Game Design for Stroke Rehabilitation Robots:
A Pilot Study Comparing
Masked Versus Unmasked Game Designs

by

Howard T. Chiam

A thesis submitted in conformity with the requirements
for the degree of Master of Health Science in Clinical Engineering
Institute of Biomaterials and Biomedical Engineering
University of Toronto
© Copyright by Howard T. Chiam 2017
Abstract

Appropriately-designed computer games can improve rehabilitation therapy outcomes by motivating stroke survivors to follow through with exercises. Supplementing conventional therapy with computer games and robotics enables patients to continue without therapists constantly present for repetitive parts of therapy. Studies have used “Basic” (framed as target exercise), “Unmasked” (simulating everyday activity), or “Masked” (fantasy activity) games, without addressing which design best maintains stroke survivors’ motivation and time on task. Three such games were developed for an upper limb robot and analyzed with gameplay measurements, questionnaires, and observations determining which game users preferred, played longer, and performed more repetitions. The “Masked” game had the greatest preference “votes” (5), time on task (24.2 ± 11.6 minutes), and motion repetitions (1191 ± 778), but also misses (180 ± 211) and standard deviations. We should individually tailor games, but continue developing “Masked” games, possibly blending game types. Future design and process recommendations were listed.
Acknowledgments

I would like to thank those who were part of my thesis committee: Alex Mihailidis, Rosalie Wang, Patricia Trbovich, Elaine Biddiss, and José Zariffa. I would like to thank Marge Coahran for all her technical support and patience in explaining and onboarding. I would like to thank Cindy Ho and Scott Tomlinson for their help in monitoring patients and support while I put together the protocol. I would like to thank Jessica Babineau for her help with doing a literature search into rehabilitation game design comparisons. I would like to thank my lab mates in IATSL (a.k.a. AIRR) at TRI for their support and feedback, especially to Michael Li for his helpful suggestions. I would like to thank Quanser Inc. for in-kind support, as well as Toronto Rehabilitation Institute, the AI and Rehabilitation Robotics Team, and the University of Toronto for the funding and resources that made this work possible. And finally I would like to thank the one who makes this all worthwhile. You know who you are.

Disclosure

The investigators involved in this research may be involved in the commercialization of therapy devices resulting from this research through Toronto Rehabilitation Institute and the industrial partner, Quanser Inc., which is involved in the design and manufacture of the haptic robots used in research projects such as this one. However, the industrial partner is not involved in the clinical evaluation of the systems, and has no access to any data collected in this research project.
# Table of Contents

Abstract ......................................................................................................................... ii
Acknowledgments .......................................................................................................... iii
Disclosure ....................................................................................................................... iii
Table of Contents .......................................................................................................... iv
List of Tables .................................................................................................................. vii
List of Figures ................................................................................................................. viii
List of Short Forms ...................................................................................................... x
1. Introduction ............................................................................................................... 1
   1.1. Background ........................................................................................................... 1
   1.2. Rationale .............................................................................................................. 1
   1.2.1. Why Retrain the Brain in Upper Limb Control Motor Recovery? ....................... 1
   1.2.2. Why are Movements that are Goal-Oriented and Repetitive Critical for Post-Stroke Rehabilitation? ................................................................................... 2
   1.2.3. Why Robots and Games for Therapy? ............................................................... 3
   1.3. Overall Goals ..................................................................................................... 4
   1.4. Thesis Roadmap .................................................................................................. 4
2. Goals – Elaboration ................................................................................................... 5
   2.1. Specific Research Questions .............................................................................. 5
3. Related Literature ..................................................................................................... 8
   3.1. Upper Limb Rehabilitation Robotics in Stroke Therapy ...................................... 8
   3.2. Gamification in Robotic Stroke Therapy ............................................................. 13
   3.3. Gamification in Robotic Stroke Therapy - “Masking” .......................................... 15
4. Research Method ...................................................................................................... 20
   4.1. Robot Platform .................................................................................................. 20
   4.2. Game Development ......................................................................................... 21
4.2.1. Pre-Testing for Masked-Unmasked Game Equivalencies ................................. 25
4.2.2. Mock Trials ................................................................................................. 28
4.3. Study Design................................................................................................. 28
4.4. Participants .................................................................................................... 29
4.5. Data Collected ............................................................................................... 29
4.6. Procedures....................................................................................................... 32
4.7. Analysis ........................................................................................................... 35
5. Results ............................................................................................................... 39
5.1. Demographics Data ....................................................................................... 39
5.2. Research Question #1: Time ....................................................................... 40
5.3. Research Question #2: Repetitions ............................................................. 41
5.4. Research Question #3: Preferences .............................................................. 47
5.4.1. Preferences - Non-IMI Question Scores ..................................................... 51
5.5. Additional Data - Statistical Analysis .......................................................... 53
5.6. Observation Logs – Verbal Feedback & Observations: ................................. 56
5.7. Questionnaire - Reasons for Preferences: .................................................... 58
5.7.1. Questionnaire - Other Open-Ended Questions ......................................... 58
6. Discussion of Results......................................................................................... 60
6.1. Masked or Unmasked? .................................................................................. 60
6.2. Recommendations and Lessons Learned for Future Research ................. 65
7. Limitations ......................................................................................................... 69
7.1. Sessions and Follow-Up ............................................................................... 69
7.2. Statistical Significance .................................................................................. 69
7.3. Equivalency Testing between “Masked” and “Unmasked” Games .............. 69
7.4. Limitation Related to Permutations of Order of Presentation of Games ....... 72
7.5. Limitations Related to Participant Sample .................................................... 73
List of Tables

Table 1 - Sample of the Variety of Stroke Rehabilitation Robots Tested with Games Tested..... 10
Table 2 - Game Types Used in Stroke Rehabilitation Robotics ........................................ 16
Table 3 - Game Features to Match across Games .......................................................... 23
Table 4 - Anchors for 7-Point Scale ............................................................................. 35
Table 5 - Participant Demographics ............................................................................. 40
Table 6 - Participant Game Preferences and Intrinsic Motivation Inventory (IMI) Scores ..... 47
Table 7 - Questionnaire, 7-Point Scale Questions Only ................................................ 48
Table 8 - Comparing IMI Subscale Scores - by Participant .......................................... 50
Table 9 - Comparing Explicit Game Preference with Other Measures - by Participant ...... 50
Table 10 - T-Tests for Significant Difference Between Games .................................... 53
Table 11 - Tests for Effect Size of Differences Between Games .................................. 54
Table 12 - ANOVA Tests for Significant Difference Between Games ....................... 55
Table 13 - Sources of the Questions .............................................................................. 95
Table 14 - Example assignment of permutations to participants ................................ 98
Table 15 - CMSA Stages Notes ................................................................................... 100
Table 16 - Games with Highest Values per Subject ....................................................... 101
List of Figures

