination rack is read aloud to the subjects by the test conductor in real time. Both subjects begin each individual motion at the same time, at the end of the test conductor reading the commands. The purpose of this testing is to artificially ensure specific conflicting motions occur independently of the speed that subjects complete the tasks. For the consecutive motion sequences, subjects are given a task sheet that provides all destination racks for the entire sequence and are allowed to move at their own pace. This testing is included to simulate a more realistic habitat task. Additionally, within each category, there are both standardized and randomized sequences. Standardized sequences



Motion #	Subject #	Starting Rack	Ending Rack
1	1	1	7
1	2	11	3
2	1	7	4
2	2	3	7
3	1	4	9
3	2	7	10
4	1	9	11
4	2	10	7
5	1	11	5
5	2	7	1

Figure 4. Scripted sequence visualization in an eight rack configuration. Subject 1 in blue, subject 2 in red.



Motion #	Subject #	Starting Rack	Ending Rack
1	1	1	7
1	2	11	3
2	1	7	4
2	2	3	7
3	1	4	9
3	2	7	10
4	1	9	11
4	2	10	7
5	1	11	5
5	2	7	1

Figure 5. Scripted sequence visualization in a twelve rack configuration. Subject 1 in blue, subject 2 in red.

consist of the same motions for each sequence, as closely as can be replicated in the different habitat sizes. As such, there are four distinct standardized motions in the current testing, one for each habitat size, but they are structured such that the motions being done are as similar as possible between configurations. This ensures consistency in comparing the datasets between dissimilar configurations and conditions. Figures 4-5 show the standardized sequence motions in the eight and twelve rack configurations, respectively. Randomized testing, by contrast, is pseudo-random for each test in order to ensure that the subjects cannot utilize prior knowledge of the testing motions to influence their behavior. Note that the randomized motions are not the same in sequences five and seven above, as both are generated independently. While this methodology was designed to be expanded to four subject testing in the future, all testing to date has utilized a two subject configuration.

The pseudo-randomness as opposed to true randomness comes from two factors. First, the randomization is accomplished by computer which, while seed numbers are changed each time, are inherently only pseudo-random. Second, certain motions are not allowed, namely motions where the subject stays at their current rack or where both subjects are sent to the same rack. While a subject staying at the same rack may be more realistic, these are removed in order to keep the subjects moving within the habitat due to the compressed nature of the testing that is achievable. Similarly, while both subjects needing to go to the same rack may be more realistic, it was determined that the increased realism was outweighed by the increased difficulty in the data analysis that was introduced by needing to account for this inconsistent delay; thus, any randomly generated rack numbers that resulted in one of those cases were replaced to achieve higher consistency between otherwise similar data points.

At each rack, a 6x6 grid of four digit numbers was printed, with the rows being identified by the numbers 1-6 and the columns by letters A-F. In addition to the destination rack, the subject is given a number position to read as part of the task. For example, a command given by the test conductor during the scripted portion might be “Subject one go to rack eight and record delta-two. Subject two go to rack twelve and record alpha-six.” After the full command is given, time begins and both subjects start their task. Once they have finished recording the number, the time ends. In practice, the recording of the number was removed due to the inconvenience in recording values underwater and the limited testing value it was deemed to have, but subjects were still required to find the correct panel position before signaling their task was complete. For the consecutive motion sequences, panel positions were printed on the task sheet next to the destination racks. For all motions and sequences, including for the standardized testing, these panel positions to be recorded were randomized.

B. Data Collection

Two methods have been used to collect data during this testing. Initially, data was collected by human operators utilizing stopwatches. For the scripted testing, this was measured from the completion of the entire command by the test conductor for both subjects to the end of the motion as indicated by each subject. This indication was done verbally when utilizing underwater communications or via a hand wave when this was otherwise not possible. During the consecutive motion trials, the ending time was the same as the scripted testing, while the starting time began immediately following the end of the previous motion time (except for the first motion, which was started at the end of the countdown from the test conductor). However, this collection method was determined to be non-ideal due to both the time-intensive nature of the data collection, and the introduction of human error that resulted.

As such, a data collection system was created to generate the timing data. This was accomplished by the addition of start and end AprilTags[9] to each of the racks. Utilizing a Raspberry Pi Zero v1.3 and a Raspberry Pi Camera Module v2.1 running AprilTag recognition software, subjects are able to point the system at the start AprilTag of their starting rack and then the end AprilTag at their destination rack to generate the timing data. The Pi records both the time duration between the reading of the two AprilTags and the rack number for each.