Figure 1 - The "Basic" Game Used in This Study ................................................................. 6
Figure 2 - "Masked" Game Used in This Study ................................................................. 6
Figure 3 - "Unmasked" Game Used in This Study ............................................................ 7
Figure 4 - Haptic Stroke Rehabilitation Robot System ....................................................... 20
Figure 5 - Time on Task (minutes) .................................................................................. 41
Figure 6 - Number of Motion Repetitions ....................................................................... 43
Figure 7 - Number of Hits ................................................................................................. 43
Figure 8 - Repetitions per Minute .................................................................................... 44
Figure 9 - Number of Misses ........................................................................................... 45
Figure 10 - Hits (percent of targets presented) ................................................................. 45
Figure 11 - Hits per Minute .............................................................................................. 46
Figure 12 - Repetition Sizes (percent of axis out from centre of ROM) ......................... 46
Figure 13 - IMI Sub-Scale Scores .................................................................................... 49
Figure 14 - Question 1 from the questionnaire ............................................................... 51
Figure 15 - Question 4 from the questionnaire ............................................................... 51
Figure 16 - Question 8 from the questionnaire ............................................................... 52
Figure 17 - Question 10 from the questionnaire ............................................................. 52
Figure 18 - Question 12 from the questionnaire ............................................................. 52
Figure 19 - Example Schematic of ROM Use Percentages in Four Directions .............. 91
Figure 20 - Time on Task, Within Subjects

Figure 21 - Hits (#), Within Subjects
List of Short Forms

TRI = Toronto Rehabilitation Institute

IATSL = Intelligent Assistive Technology and Systems Lab

REB = Research Ethics Board

ADL = Activity of Daily Living

ADLs = Activities of Daily Living

IMI = Intrinsic Motivation Inventory

ROM = Range Of Motion

CMSA = Chedoke-McMaster Stroke Assessment

RCT = Randomized Control Trial
1. Introduction

1.1. Background

Appropriately-designed computer games can engage stroke survivors and motivate them to follow through with therapy exercises, hence leading to improved outcomes for the upper limb (Laver et al., 2011). Supplementing conventional upper limb therapy with computer games and robotics enables patients to continue exercises without requiring therapists to be constantly present for repetitive parts of therapy.

1.2. Rationale

1.2.1. Why Retrain the Brain in Upper Limb Control Motor Recovery?

After a stroke, it is essential that the upper limb is retrained for motor recovery. Stroke is the leading cause of disabilities in motor function, and can significantly interfere with performing Activities of Daily Living (ADLs), such eating, walking, or dressing (Johnson, 2006). Most stroke survivors experience a loss of control of the upper limb and a negative impact on quality of life (Lum et al., 2009). Two types of events can cause a stroke: when blood is obstructed from delivering oxygen to a part of the brain (causing ischemic stroke), or when there is a broken vessel bleeding in the brain that damages brain cells (causing hemorrhagic stroke), with ischemic stroke being more common of the two (NHLBI, 2016). Both of these events can cause brain cells to be damaged or to die, in turn negatively affecting the parts of the body that are controlled by the corresponding brain cells (NHLBI, 2016). Impairments that result sometimes include partial paralysis on one side of the body, and can significantly impair functional performance in ADLs (Johnson, 2006). The upper limb is a top priority in recovery, because an estimated 65% of stroke survivors cannot perform daily activities due to loss of control of their upper limb (Lum et al., 2009). It is thus essential to retrain the parts of the brain that control the upper limb. This can be done through repeated motions, and these motions should be both goal-oriented and repeated many times over.
1.2.2. Why are Movements that are Goal-Oriented and Repetitive Critical for Post-Stroke Rehabilitation?

Disabilities or impairments caused by stroke can be reduced with goal-oriented and repetitive movements which can help to restructure neural pathways and improve motor control (Broetz et al., 2010; Lum et al., 2002). Based on studies using animal models, repetitions in the order of hundreds per day, or at the very least 100 repetitions per day, are required for effective motor recovery (Kimberley et al., 2010), with some studies further citing 400 to 600 skilled movement repetitions per day (Cameirão et al., 2016). Performing such movements are critical for post-stroke rehabilitation therapy. However, research shows that patients also need to feel that such “practice” is aimed towards a task or goal that is meaningful to them, in order for there to be significant cortical reorganization in the brain to remap motor control, and hence recovery of motor control (Bayona et al., 2015).

Traditional therapy involves physiotherapists and occupational therapists working directly with stroke survivors one-on-one (Perry et al., 2010), with intense care over the long term, but this cannot always be done. For example, a stroke survivor might practice moving an affected limb while a therapist provides assessment, assistance, (NIH, 2015) or resistance to the arm (Valero-Cuevas et al., 2016). Therapists perform a large number of services that can be grouped under evaluation, goal-setting, intervention, and re-evaluation (Gillen, 2015), and cover activities such as remediating impairments, teaching compensatory techniques, providing preventative intervention, training to do ADLs, tailoring programs, and performing assessments throughout the process (Richards et al., 2005). Therapists also help stroke survivors perform progressively more complex tasks while supervising, assessing ROM and endurance, and even helping make changes in their living environment for ADLs (NIH, 2015). A majority of stroke rehabilitation therapists seem to frequently use techniques from multiple approaches (Schriner et al., 2014). These approaches are conceptually termed as different neurological frames of reference: Brunstrom, Constraint-induced Movement Therapy (CIMT), Neurodevelopmental Treatment (also known as NDT or the Bobath method), Proprioceptive Neuromuscular Facilitation (also known as PNF), Rood, and Task-Oriented (Schriner et al., 2014). For example, in CIMT, the unaffected limb is restrained while the affected limb is encouraged to be used intensively to counteract “learned non-use” (Grotta et al., 2004), hence involving investment of time and effort on
the part of both stroke survivor and therapist. CIMT is effective at improving function and causing cortical changes (Peurala et al., 2012; Miltner et al., 2016). However, as shown above, rehabilitation therapy demands a lot of interaction between therapists and stroke survivors on repetitive tasks, and some therapists have between 7 to 16 stroke patients to attend to (Lowry, 2010), which can be challenging for therapists to offer the needed repetition over long periods of time.

In brief, there is a high demand for, yet low supply of available therapist time, while the time that is spent in therapy may be overwhelmed by time spent on tasks such as repeating movements, taking up time that could be more effectively allocated to other parts of therapy that require more involvement of a therapist’s presence, such as consultation for goal-setting and personalized guidance that can vary between patients and even day-to-day for a single patient. A modern form of stroke rehabilitation therapy may address these issues: robotic stroke therapy.

1.2.3. Why Robots and Games for Therapy?

Robots can provide haptic assistance for stroke survivors to perform the much-needed repetitions (Lum et al., 2002), provide reproducible motions and objective measurements of performance (Volpe, 2014), give feedback to stroke survivors (Alves & Samson, 2014; Putrino, 2014), and provide more time for therapists to focus on the non-repetitive aspects of their work with stroke survivors (Bergeron & Lam, 2014). Repetitive exercises can be “gamified”, i.e. re-framed as games, to make them more enjoyable and increase player motivation (McGonigal, 2011), sometimes without stroke survivors even being aware of how repetitive their therapy exercises are (Alves & Samson, 2014). So far the literature suggests that more research is needed to investigate gamified robotic stroke therapy. Only about a third of stroke survivors seem to adhere to exercises recommended to them by therapists (Shaughnessy et al., 2006). This may be due to attitudes (such as self-efficacy and outcome expectations) towards the design of the exercise activities, but other design aspects need to be explored to fully explain and increase adherence to exercise (Shaughnessy et al., 2006). The current study investigates one design aspect of the system by looking into different ways to reframe exercises as a game (namely via “Masking”), in order to increase motivation, time on task, and motion repetitions, and to set the stage for future research to investigate recovery and transfer to daily living. To the author's knowledge, there are no previous studies to date that examine the idea of masking and directly compare the three game designs used in this study, only studies comparing games of one or two of such design types (for example studies, see the state of the art survey by Ma et al., 2014).
Testing all three games design types directly can reveal overall game preference for stroke survivors themselves. However, also making objective measurements can help bring to light any discrepancies between user preference and performance, and enable further insights into improving both motivation and recovery outcomes in general.