Due to the requirement that this system be able to operate underwater, a method of waterproofing the electronics was required that simultaneously allowed for fast, repeated access to remove SD cards and change batteries. Accessibility is required since, for longer tests, battery changing needs to occur in the middle of a dive. As such, the entire system was designed to fit within the waterproof housing of a GoPro HERO9 camera. This allows for the transparency required by the camera, the size required by the Pi and its associated battery, and the fast and easy access to the system. Within the housing, the camera module is held in place by a 3D printed structure to ensure it remains in position within the camera volume regardless of the movement of the subjects. Furthermore, the utilization of a housing intended for cameras also allowed for the addition of an integrated pushbutton so that the subjects can manually trigger the recording of the AprilTag. Some subjects found this mode preferable as compared to having to ensure that the camera module did not point at any other AprilTags and thus generate an incorrect reading. However, the camera and AprilTag method was still maintained as opposed to simply recording the time between pressing of the pushbutton to ensure that accidental button pressing did not compromise the data. Since the time does not end unless the system sees an end AprilTag, it helps to prevent potential errors from propagating through the test. Additionally, the recording of the rack numbers allows for post-processing to ensure that subjects went to the correct racks, as in the event of an error the data was still able to be utilized by correcting for the difference in distance traveled.

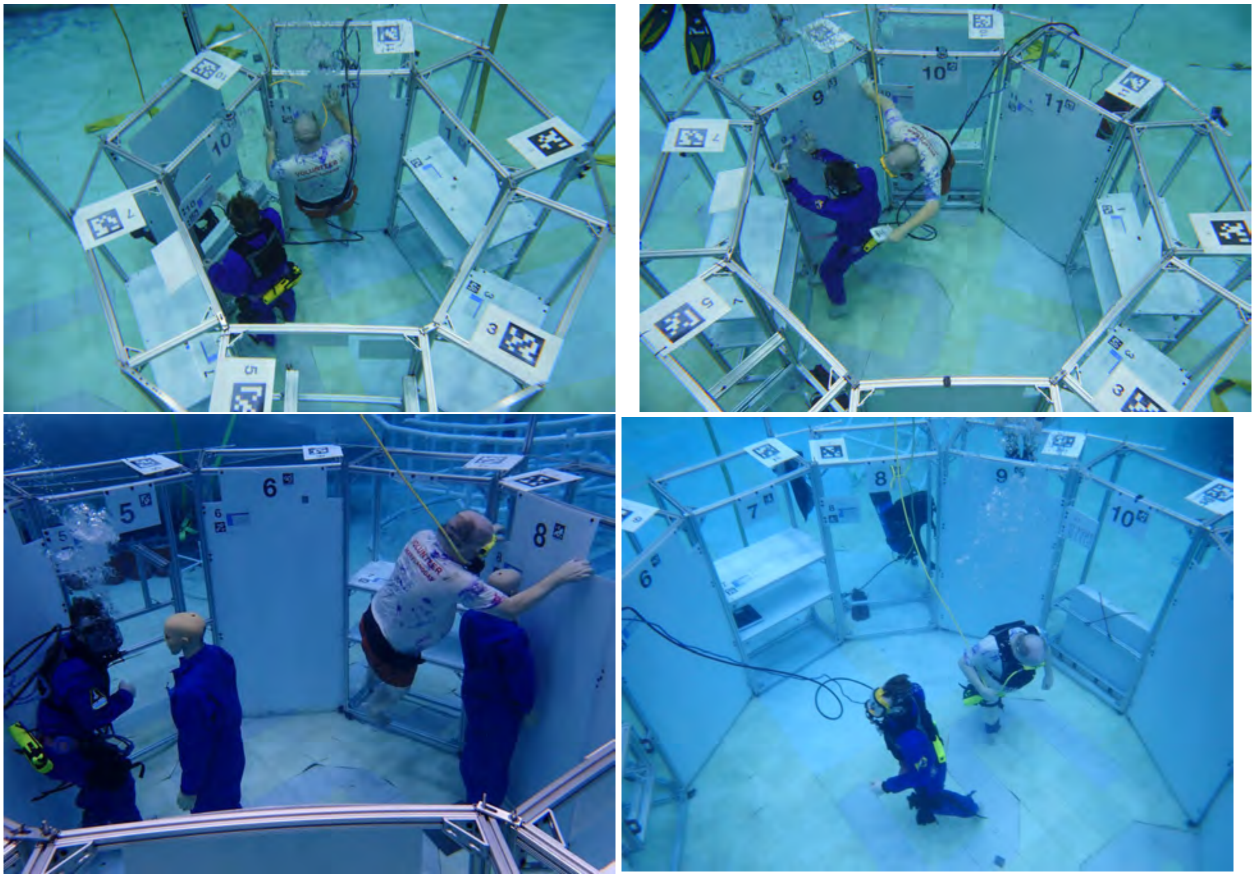


Figure 6. Testing of vertical habitats formed by 6- (upper left), 8- (upper right), 10- (lower left), and 12- (lower right) rack configurations.