1.3. Overall Goals

The main goal is to determine which game design (“Basic”, “Masked”, or “Unmasked”) is better at getting stroke survivors to be willing to play, to play longer, and to perform repetitions (of greater number and intensity) during the time that they play. The willingness to play, longer play times, and greater number and intensity of repetitions are promising indicators of the amount of exercise they receive. The secondary goal is to find beneficial game design aspects or features from each game for future game development. Future work will apply the findings to develop more games and compare different games’ outcomes in motor recovery and transfer of gains to daily activities.

1.4. Thesis Roadmap

Chapter 2 elaborates on the goals of this study and the specific research questions to be answered, as well as provides definitions of the three games developed and tested in this study. Chapter 3 describes related literature covering different stroke rehabilitation robots, gamification, and “masking” in stroke rehabilitation games, and provides background for the three groupings of measures used to answer this study’s questions. Chapter 4 describes the study design, the research method for game development, the selection of participants to test the games, and defines the data and measurements collected and how they are analyzed to answer the three specific research questions. Chapter 5 provides the results, including demographic data, tables and plots for each measurement grouped by their respective research questions, and participant feedback (quotes) and observations to support the quantitative data. Chapter 6 discusses the results, answers the research questions, and provides recommendations based on findings from this study. Chapter 7 discusses the limitations on the interpretations of the results and recommendations, and discusses the relation of these results and recommendations to other literature. Chapter 8 discusses the relevance of this study to game design for robotic stroke therapy and its clinical implications. Chapter 9 recapitulates with a conclusion.
2. Goals – Elaboration

2.1. Specific Research Questions

For the main goal of comparing the three game designs (“Basic”, “Masked”, and “Unmasked”), the research questions are as follows:

1) Which game design has users wanting to keep playing for the longest period of time?

2) Which game design contributes to the greatest number, rate, and size (i.e. Range of Motion used in workspace) of motion repetitions?

3) Which game design has the most favourable stroke survivor self-reported preference scores (perceived usefulness, enjoyment, explicit indication of game preference)?

For each of the three games used in this study, various beneficial aspects of those games are identified based on user preferences, and those design aspects will be recommended to be incorporated into future game designs. The hope is to combine those beneficial aspects that contribute to stroke survivors wanting to keep playing those games, and to receive more exercise through playing longer and doing more repetitions.

The game design aspects that this study looks at include general game features captured from both participant feedback and researcher observations. Examples of general game features include gameplay mechanics and different kinds of graphical interface elements. Features were pointed out by users in their open-ended feedback during gameplay and while answering a custom questionnaire. Any elements perceived by the researcher to have an effect user experience were also noted. However, the main game design aspect that this study looks at is the effect of “masking”. This study compares three games: a “Basic” game, a “Masked” game, and an “Unmasked” game.

The “Basic” game refers to a game that does not simulate an environment or ADLs, and has explicit exercise goals. For example, there are only set points (targets) on a screen, with no game or “story” overlaid (Figure 1).
The “Basic” game refers to a game with non-realistic environments, and involves activities that are not obviously directly related to ADLs (Figure 2).

The “Unmasked” game refers to a game with realistic environments and involves activities that are obviously directly related to ADLs (Figure 3).
Figure 3 - "Unmasked" Game Used in This Study. The user moves the cup to fill the soup with ingredients. Note that there is a timer in higher game levels.

For the “Masked” and “Unmasked” games, note that the difference is not simply different visuals overtop the same motions made by the user, but having the user feel that the tasks they are trying to do with those motions are different. The “Unmasked” game intends to present or frame the exercise as an ADL, while the “Masked” game frames the exercise as an activity that is not an ADL. Both the “Masked” and “Unmasked” games try to hide the fact that the user is doing exercise, while the “Basic” game only displays targets on the screen interface for the user to see.
3. Related Literature

3.1. Upper Limb Rehabilitation Robotics in Stroke Therapy

This study uses a rehabilitation robot designed for upper limb stroke therapy as a platform on which to design and test computer games for stroke rehabilitation. In the Encyclopedia Britannica online, rehabilitation robotics is defined as “any automatically operated machine that is designed to improve movement in persons with impaired physical functioning” (Reinkensmeyer, 2016).

Robotic stroke therapy uses robotic devices to assist in delivering therapy, and offers potential benefits in rehabilitation therapy for both stroke survivors and therapists. A Cochrane Review updated in 2015 found that there is some limited evidence suggesting that robotic stroke therapy can improve stroke survivors’ arm and hand function, their arm and hand muscle strength, and their ADL performance (Mehrholz et al., 2015). In a more recent review (Veerbeek et al., 2016), 44 RCTs in English, German, or Dutch were found, of which 38 trials (N = 1206) had promising but limited results, especially in terms of ADLs and generalization, with limitations possibly due to inadequate trial designs that did not account for the timing, contrast, and recovery potential from baseline. Electronic devices can collect detailed information about a stroke survivor’s performance in therapy, and display the data back to them as another way to motivate them to improve (Alves & Samson, 2014; Putrino, 2014). Using robots in rehabilitation therapy can allow therapists to focus on non-repetitive aspects of their work (Bergeron & Lam, 2014) such as individualized goal-setting and counselling, reduce the high demand on therapists, and serve as a complement to traditional therapy. This is because robots can provide stroke survivors with reproducible precise motion and force control for longer periods of time, give objective numerical feedback and history of performance even when a therapist is not available, and can optimize exercise settings for individual stroke survivors (Volpe, 2014). These kinds of assistance can complement the other services that traditional methods provide, such as the various kinds of feedback, goal-setting, and personalized encouragement that therapists provide to individual stroke survivors.

It might be argued that some non-robotic systems can be easier to access, set up, and use, however robots designed for rehabilitation therapy are able to provide “direct” physical control
(e.g. of joints), haptic feedback, assistance, and resistance, which are not available in systems controlled by computer-vision (such as the Kinect) or even by “haptic remote” (such as the Wii). The Kinect may be simpler to set up and use than a robotic system because it is controlled hands-free, but is limited to large, non-occluding movements, and is prone to “cheating” via trunk compensation and has only been used in short-term studies with small sample sizes (Webster and Celik, 2014). On the other hand, Soli radar technology can overcome the problem of compensation, since it has sub-millimeter accuracy detection of hand gestures even with occlusion (Yeo et al., 2016), and can even detect different materials to expand the possible game applications (Lien et al., 2016). However, by the very fact that these technologies are not haptic-controlled, they are not able to provide haptic feedback, assistance, and resistance that therapists provide.

According to a review by Veerbeek et al. (2016) there is no globally-accepted classification for rehabilitation robots, but one way to classify rehabilitation robots is through a dichotomy of end-effector robots and exoskeleton devices. End-effector robots have one attachment point controlled by the outermost part of a user’s limb (usually their hand), while exoskeleton devices directly control joints that correspond to anatomical joints (Veerbeek et al., 2016). The review indicates that we need more research into the effects of both end-effector robots and exoskeleton devices (Veerbeek et al., 2016).

Besides the categories of end-effector robots and exoskeleton robots, there are also other facets of (sub-)categorizing upper limb stroke rehabilitation robots such as robots targeting the arms, multiple joints, or fine motor control in the hands; passive control, or active control; or even a combination of a robot with virtual reality and simultaneous electroencephalography (EEG) recordings. Examples of such robots are described in a state of the art survey of video games for upper limb stroke rehabilitation (Ma et al., 2014): the MIT-MANUS, the ARMin, the UL-EX07, the T-WREX, the Trackhold, as well as a combination of the Cyberglove, CyberGrasp and Haptic Master. These robotic systems use anywhere from 2 to 7 degrees of freedom and most of them provide movement assistance and gravity compensation (Ma et al., 2014), and have generally promising results. What follows is a description of this range of robots used in gaming studies and their pros and cons, the robot used in this study, as well as a final synthesis of the literature on upper limb stroke rehabilitation robotics. See Table 1 below for a summary of the variety of robots found in a state of the art survey (Ma et al., 2014):
### Table 1 - Sample of the Variety of Stroke Rehabilitation Robots Tested with Games Tested

<table>
<thead>
<tr>
<th>Name</th>
<th>Studies</th>
<th>Example of…</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIT-MANUS</td>
<td>(Huq et al., 2013; Veerbeek et al., 2016; Krebs et al., 1998)</td>
<td>Early example of table-top end-effector robot</td>
</tr>
<tr>
<td>ARMin</td>
<td>(Staubli et al., 2009)</td>
<td>Exoskeleton robot targeting shoulder, elbow, and wrist</td>
</tr>
<tr>
<td>UL-EX07</td>
<td>(Kim et al., 2013)</td>
<td>“Bilateral” robot that can be used on one or both arms for maximum ROM workspace</td>
</tr>
<tr>
<td>T-WREX</td>
<td>(Housman et al., 2009)</td>
<td>Passive robot</td>
</tr>
<tr>
<td>Trackhold</td>
<td>(Steinisch et al., 2012)</td>
<td>Passive robot with virtual reality and simultaneous EEG</td>
</tr>
<tr>
<td>Cyberglove with CyberGrasp and Haptic Master</td>
<td>(Merians et al., 2011)</td>
<td>System for both arm and hand, to target fine motor control</td>
</tr>
</tbody>
</table>

There have been different robots of various levels of sophistication and cost on which stroke rehabilitation games were tested, with robots designed to consider the amount of assistance and targeting of joints in the upper limb. A table-top end-effector robot such as the MIT-MANUS (Huq et al., 2013; Veerbeek et al., 2016) can be lower cost than an exoskeleton robot, but can have less flexibility in the kinds of movements and activities it can provide for exercises. On the other hand, an exoskeleton robot such as the ARMin (Staubli et al., 2009) can target the movements of more anatomical joints than an end-effector robot, thus reducing compensatory movements ("cheating") during gameplay, but can be more costly and difficult to set up than an end-effector robot. Passive robots can be both cheaper and safer because of their lack of active motors, such as the T-WREX (Housman et al., 2009). However, due to the very fact that passive robots does not use active motors, they cannot provide movement assistance using forces, which may especially be a drawback for use with stroke survivors with severe movement impairments. Even though Housman et al. (2009) cites lack of evidence for the effectiveness of motion assistance and even negative effects of using assistance, this does not consider the benefits of resistance, let alone a responsive artificially intelligent system using a combination of assistance,
resistance, and absence of force. Targeting the hand of the upper limb is helpful as hand impairments have been found to be a main limiting factor for functional arm use (Housman et al., 2009; Lum et al., 2009), but one potential drawback is the question of cost and complexity of setting up and maintaining the three different systems in the case of the combination of the Cyberglove, CyberGrasp, and Haptic Master (Merians et al., 2011).

There have also been alternative means of enriching interaction with stroke rehabilitation robots, and hence with games. One way was to involve use of both arms. A “bilateral” robot like the UL-EX07 (Kim et al., 2013) can be controlled by one or both arms, being able to target the full ROM of the arm(s) because it can support almost all of the upper limb work space, and hence realistically simulate movements of a wider range of ADLs. In “bilateral” use, one arm can be used to support the motion of the other arm, however there is not enough evidence to show that “bilateral” use provides better recovery than using the robot for only one arm (Kim et al., 2013). Using a “bilateral” robot opens up the possibility of activities that involve both arms, or enables users to safely support the other arm during gameplay, but can be costly or require more effort to set up than other systems. Another way to enrich interaction with games was using virtual reality, such as was done with the Trackhold robot, which was also tested with simultaneous EEG recordings (Steinisch et al., 2012). Using virtual reality can add a more immersive experience, and using simultaneous EEG could add better measurements of responses during gameplay, but combining one or both adds more to the set up of a system than simply setting up only a robot.

The robot used in this study is an example of an end-effector robot that can be used with one of either arm. One difference between this robot and the MIT-MANUS is its design for portability. It was designed to address upper limb stroke rehabilitation therapy, as it was specifically created in consultation and tested to address stroke survivors’ needs. It can provide motion resistance, assistance, and haptic feedback, with directionally (vs. the “on/off” of vibrating controllers). It can also directly monitor and measure (instead of “interpolating”) movement and forces from the arm. It was also explicitly designed to be lower cost than typical rehabilitation robots on the market, and has AI capabilities in its software to adjust settings automatically in response to measured gameplay characteristics such as user fatigue. Being an end-effector robot, it is simpler to construct and hence less costly than more complicated devices such as exoskeletons, but requires motors and can be more costly than passive robots, for the trade-off of offering haptic
feedback, assistance, and resistance that therapists provide. Its simple planar design does not require gravity compensation in its table-top mode and the arm moves with two degrees of freedom. This robot serves as the platform for the games tested in this study.

Finally, although results from various studies are promising, current rehabilitation robots still lack certain features needed in stroke rehabilitation therapy, and more research into stroke rehabilitation robots is needed. Many of the robots mentioned are reported to be able to provide motion assistance, but almost all seem to have no mention of being able to provide resistance, even though resistance training in activities are one of the things therapists may provide during rehabilitation sessions (Valero-Cuevas et al., 2016). Robot controllers exist with many degrees of freedom, but at greater cost. Others may have relatively lower costs, but have limited ability to increase the range of upper limb motion for the stroke survivor, such as the stationary joystick design in Reinkensmeyer et al. (2002) with a planar workspace of 10 cm x 10 cm. Although there are promising results for stroke rehabilitation robotics, more research is needed into both end-effector and exoskeleton classes of upper limb stroke rehabilitation robots, and into whether the added benefit of exoskeletons (and the “bilateral” subtype) outweighs their cost and complexity of setup (Veerbeek et al., 2016). Making a robot more affordable may remove part of the roadblock to the actual use of a robot, and is part of the rationale for passive robots that are low-cost because of their lack of actuators (Housman et al., 2009). However, passive robots cannot replicate the assistance and resistance that therapists provide. Ongoing research on the robot used in this study will aim to at least help to confirm results in existing literature. Ongoing research at Toronto Rehabilitation Institute is also underway to design a hand module for the robot in this study, which is important as hand impairments have been found to be a main limiting factor for functional arm use (Housman et al., 2009). In general, more research is needed for optimal upper limb stroke rehabilitation robots.