C. Testing Matrix

The original test matrix at the beginning of this phase of testing included four habitat configurations and three underwater gravity levels. The four habitat sizes consisted of the six, eight, ten, and twelve rack vertical configurations, as shown in Figure 6, with the three gravity levels being micro, lunar, and martian gravity. As previously described in Section III-B, the limitations of the underwater environment and the ongoing COVID-19 pandemic resulted in a matrix consisting of only one subject group. Additionally, between the overhead required for dives, the pandemic necessitating a reduced diving schedule due to personnel limitations, and other ongoing projects requiring underwater testing time, it was determined part-way through testing that only one full iteration through the original test matrix could be achieved. In an effort to increase the number of repetitions for at least part of the testing, martian gravity testing was removed from the testing matrix so that more effort could be devoted to the other two gravity levels, namely in allowing for another repetition of the lunar testing. While two sets of testing data is not enough to draw conclusions regarding statistical significance, repeating the lunar gravity testing was nonetheless determined to be worthwhile to help establish a (very) preliminary understanding of the deviations that can be expected in the data when future testing is possible. Lunar testing was deemed preferable to martian testing due to the higher expected utility of this research resulting from efforts underway for the Artemis program at the time of writing.

However, at the time of writing it was only possible to achieve full inclusion of the 6 and 8 rack configurations, a majority of the 12 rack configuration, and limited 10 rack configuration data. The investigation of the 12 rack configuration lacks the inclusion of random and consecutive motion testing for both gravity levels, and the 10 rack configuration was only tested with a single weighted test subject due to equipment limitations (see Section IV-D below for more details). While it is planned to conduct the requisite testing to complete the testing matrix as planned for this phase of the research, these equipment and logistical limitations prevented this by the publication of this paper. However, the data for the 6 and 8 rack configurations consists of two full iterations for both gravity levels and partial data from two (6 rack) and three (8 rack) other tests.

D. Other Testing Investigations Explored

In addition to the previously mentioned testing sequences (Section IV-A), earlier tests also included three additional standardized scripted sequences: both divers with a table, both divers with two mannequins, and both divers with both a table and two mannequins. These tests occurred between sequences four and five above. Additionally, sequence five (the randomized scripted sequence) included both the table and the two mannequins, and these tests did not include the consecutive motion sequences (as it had not been developed at the time). The table measured 72”x30”x29” (length, width, height) and was placed into the habitats such that the length dimension was in the radial direction. The mannequins were placed standing in front of random racks within the habitat. The concept of the mannequins was to replicate the presence of additional, stationary subjects to simulate crew members who may be working at a single rack while these other tasks are occurring. Mannequins were preferable over additional subjects due to diver limitations, as described previously, and COVID-19 concerns resulting in the need to minimize the number of people involved in testing. However, after several tests utilizing these additional obstacles, it was determined that the effect of these obstacles was not especially interesting: more obstacles in the habitat unsurprisingly increases the motion times. It was thus determined that these tests were not worthwhile to continue due to the time involved and the limitations that underwater testing entails (as described in Section III-B). Since the limitations to number of tests precluded running these sequences enough times to draw more meaningful conclusions, they were removed from later tests to reduce the required overhead of the overall testing program. As such, while a brief description of these results is discussed in Section V, the main focus of this research, and thus this paper, is on the later testing that omitted these sequences. For tests where these sequences were included, these results were discarded prior to inclusion with the other data for consistency (including the random trials of these tests, as they included the mannequins and table and are thus inconsistent with the later tests that did not include these).

Additionally, diver equipment limitations and associated safety concerns prevented both test subjects from being weighted for the initial testing that was conducted. Thus, several of the initial tests had subject 1 weighted at lunar or martian gravity and subject 2 at microgravity. It was hoped that the errors this mixed gravity methodology introduced could be normalized later in the study to allow for this data to be utilized, but the limitations to the total number of tests prevent this at present. Once the equipment issues were resolved and safely conducting dual-subject weighted testing was possible, it was decided to focus this research on the results from these later tests. As such, this paper focuses almost exclusively on the results from these later tests. However, since the individual testing in sequences 2 and 3 are unaffected by mixed gravity impacts, as only one subject is present at a time for these tests, these results are incorporated where possible. In the future, it may be possible to revisit and successfully normalize the remainder of this data to allow for its use in the study.

E. Earth Gravity Surface Testing

In addition to the underwater testing, a limited series of Earth gravity surface tests were also conducted. This testing similarly included only one subject group due to Covid-19 concerns at the time, as this testing occurred in the middle of a large spike in the number of cases and a return to higher university pandemic restrictions. Subjects were allowed to wear their preferred clothing (with the added requirement of a KN95 or N95 face mask for Covid-19 prevention), and as such did not utilize the same equipment as is required underwater. This may have an unknown impact on the comparative results between the underwater and surface testing, but it is not possible to outfit the subjects similarly: floating umbilical lines are not possible except underwater and the impact of the diving weight system, which does have an impact on range of motion underwater, would result in surface subjects being significantly weighed down and curtail motion much more significantly than it does underwater as a result. Mannequins and the table were still included in the testing matrix at the time of this testing; thus, the random sequences of this program included these elements and thus cannot be directly compared. Furthermore, while these tests did include consecutive motion trials, these too included the mannequins and table. As a result, for the purposes of this paper, only sequences one through four are utilized and compared from this testing program.