However, despite the advantages of robots in providing long-term numerical data and freeing up therapists’ time from repetitive parts of therapy, there is also a need to motivate stroke survivors to participate in activities so that they continue using these robots in the first place.
3.2. Gamification in Robotic Stroke Therapy

Gamification can be useful for rehabilitation. Gamification is the use of games to increase engagement for an activity, and hence increase motivation to perform a task, even if it is repetitive (Alves & Samson, 2014). For instance, there are many exercise games made for the general public, such as Wii Fit, Kinect Sports, and Dance Dance Revolution. These games apply the concept of gamification to exercise tasks by framing them as a game, making the tasks more enjoyable and motivating for players to perform (McGonigal, 2011). Playing games can help motivate stroke survivors to follow through with repetitive therapy exercises, sometimes without them being aware of how repetitive the exercises are (Alves & Samson, 2014). For stroke survivors in the later stages of recovery (i.e. for which stroke occurred more than one year in the past), gamified rehabilitation can improve motor recovery, as shown in fMRI scans of stroke survivor brain behaviour before and after playing Pong with a one-dimensional mouse (Azpiroz et al., 2005). In that study, play was shown to increase spatial activation and activate previously inactive areas in the pre-motor cortex, as well as produce statistically significant improvement, even though participants were originally considered to be at their maximum recovery level (Azpiroz et al., 2005).

For stroke rehabilitation to effectively make use of gamification, we need to bridge the disconnect between popular commercial games and games for stroke rehabilitation. Many commercial “off-the-shelf” games were not designed for players with severe physical impairments caused by stroke, with impairments that can influence or interfere with gameplay, such as a limited range of motion (ROM) experienced by the majority of stroke survivors (Alankus et al., 2010). In addition, various design features, such as complex coloured backgrounds found in some games can sometimes be problematic for stroke survivors who often have visual or cognitive impairments (Bonnechère et al., 2013). While popular games have lacked essential components for effective rehabilitation, rehabilitation games have lacked the entertainment qualities of popular games (Flores et al., 2008). It would be beneficial to offer the same kind of motivation to exercise for stroke survivors as well, for therapy and entertainment. This study seeks to use games to increase motivation in stroke survivors to continue doing therapeutically-relevant motion exercises.

For robotic stroke rehabilitation, gamification can be combined with robotic therapy to provide both automated prompts and objective measurements during gameplay. When a stroke
survivor plays games on a robotic system, the game part of the system can automate requests that a therapist would normally ask of them (i.e. ask them to repeat certain motions, instead of requiring a therapist to prompt repeatedly), while the robot continuously takes objective measurements of movement characteristics, force, or range of motion.

However, more research is needed into gamified robotic stroke rehabilitation. Gamified therapy is a relatively new concept (Rahmani & Boren, 2012). There is insufficient research on which kinds of games are most effective and how appropriate gamified therapy is for stroke survivors, even though meta-analysis shows generally positive results that gamified therapy has an effect on improved outcomes for the upper limb (Laver et al., 2011). Furthermore, some studies seem to only touch on the idea of turning activities into games in stroke therapy and focus on the design of the robot in the system (such as Staubli et al., 2009; Reinkensmeyer et al., 2000), and not on investigating which specific components of the design of the games help to best engage and motivate stroke survivors. Examples of such components to consider would be specific types of reframed activities, i.e. game environments (Rego et al., 2010), as well as feedback types, time constraints, levels, clear goals, challenge, and ensuring adherence and piquing curiosity (Deterding et al., 2011). This gap in literature is mentioned as a topic of required future work by some studies (Alankus et al., 2010). Additionally, few studies have put significant focus into designing games with high usability for both stroke survivors and therapists (Putrino, 2014) and into investigating game scenarios, interface, and overall effectiveness (Flores et al., 2008).

There has been research done into gamified robotic stroke rehabilitation at TRI, and findings from previous work that this study took into consideration. The IATSL team at TRI conducted previous studies on gaming environments for robotic stroke rehabilitation, and findings indicated that stroke survivors who tested several game ideas with the robotic system perceived various factors that might improve their motivation and engagement in robotic rehabilitation (Alves & Samson, 2014). Some suggested ways to improve the games included: building in game characteristics that involve both cognitive and motor elements, offering visual and auditory feedback as incentives, providing choices in games to suit personal functional goals, offering short-term and long-term performance targets within games, and avoiding causing a player to over-exert themselves beyond their arm’s range of motion (Alves & Samson, 2014).
There seems to be a lack of literature on the categorization of games played using robots for stroke rehabilitation, but it is clear that more research needs to be done to examine different game characteristics for effective games, including game environment. A survey paper focusing on serious games in rehabilitation found that there is no single definition for serious games (let alone for robotic stroke rehabilitation games), but different authors agreed that serious games are computer games with pure entertainment not being the main purpose (Rego et al., 2010). Since there is no single way to define serious games, one would suppose that there is no single way to categorize or classify serious games. This in turn can make it difficult to compare games, especially since different rehabilitation systems can be very different, so (Rego et al., 2010) proposed a classification that is based on a set of criteria for designing more effective rehabilitation games and that is framed in terms of their fundamental characteristics. These characteristics include the application area, interaction technology, game interface, number of players, game genre, adaptability, performance feedback, progress monitoring, and game portability. Within “game genre” (i.e. game environment), there were games found to be based on movement evaluation, simulations, strategy, or combinations of game types (Rego et al., 2010). However, the games within that paper were either not played using robots, or were only compared to other games of similar “game genre”. Further research is needed to fully explore all entertainment characteristics, including “game genre”, for rehabilitation games to attain higher levels of motivation in users (Rego et al., 2010).

As identified above, more research needs to be done to examine game environments in order to improve user motivation. The next section describes how there is a lack of data on certain comparisons between different game environments and hence between different ways for gamification to reframe the exercises presented to stroke survivors.

3.3. Gamification in Robotic Stroke Therapy - “Masking”

No other study has specifically investigated the game classification used in this study, nor compared all the three game designs together (“Unmasked”, “Masked”, and “Basic”). Previous studies have used games that simulate real-life activities (“Unmasked”), games reframed as non-realistic activities (“Masked”), or computer-screen activities that have no game or story overlay at all but are explicitly presented as exercises, such as a set of targets on a plain background
The “Basic” game is how robotic stroke rehabilitation games had initially looked like in the past, including one of the initial games designed for the robot used in this study.

There are many examples of previous studies that tested their respective robots while using games that fall under either “Basic”, “Masked”, or “Unmasked” (or two of the three, but not all three), such as the many games tested on the six stroke rehabilitation robotic systems mentioned earlier (the MIT-MANUS, the ARMin, the UL-EX07, the T-WREX, the Trackhold with EEG, and the combination of the Cyberglove, CyberGrasp, and Haptic Master), and hence were a factor in their respective robotic systems’ research results. See Table 2 below for the game types that each study examined while testing their respective robots. For further descriptions of the games used in these studies, see Appendix H - “Game Design Types in Previous Studies”. None of these previous studies have simultaneously tested all three game design types (based on the definitions used in this study), and usually did not directly compare game types as an explicit research goal.

**Table 2 - Game Types Used in Stroke Rehabilitation Robotics**

<table>
<thead>
<tr>
<th>Robots</th>
<th>References</th>
<th>Game Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIT-MANUS</td>
<td>(Reinkensmeyer et al., 2000; Ma et al., 2014)</td>
<td>Masked (several).</td>
</tr>
<tr>
<td>ARMin</td>
<td>(Staubli et al., 2009)</td>
<td>2 Masked.</td>
</tr>
<tr>
<td>UL-EX07</td>
<td>(Kim et al., 2013)</td>
<td>4 Basic and 4 Unmasked.</td>
</tr>
<tr>
<td>T-WREX</td>
<td>(Housman et al., 2009)</td>
<td>3 Unmasked.</td>
</tr>
<tr>
<td>Trackhold &amp; EEG</td>
<td>(Steinisch et al., 2012)</td>
<td>5 Masked or Unmasked.</td>
</tr>
<tr>
<td>Cyberglove, CyberGrasp,</td>
<td>(Merians et al., 2011)</td>
<td>2 Masked and 2 Unmasked.</td>
</tr>
<tr>
<td>Haptic Master</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It is important to consider the three game designs to see whether the way that these games frame the exercise task impacts usability, or what can be learned from each of these designs, in order to help motivate stroke survivors to perform the necessary motion repetitions for motor recovery. Among studies that compared stroke robotic rehabilitation games, there seems to only
have been studies that compared games that fall under one or two of such design types (examples: Laver et al., 2012; Shin et al., 2014) such as the stroke robot game studies described in the state of the art survey mentioned earlier (Ma et al., 2014). Research seems to still be in the preliminary stage and hence only the following comparisons have been found in a literature search:

- comparing one custom game design with traditional therapy (such as Shin et al., 2014),

- comparing custom games with each other, but all of only one game design type, with games being one of either “Basic”, “Masked”, or “Unmasked” (such as in Reinkensmeyer et al., 2000; Staubli et al., 2009; Housman et al., 2009), and sometimes

- comparing custom games of only two of the game design types, with one from “Basic”, “Masked”, or “Unmasked”, and the other of a different type (such as in such as in Kim et al., 2013; Steinisch et al., 2012; Merians et al., 2011).

The studies mentioned earlier seem to conclude with favouring one game design over the other but without having directly compared all three designs used in this study to gain preliminary insight, which this study attempts to do. Previous studies’ publications also sometimes do not provide any rationale for their game choice(s), but those that do cite evidence for simulating ADLs increasing the likelihood of relevance and meaning to patients as well as evidence for a task-specific approach and use of context (such as Laver et al., 2012). However, studies have not specifically compared games the way this study does to address which design helps better maintain stroke survivors’ motivation and time on task.

There are perceived benefits in both "Unmasked" and "Masked" games. The “Unmasked” and “Masked” games basically test the idea of the need for near transfer of abilities to ADLs versus entertainment value and distraction from mobility limitations post-stroke. Comparing the “Unmasked” and “Masked” games with the “Basic” game tests whether adding any game or story overlay at all is superfluous when it comes to providing therapy to stroke survivors (the “Basic” game can be thought of as a control condition, but it is treated equally as one of the three conditions for the purposes of this study). Anecdotally, from initial talks with therapists, there is no apparent consensus among therapists as to whether an “Unmasked” or a “Masked” game design would be the best for the majority of stroke survivors. On the one hand, an “Unmasked” game could provide practice with motions in the context of activities that are similar to ADLs,
and hence make it easier to apply in real life, i.e. providing near transfer to target tasks or easy generalization of skills (Toglia, 1991). There is evidence that the therapy needs to be task-specific and meaningful to the individual stroke survivors (Bayona et al., 2015), which would support use of the “Unmasked” game since it simulates an ADL, especially if individual stroke survivors find doing such as meaningful to them. Intuitively, it is just like how spending time practicing a motion for a sport helps improve motor skills, not simply the repetitions of motions (Bayona et al., 2015). On the other hand, a “Masked” game could provide the entertainment value of novel scenarios as a kind of “escapism”. One may also consider from personal experience the effect of the suspension of disbelief in films that are clearly not of realistic scenarios, but that naturally engage one's attention for the entire duration of the film—often lasting over an hour. A “Masked” game could also help with self-efficacy by distracting a stroke survivor (Taylor & Griffin, 2014), instead of discouraging them by inadvertently reminding them of activities they cannot easily perform, reminiscent of distraction therapy used to reduce anxiety in cognitive rehabilitation (Stone et al., 2014). This may help stroke survivors be more willing to participate in activities in the first place if they are framed more as fantasy games and not so much as practice, especially by reducing the boredom of training repetitive motions (Taylor & Griffin, 2014).

In cognitive rehabilitation studies, it seems that helping patients complete sessions and adhere to long term therapy is as beneficial as the therapy itself (Bruckheimer et al., 2012), and that games providing pure enjoyment for its own sake can help motivate patients to participate for a long time (Kizony et al., 2006). There is preliminary cognitive rehabilitation evidence that “Masked” games can help patients unconsciously perform desirable movements, have raised levels of attention and motivation, and thus may forget about time and hence complete therapy sessions and be willing to adhere to future sessions (Bruckheimer et al., 2012). Given a mix of undocumented anecdotal opinions on this matter from people other than the users themselves (i.e. only therapists, not stroke survivors) and lacking data on this for the specific group of users with their therapy needs, this study explores this aspect of game design environment. As already discussed, there are examples of both “Unmasked” and “Masked” games in literature with promising results, but have not been directly tested in terms of “masking” and compared to a “Basic” game. As an additional note of clarification, while other studies have used either “Masked” or “Unmasked” games (similarly defined around ADLs and non-ADLs) for cognitive rehabilitation (Vourvopoulos et al., 2014; Shapi’i et al., 2015; Stone et al., 2014), this study focuses on motor rehabilitation. However, even in cognitive rehabilitation, there are examples of both “Masked” games and
“Unmasked” games finding use or at least being developed, such as a realistic navigation of tasks in a simulated city (Vourvopoulos et al., 2014), quizzes and mazes (Shapi’i et al., 2015), and “SnowWorld”, which had the intentional goal of re-directing attention (Stone et al., 2014).
4. Research Method

This study used an exploratory mixed-methods design (Johnson et al., 2007). Consulting stroke survivors and therapists generated game parameters and ideas. Three games requiring the same motions, but different designs were developed for an upper limb rehabilitation robot (Lu et al., 2011). Stroke survivors tested the games and game design aspects were analysed using robot data, questionnaires, and observations to determine contributions to motivation, time on task, and repetitions. Discrepancies between subject preference measures and objective performance measures were also observed and analyzed.

4.1. Robot Platform

IATSL in collaboration with Quanser Inc. (Markham, ON, Canada; www.quanser.com) have developed a haptic robotic system designed for upper limb stroke rehabilitation. IATSL has begun development on rehabilitative exercise software for this device, including rehabilitative games that offer therapeutic exercise framed as a game. The basic components of the hardware of the robotic device are shown in Figure 4.

![Haptic Stroke Rehabilitation Robot System](image)

Figure 4 - Haptic Stroke Rehabilitation Robot System
The robotic system has capabilities and functions that are outside the scope of this study, but currently the robotic system is able to apply haptic feedback, resistance, and assistance, and is also able to record a variety of parameters of motion. It is a low-cost robot with two degrees of freedom in a planar workspace, and can be controlled using either the left hand or the right hand with the same symmetrical ROM, despite the physical appearance of its arm.