V. Results to Date

A. Testing Summary

Table 1. Summary of dives conducted to date (CM = consecutive motion).

# of Racks	Orientation	Subject Gravity	Testing Description
6	Vertical	2 Micro	Qualitative assessment
8	Vertical	2 Micro	Qualitative assessment
8	Vertical	2 Micro	Qualitative assessment
10	Vertical	1 Micro, 1 Lunar	Qualitative assessment
6	Vertical	1 Micro, 1 Martian	Qualitative assessment
8, 10, 12	Horizontal	2 Micro	Qualitative assessment
12	Vertical	2 Micro	Qualitative assessment
8	Vertical	1 Micro, 1 Lunar	Table testing, no CM testing
12	Vertical	2 Lunar	Table and mannequin testing, no random or CM testing
8	Vertical	1 Micro, 1 Martian	Table testing, no CM testing
12	Vertical	1 Micro, 1 Martian	Table and mannequin testing, no CM testing
10	Vertical	1 Micro, 1 Martian	Table and mannequin testing, no CM testing
10	Vertical	1 Micro, 1 Lunar	Table and mannequin testing, no CM testing
12	Vertical	2 Micro	Table and mannequin testing, no CM testing
8	Vertical	2 Micro	Table and mannequin testing, no CM testing
8	Vertical	1 Micro, 1 Lunar	Table and mannequin testing, no CM testing
6	Vertical	1 Micro, 1 Lunar	Table and mannequin testing, no CM testing
6	Vertical	1 Micro, 1 Martian	Table and mannequin testing, no CM testing
8	Vertical	2 Micro	CM testing only
6	Vertical	2 Lunar	Finalized sequence
8	Vertical	2 Lunar	Finalized sequence
6	Vertical	2 Micro	Finalized sequence
8	Vertical	2 Micro	Finalized sequence
8	Vertical	2 Lunar	Finalized sequence
6	Vertical	2 Lunar	Finalized sequence
6	Vertical	2 Micro	Finalized sequence

Table 1 shows all of the tests that were conducted as part of this study as of the time of publication, and provides associated descriptions of each test. As can be seen, initial testing focused on qualitative assessment of habitat sizing, which was utilized in conjunction with input from NASA reviewers to influence both the final test matrix and the sequences and tasks that were utilized in later testing. Some results of this initial testing were described by Akin et al.[5] All other tests utilized the task methodology described in this paper, although as previously described some of the testing included sequences with mannequins and/or a table which were later omitted and the consecutive motion tasks were created and added later in the testing program.

For all data detailed below, the total time for each subject in a given sequence is normalized by the distance of the total number of racks transited by the subject in that sequence to arrive at the rate data used for this analysis. Note that this rate is analogous to the inverse of velocity; therefore, lower values are indicative of higher task efficiency.

B. Table and Mannequin Testing Results

While not the focus of this research, Table 2 shows the results of the table and mannequin testing. This table only includes tests where both subjects were at the same gravity level, as the majority of the testing with these elements occurred when only one weighted subject was possible as previously described. All data presented utilized the standard scripted tasks, and all habitats were in the vertical orientation. For each of the data columns in Table 2, the data is presented in the form [subject 1 value, subject 2 value]. The ”Empty” column corresponds to sequence four, and the other sequences were conducted consecutively in the order presented in the table before the random testing in sequence five. As can be seen, generally the mannequins provided a significant obstacle in microgravity. This is because circumferential motions are required in the microgravity testing; thus, the mannequins could not be easily navigated around. In the weighted testing, the mannequins were a much less significant obstacle, showing very little distinction between the empty test and the mannequin test. Similarly, the effects of the table were also more pronounced in microgravity as opposed to gravity, although the mannequins were more impactful in microgravity than the table due to the ability to travel over the table in a weightless environment. While a weighted participant could have vaulted over the table in practice, subjects were instructed not to do so. Finally, the table and mannequin testing shows more significant impacts overall than all of the other sequences, as there are physically more obstacles impeding motion. These results were not deemed to be especially interesting as they are consistent with intuition, which is why they were omitted from later tests.

Table 2. Results from table and mannequin testing. Values given are total time / total number of racks transited (seconds per rack) and are of the form [subject 1 value, subject 2 value].