4.2. Game Development

Game Development Overview - for the "Masked" and "Unmasked" Games:

Stage 1: Consultation for Generating Game Ideas and Features

Stage 2: Grouping of Game Features for Development Prioritization

Stage 3: Creation of Checklist of Features to Match between Games

Stage 4: Development of Initial Game Iterations

Stage 5: Volunteer Testing of "Equivalency" between Games

Stage 6: Development of Later Game Iterations

Stage 7: Final Testing with Stroke Survivor Study Participants

An advisory panel composed of stroke rehabilitation therapists, stroke survivors, and game design students met together. They were asked questions regarding game design considerations, and were also asked to contribute initial game ideas (see Appendix A - Advisory Panel Questions and Brainstorming). Groupings or patterns in the advisory panel’s input were sought in order to organize the information and make it easier to prioritize the large number of features for later development. Most of the advisory panel’s input could be grouped under the different conditions of flow used in human factors design of fun user experiences (Human Factors International Inc., 2013), which are also worded as questions in the Flow Condition Questionnaire (Human Factors International Inc., 2013). Flow is “a state of peak enjoyment, energetic focus, and creative concentration” (Csikszentmihalyi, 2000), when one experiences being “in the zone”, enjoying and being intrinsically motivated to do something, where attention is so focused that the passage of time is not noticed (Human Factors International Inc., 2013). The groupings that the
input fell under were categorized as: “it’s clear what to do”, “it’s clear how to do it”, “it’s clear how well you are doing”, and “there’s a clear feeling that you are able to do it”. As well, there was some input that did not fit neatly into these groupings, and so some ad hoc groupings based on the context of this study were made. These are the other groupings: “time”, “choice”, “robot/game features to account for cognitive deficits”, and “miscellaneous”. These groupings and the conditions of flow mentioned in Human Factors International Inc. (2013) also inspired some of the questions in the custom questionnaire used in this study. Each of the game design considerations from the advisory panel’s input would be later incorporated into the final designs for the Masked and Unmasked games. The game design considerations included things along the lines of accounting for cognitive deficits (such as visual complexity and subtle reminder of the task), motivating stroke survivors (such as managing challenge and non-patronizing feedback methods), and targeting physical needs (such as fatigue, ROM, and different needs and targets for Chedoke-McMaster Stroke Assessment stages). The Chedoke-McMaster Stroke Assessment (CMSA) assesses the physical impairments and disabilities that impact stroke survivors’ lives (Gowland et al., 1993; Miller et al., 2008).

This project focused on development of the games and graphics in Java (Oracle Corporation, California, United States), using Eclipse (The Eclipse Foundation, Ottawa, Canada) as the Java programming interface, and a game development framework called libGDX (Mario Zechner, Badlogic Games) for the graphics.

Three games were tested for this study: one with the existing “Basic” design already being used with the robot, one developed with the “Unmasked” design, and one developed with the “Masked” design. Images of the final designs of these games were shown in Chapter 2. Initial iterations for the Masked and Unmasked games were generated during the advisory panel meeting, but were revised after some practical concerns such as streamlining features and “equivalency” between the two games to try to isolate the “masking” feature, and after subsequent design considerations and suggestions obtained from “equivalency” testing. This happened during an “equivalency” pre-testing phase where five volunteers from the lab tested the “Masked” and “Unmasked” games, answered questionnaires evaluating how “equivalent” the two games were (besides the “masking” aspect), and also gave feedback to improve the games.

The three games in this study were designed to all make use of the same therapeutic motions with the robot system. To ensure that the same therapeutic motions are used in the games
and that we are isolating differences between the “Unmasked” versus “Masked” games (for “equivalency” between them), the two games were designed “side-by-side” in order to have features concurrently compared to match as closely as possible in several aspects. These aspects included the same level of cognitive requirements (same complexity of virtual environment, actions, features, etc.), same accessibility for vision/perception (same modes of feedback), same game categories (for example, both games are either strategy or action), and same performance markers/motivations (same system for points, progress, etc.). These two games also shared the same game mechanics and were designed step-by-step with each aspect in mind. A checklist was used during game design to ensure comparable game features. See Table 3 - “Game Features to Match across Games”.

Note that this checklist is not exhaustive and was used as a guide. Features were to be added to this checklist as seen fit during development.

Table 3 - Game Features to Match across Games. Note: x indicates not applicable.

<table>
<thead>
<tr>
<th>Basic</th>
<th>Masked</th>
<th>Unmasked</th>
<th>Game Features to Match:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Motions desired</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Types required (to enable motion counts, limit to those with a clear beginning &amp; end, not possibly difficult-to-evaluate circular paths)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Types of motions users would likely use intuitively</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Number of repetitions required to reach game goal</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Order of motions used (if applicable)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Number of possible actions of any in-game avatar</td>
</tr>
<tr>
<td>x</td>
<td></td>
<td></td>
<td>Number of options in user menu (as applicable)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Environment complexity (visual)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Environment objects in same locations</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Size of objects relative to screen size</td>
</tr>
<tr>
<td>x</td>
<td></td>
<td></td>
<td>- Colour palette complexity (number of colours)</td>
</tr>
<tr>
<td>x</td>
<td></td>
<td></td>
<td>- Number of visuals in environment</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Feedback</td>
</tr>
</tbody>
</table>

Note: x indicates not applicable.
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Types (points, levels, progress bars; can be themed/stylized per game)
- Senses used (sound, visual, haptic, or combination used)
- Number of feedback items used for each sense
  Different levels of game
  - Amount of resistance provided
  - Variation in speed of game
- Environmental factors (e.g.: non-player character as “social pressure”)
  - Presence
  - Amount, change
  Speed/tempo of background music (if any).
  Levels
  - Progression: types, number of changes
- Story development

Certain features did not apply to the “Basic” game, because by definition it does not overlay a game nor simulate a story overtop of the exercise as the “Masked” and “Unmasked” games do. Certain features were omitted from all games, namely haptics (such as resistance, assistance, and shaking) and background music. The omission of music and haptics from all games does not provide any one game an advantage or disadvantage. Haptics were omitted because it could bring an extra source of variability between sessions, adding unnecessary complexity for this study’s purposes of analyzing the differences between the games. Music was omitted because any existing associations that a participant may have with certain music or music genres may have an undue impact on their preference and/or impression of any one of the games presented to them, especially if games may use different background music to fit into the overall designs distinct to each game. Also, the presence of music has already been shown to have the potential of improving gameplay experience (Lipscomb & Zehnder 2004) and may improve gameplay experience for all the games in this study, although participants are more likely to perform optimally
when they choose their own music, as opposed to when researchers choose the music (Cassidy & Macdonald, 2009). As well, the relationship between music tailored for a game and gameplay experience is complex (Lipscomb & Zehnder 2004), and hence outside the scope of this study.

Accessibility considerations were also made with specific considerations made for possible cognitive needs of the stroke survivor participants. Listed are examples of accessibility design considerations included in the study’s games (Ellis et al., 2015):

- Showing large virtual “buttons” for clear visibility and responsiveness.
- Teaching game mechanics within context or within the game itself.
- Allowing play to start without requiring navigation of multi-level menus.

4.2.1. Pre-Testing for Masked-Unmasked Game Equivalencies

Game equivalency was evaluated by five volunteers on the research team prior to testing with stroke survivor participants. Each volunteer tested the two games and provided ratings and free-form feedback. Their suggestions for design improvements were also considered. Between some of the testing sessions, there was time to address obvious design issues before the next volunteer tested the games. For example, certain graphics were updated to make it clearer what they were without explanation, and the basic “story” in each game was modified so that the tasks were more likely to implicitly require the same motions in both games.