# of Racks	Subject Gravity	Empty	Table Only	2 Mannequins Only	Table + 2 Mannequins
8	2 Micro	[3.44, 3.06]	[4.06, 3.29]	[4.72, 3.90]	[5.22, 3.68]
12	2 Micro	[3.46, 3.49]	[3.50, 3.49]	[4.81, 4.51]	[4.62, 4.21]
12	2 Lunar	[2.78, 3.31]	[3.13, 2.95]	[2.93, 3.53]	[3.46, 2.89]

C. Habitat Size Results and Analysis

To determine the impact of habitat size in the different gravity environments, the data is compared across the tested rack configurations to arrive at sizing conclusions for microgravity, Figures 7-8, and lunar gravity, Figures 9-10. For configurations with multiple different tests, the total times and number of racks transited were aggregated and then divided. This was done as opposed to averaging results from different tests for simplicity due to the low number of tests precluding any meaningful analysis of standard deviation and other statistical quantities.

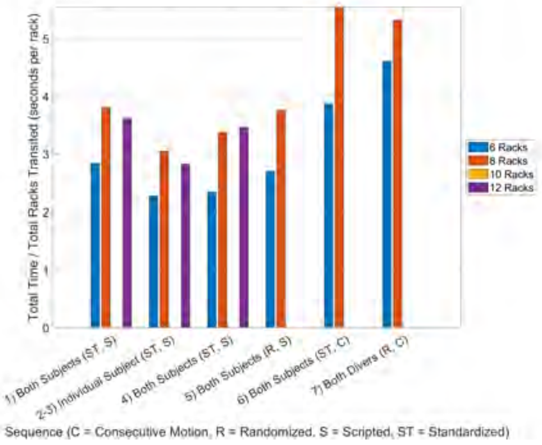


Figure 7. Microgravity timing data for subject 1.

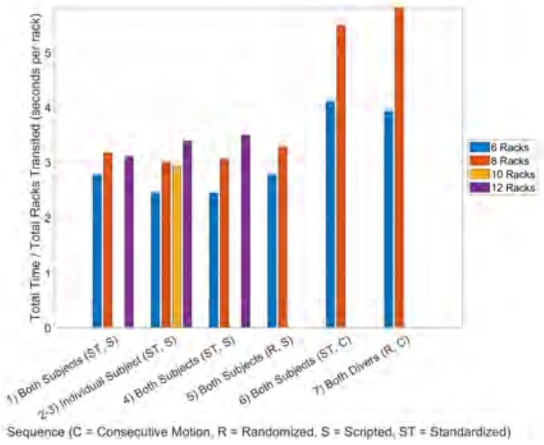


Figure 8. Microgravity timing data for subject 2.

With the caveats regarding statistical significance previously mentioned, this data suggests that smaller habitable volumes are better suited to the short-term, high motion tasks that were the central focus of this testing. As a result, it can be concluded that based on the current testing, the shorter motions that are required for smaller habitats outweigh the negative impacts of greater motion conflict between subjects in short-term cases. Furthermore, it is noteworthy that both the total translation times for all habitat sizes and the magnitude of the difference in translation rate between the different sizes appear to be greater for the microgravity cases as compared to the lunar gravity cases. The data therefore indicates that navigation of the habitats becomes more difficult in microgravity compared to lunar gravity as habitat size increases. This result is logical considering that, in this testing where hydrodynamic drag is a major factor, direct motions are only possible in the lunar case whereas microgravity necessitates longer circumferential motions

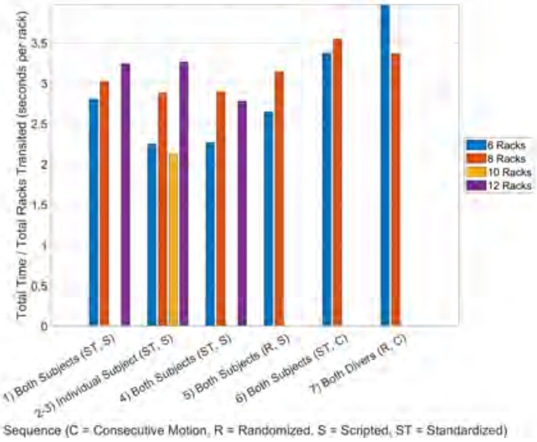


Figure 9. Lunar timing data for subject 1.

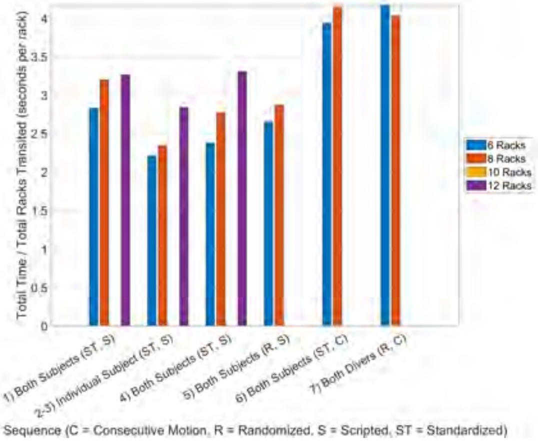


Figure 10. Lunar timing data for subject 2.

(as circumferential distance grows more quickly with increasing size than direct distance). It is unclear whether this difference between the lunar and microgravity cases would be entirely eliminated without the presence of drag, but this is likely a significant contributing factor to these trends.

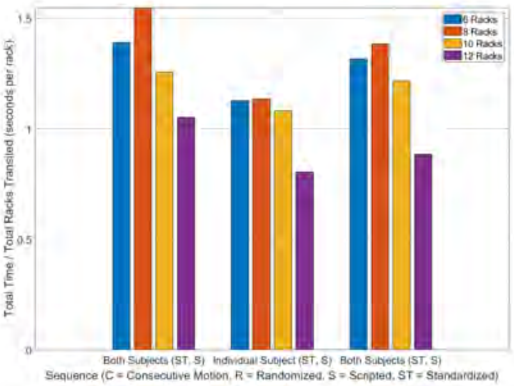


Figure 11. Earth gravity timing data for subject 1.

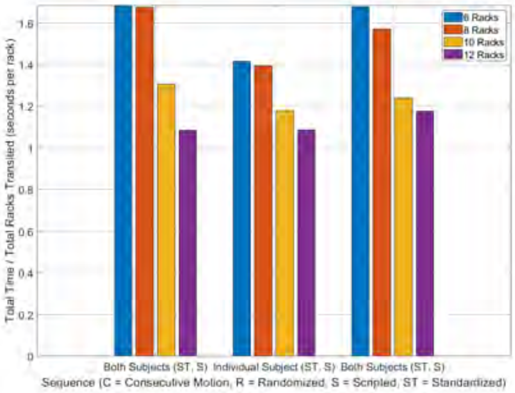


Figure 12. Earth gravity timing data for subject 2.

Additionally, in the majority of cases subjects showed significant timing improvement between sequences one and two/three and performance reduction between sequences two/three and four, indicating that the motion conflicts between subjects are a significant negative impact on crew performance in small space habitats. As such, this research suggests that the work environment and interaction of the crew are a significant factor in the sizing of habitats for crew efficiency: habitats where crew members will be working individually and/or can be artificially kept out of the way of each other through control of the tasks can be kept smaller than those where these cannot be ensured while still maintaining crew efficiency. However, in all cases subjects performed significantly faster in sequence four as compared to sequence one, which are identical except for their position in the overall test sequence. This shows an apparent strong impact from a short-term learning curve in the standardized scripted testing. Some underwater tests showed up to a 20% reduction in the required time per rack transited, with the average being approximately 10% improvement. It is theorized that this is the result of subjects learning both their own motions and the motions of the other subject and thus being able to plan their motions with greater efficiency. While more data would be needed to make a definitive claim, this does suggest that crew members in small habitats where the daily tasks are relatively repetitive can learn and adapt to the confined space extremely quickly, thereby recovering a significant amount of the efficiency lost due to motion conflicts with other personnel. It is important to note however that, for each motion, there was only the potential for a singular conflict to occur. In practice, conflict would likely be occurring with higher frequency and less regularity as crew live and move throughout a habitable space; thus this simulation methodology appears unable to fully capture these effects. Additionally, the presence of drag slows motion quite considerably, which may make adaptation to the presence and motion of others more efficient due to allowing for slower reaction times. This latter factor is explored through the Earth gravity testing comparisons.

The dry-land Earth gravity testing, shown in Figures 11-12, show the opposite trend: as habitat size increases, translation rate generally decreases. There are two factors that could contribute to this apparent difference in trend. First, it is possible that the underwater environment is reducing the impact of motion conflict between subjects. Since motion is significantly slower underwater, if the amount of time a conflict consumes is relatively consistent between the testing, then it would constitute a substantially lower proportion of the total translation time underwater as compared to the surface testing. Alternately, this could be a result of the drag increasing the reaction time available to subjects. Thus, unlike in the underwater testing, the relative reduction to the inter-subject motion conflict that larger habitats provide could be outweighing the increased translation distances in the surface testing. Second, this could also be an unanticipated artifact of the metric being used to analyze the data. The normalization used is the number of racks translated, which is analogous to a circumferential distance. However, subjects in a gravity environment move in direct motions: thus it is possible that the normalizing metric is growing faster than is representative as habitat size increases in the Earth gravity case. Despite this possibility, it was determined to maintain the racks transited as the normalization metric due to its utility in the underwater testing, especially in allowing for comparison between microgravity and lunar gravity data. Since this data is the focus of this research, the possible error that this may introduce in the Earth gravity data was deemed to be acceptable in the current research.

Additionally, the surface Earth gravity testing also does not show the same large improvement to performance between sequences 1 and 4. While some improvement is seen in most cases, the performance between the two sequences is much more similar than in the underwater environment, and certain configurations even show a performance decrease. As a result, this data suggests that, unlike in the underwater environment, the impact of motion conflict is not as easily adapted to. However, it is important to note that Earth gravity surface testing can be conducted at a faster translation pace than would likely be achievable in microgravity or lunar gravity habitats due to the higher translation rates that are possible in a higher gravity environment and the extreme familiarity and efficiency with which test subjects can move around on Earth, given that this is the environment where humans have naturally evolved and live, as compared to motion in space. Thus, it is theorized that this difference between surface and underwater testing is because of the significantly faster speeds that are achievable in surface testing without the presence of hydrodynamic drag, which decreases the relative effect of deviations to the optimal path on the timing (and thus rate) data. While some tests do show a loss of efficiency, the lack of high repeatability means that these could be the result of external or otherwise uncontrolled factors that are independent of the learning curve over an individual test. Thus, the conflicting trends between the two methodologies cannot be immediately reconciled; more testing is needed to further explore the nature of these dissimilar trends to determine the causes and where each methodology provides higher accuracy to the space environment as it pertains to this investigation. However, in both the surface and underwater testing, the net change between sequences one and four was similar compared to others in the same environment even after multiple tests on the same day. This indicates that a change to habitat and/or gravity configuration may be enough to necessitate subjects re-learn optimal motions and thus reset, at least partially, the effect of any short-term learning curve present in the test.

Compared to the scripted motions, the consecutive motion tasks show a significant decrease to efficiency. Due to the design of the scripted motion, the test subjects have some portion of time that they are able to utilize to plan their motions as the test conductor finishes the command. While this was discouraged prior to testing in the scripted motion, the consecutive motion trials better measure the impact that the need to determine the optimal path has on crew efficiency. As such, this data is likely more representative of the overall effort required to navigate the different habitat sizes due to factoring in the cognitive impacts. However, due to the nature of how the timing was collected for both test types, the time required to read the task sheet is included in the data for the consecutive motion testing, but the time to communicate the task is not for the scripted motions. As such, this accounts for an unknown portion of the time difference between the two types of testing.

There are several factors that were not analyzed in this testing that may impact the results. These include the potential presence of a long-term learning curve spanning over the course of the testing program and fatigue over the course of an individual dive. While efforts were taken to mitigate the effects of variables like fatigue to the extent possible, these factors remain unaccounted for, and are thus potential sources of error to the results detailed above. In the future, a dedicated ground-based Earth gravity testing sequence would be beneficial to explore these impacts and determine the time required to negate the significance of any potential long-term learning curve, if it exists, which could then be utilized to reduce this bias in future testing (either experimentally or through data normalization).

VI. Conclusions

In this work, a new methodology is proposed and explored by which investigation into minimum habitat sizing can be conducted through experimental means. Utilizing modular rack structures to form habitats and the underwater environment to simulate differing gravity levels, the impact of habitat size on crew efficiency can be investigated without the need for predictive or heritage design regression methods. While the research that has been conducted to date is still preliminary, the data shows potential trends in habitat sizing favoring smaller modules for short-term, two crew vehicles to maximize efficiency. It also shows that, at least in the underwater environment, the impacts of inter-crew motion conflict are reduced and are more easily adapted to, thus necessitating future underwater testing (both in this research and in other underwater testing programs) to consider these impacts to the fidelity of the data. Ultimately, the question of the minimum required habitat size for various mission types still remains unanswered; however, the modular nature of this methodology lends itself to a variety of different forms of study through the ability to rapidly change habitat size and configuration. While the question is still open for the moment, further testing with this system can be utilized to generate a more comprehensive experimental basis for future sizing requirements and guidelines, thus enabling space habitats to be sized more appropriately, and ensuring a more optimal balance between mission cost and crew performance.

VII. Future Work

Of perhaps the highest benefit to future research utilizing this methodology would be the expansion of the subject pool and conducting a high number of repetitions with each group, as this would allow for a substantially greater degree of statistical significance to the data. While the COVID-19 pandemic limitations that impacted laboratory tests in the current study are (hopefully) transitory in nature, overcoming the underwater limitations described in this paper will require dedicated focused research funding incompatible with "student project"-oriented funding such as X-Hab. Additionally, completion of the testing matrix with martian gravity would be worthwhile, especially as manned missions to Mars increase in interest. The testing matrix could be expanded to include horizontal habitats and/or larger crew size testing as well. Horizontal configurations would increase the number of different volumes being tested, and would also likely result in significant variation in motion data due to the reduction of the impact of circumferential vs direct motion from the walls being closer together. The inclusion of larger crew size testing would increase the complexity of each subject having to deal with the motion of additional subjects instead of only one. This cognitive load and the increase in motion conflicts would potentially have significant impacts to the data, and would also be of utility due to the interest in four crew Artemis missions.

Additionally, higher fidelity testing would be a major way in which this research path could be built further. This testing, by design, is a low-fidelity habitat simulation over a short time span. The inclusion of high-fidelity tasks, including environmental control repair tasks, medical events (standard and emergency), scientific work such as operating a glovebox, cargo transfer bag manipulation, spacesuit maintenance, and airlock operations, would result in a much more realistic simulation of habitat operations. This concept was explored in more detail at the planning level in prior activities.[5] Similarly, longer term studies would both increase fidelity and enable analysis of the psychological impacts of spending long periods of time in restrictive habitats, which is something that this study does not capture. While several programs aim to analyze the psychological impacts that result from isolation through the use of habitat simulations, such as the NASA HI-SEAS missions,[10] these often do not analyze the impact of habitat size itself and do not capture differences due to differing gravity levels. It may be possible to conduct longer-term (several hour) testing underwater in a non-academic environment, but any multi-day testing would preclude the use of underwater testing. Thus, to conduct significantly longer-term testing, surface testing would be required, and would likely need to be modeled in structure and scope to the HI-SEAS and similar projects. These paths remain currently unexplored in the scope of minimum habitat sizing, but are of great interest to the continuation of this work.

Finally, in the early stages of testing, the racks were open frames apart from some units having shelves. Despite the racks describing a simulated boundary that test subjects never crossed, the habitat still had the illusion of feeling larger and more open than it was. When opaque panelling was added to most of the racks to bound the habitat, subjects reported the space feeling smaller despite the identical volume. This same effect could be assumed of a ceiling barrier that visually bounds the vertical space in the habitat. However, this creates significant safety concerns, since divers always need a way of exiting the water in case of emergency. This could be negated by creating a partial, radial ceiling that is made of flexible material and overhangs only a couple feet from the rack tops. This becomes more critical when considering future research into multi-level habitat operations, including transfer between levels as a major extension to studies of internal crew mobility.

Acknowledgements

The authors would like to thank Athenais Culleron-Sun, Robert Fink IV, Jonathon Gabriel, Adin Goldberg, Nicholas Limparis, Jonathan Molter, Sarah Pfau, Josaiah Rodriguez, Ethan Tang, Alison Thieme, and Howard Zheng for their contributions to this project. We would also like to thank the students in the 2021 sequence of ENAE 100 who worked on this project: Dominik Blaho, Chaitanya Garg, Kyle McMahon, William Scott, and Alejandro Tovar.

This research was conducted with the support of the 2020 and 2021 NASA Moon-to-Mars Exploration Systems and Habitats (M2M X-Hab) Academic Challenge programs.

References

[1] J. T. Celentano, D. Amorelli, and G. G. Freeman, "Establishing a Habitability Index for Space Stations and Planetary Bases," AIAA 63-139, AIAA/ASMA Manned Space Laboratory Conference, Los Angeles, CA, May 2, 1963.

[2] Marc M. Cohen, "Testing the Celentano Curve: An Empirical Survey of Prediction for Human Spacecraft Pressurized Volume," SAE 2008-01-2027, 38th International Conference on Environmental Systems, San Francisco, CA, June 29-July 2, 2008.

[3] David R. Liskowsky and William W. Seitz, "Human Integration Design Handbook (HIDH)" NASA/SP-2010-3407/REV1, June, 2014.

[4] M. Rudisill, R. Howard, B. Griffin, J. Green, L. Toups, and K. Kennedy, "Lunar Architecture Team – Phase 2 Habitat Volume Estimation: "Caution When Using Analogs"," Earth and Space Conference 2008, 1–11. https://doi.org/20110023287

[5] David Akin, Zachary Lachance, and Charles Hanner, "Experimental Investigation of Minimum Required Cabin Sizing in Varying Gravity Levels," 50th International Conference on Environmental Systems, July 2021.

[6] Adam D. Mirvis, "A Dynamics-Based Fidelity Assessment of Partial Gravity Gait Simulation Using Underwater Body Segment Ballasting," Thesis, 2011.

[7] John Mularski and David Akin, "Water Immersion Ballasted Partial Gravity for Lunar and Martian EVA Simulation," 2007-01-3145, 37th International Conference on Environmental Systems, Chicago, IL, July 2007.

[8] Eric Jones, "Lunar Gaits," Apollo Lunar Surface Journal, November 2010.

[9] Edwin Olson, "AprilTag: A Robust and Flexible Visual Fiducial System," Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), May 2011.

[10] Sandra Haeuplik-Meusburger, Kim Binsted, and Tristan Bassingthwaighte, "Habitability Studies and Full Scale Simulation Research: Preliminary Themes Following HISEAS Mission IV," 47th International Conference on Environmental Systems, July 2017.