To complete the required testing for equivalency between the developed “Masked” and “Unmasked” games, a subjective questionnaire was used. The results were used for design purposes before finalizing the designs for play in the study. The questions were on a five-point Likert scale (1 = “Not At All”; 3 = “Somewhat / Unsure”; 5 = “Very”):

1. How equivalent is the quality of graphics for both games?

2. How equivalent are the player characters?

3. How equivalent are the non-player characters?

4. How equivalent are the animations when the player successfully reaches targets?
(5) How equivalent are the animations when the player does not successfully reach targets?

(6) How equivalent are the fonts?

(7) How equivalent are the sound effects?

(8) How equivalent are the visuals (graphics, etc.)?

(9) How equivalent are the audio?

(10) How equivalent is the audio feedback?

For further pre-testing for design purposes, there were also questions based on the suggestions for phrasing questions from Freeman (2015):

(11) How did you feel about the firing game mechanic? (“Masked” game)

(12) How did you feel about the wiping game mechanic? (“Unmasked” game)

(13) How did you feel about the other characters in the game?

(14) Any other comments? (If none, write “None”.)

Again, these questions are only for the “Masked” and “Unmasked” games, not the “Basic” game, as they were meant for test for equivalency between these two games so that one does not have an unfair advantage that comes from a feature unrelated to being “Masked” or “Unmasked”.

After calculating the questionnaire results of the “equivalency” pre-testing, all the Likert scored questions (questions 1-10) were found to be at least “above passing” scores (i.e., average scores for all questions were above 3), and total scores averaged at 74% (i.e. average total score was 37 out of 50), across the five volunteers. Feedback and comments on design ideas were also taken into consideration between these volunteer “equivalency” tests, in order to fix some obvious issues in the programming, to make some graphics clearer, and to further improve the equivalency for the next volunteer’s test. When the game mechanics were updated before later volunteers were to test the games, questions 11 and 12 were re-worded to match their respective games. These questions still essentially provided an opportunity for free-form feedback to inform
improvements in “equivalency” between the two games, possibly to provide more reasons behind the simple Likert questions. After the volunteer testing, it was then decided that the study proceed to testing with stroke survivor participants. It should be noted that time on task, motion repetitions, “hits” (scores), and percent of active ROM were also recorded for the short “equivalency” pre-testing sessions with volunteers, simply as an afterthought for later reference. The number of motion repetitions and “hits” were both about 1.6 times higher on average for the “Masked” game than for the “Unmasked” game for these volunteers, but this was not noticed because this was calculated after noticing how stroke survivors consistently missed a certain target in the “Masked” game (this was later addressed in the data analysis by removing one tenth of misses to adjust for the “false” target misses). The number of motion repetitions as defined in this study (Appendix B - “Measurement Definitions and Details”) can sometimes have false readings of repetitions, however the “hits” do not. The 1.6 times higher difference in “hits” was not due to a software measurement error because the “hits” was simply calculated from the number of objects successfully moved, and was straightforward to program. However, this 1.6 times higher difference did not seem significant because there was no explicit time frame provided to volunteers (making it hard to interpret, especially with person schedules and other external factors), and because the focus was on the “equivalency” between the games based on qualitative “feel”, not based on robot measurements of repetitions.

One volunteer’s suggestion for “equivalency” was major in that it made the games much more equivalent for later volunteers. This suggestion made the games use the same motions by considering the possible motions that users would intuitively think of using. For example, compare these two situations: wiping a table can use back-and-forth motions or circular motions (very different from aiming a cannon), but moving items around uses a motion from A to B (so in the context of a computer game, moving ingredients to a pot should be similar to moving fuel to a tank). Thus, with this major improvement in “equivalency”, as well as with other suggestions implemented, it was decided that we prepare for testing with stroke survivor participants (after conducting mock trials).

The final versions of the Masked and Unmasked games were titled "Fuel The Ship" and "Make The Soup", respectively. For "Fuel The Ship" (Masked game), the objective was to fill a spaceship with fuel. The ship's body would act as a progress bar that fills up to the top when
more fuel tanks are added. In higher levels, there would be two different kinds of fuel (just different graphics), and alien spaceships would move around and could “steal” fuel (removing targets). For "Make The Soup" (Unmasked game), the objective was to fill a pot with ingredients. Similarly, the pot would act as a progress bar that fills up to the top when more ingredients are added. In higher levels, there would be two different kinds of ingredients (just different graphics), and a timer that would be shown counting down until ingredients disappear (“timing out” and removing targets).

4.2.2. Mock Trials

Two mock trials were conducted by the research team to estimate the duration of set up for study sessions. The mock trials also helped to familiarize research team members with the procedures and any possible adjustments to the protocol that needed to be made before testing with actual stroke survivor participants.

4.3. Study Design

This study used a mixed methods and exploratory study design. The study design was “mixed methods” because both quantitative and qualitative data were collected (Johnson et al., 2007). The study design was also “exploratory” because the field of computer game design in robotic stroke rehabilitation is relatively new, and exploratory research helps to initially define concepts and problems (Shields et al., 2013). Even within the broader field of rehabilitation robotics, more research is still needed to better know what exactly needs to happen for the nervous system to adapt and overcome physical impairment (Reinkensmeyer, 2016). Exploratory research is conducted with provisional hypotheses that can be updated if there are unusual findings, but such hypotheses provide an initial way to structure data collection and make analysis more manageable (Shields et al., 2013). The findings gained from testing these initial hypotheses can provide a reasonable starting point for later research (Shields et al., 2013). Exploration also involves finding and sometimes inventing assessment criteria, but also requires openness to new evidence (Shields et al., 2013). Exploration is needed in order to identify key issues and variables (“Research Methods,” Retrieved July 2015) when we do not have anticipated answers (Shields et al., 2013). Such variables are the aspects of beneficial game design to be identified in this study.
4.4. Participants

This study required at least eight participants, based on the minimum of a recommended 10±2 samples from a meta-analysis of usability study sample sizes (Hwang & Salvendy, 2010). The inclusion and exclusion criteria for this study were as follows.

**Inclusion Criteria:**

- Stroke survivors in chronic (over six months post stroke) stage of recovery
- Right or left upper limb impairment as a result of stroke, stages 3-5 on Chedoke-McMaster Stroke Assessment (CMSA, Gowland et al., 1993)
- Physically and cognitively able to participate in a session of at least 10 minutes (and up to 60 minutes) of moving the affected upper limb, with a total session duration of up to 90 minutes including time for set up and questionnaire
- Able to understand and speak English
- Able to participate in an interview and respond to questionnaire questions
- Able to give informed consent
- Able to attend 3 sessions at Toronto Rehabilitation Institute (University Centre Site)

**Exclusion Criteria:**

- Presence of pain when moving the upper limb, such as significant self-reported joint pain when moving the upper limb

Stroke survivor participants were recruited through Toronto Rehabilitation Institute – University Health Network by therapist referral, through the centralized recruiting process, and through the Research Volunteer Pool Recruitment Process.

4.5. Data Collected

What follows are summaries of the data collected from participants. They include general data used for verification purposes, as well as groupings of measurements used to answer the
three specific research questions. Further definitions and details of the measurements used in this study can be found in Appendix B - “Measurement Definitions and Details”.

General Data:

- **Demographics** data.

- **Game type** played by participant for each session.

- **Verbal feedback** and responses to open-ended questions, and feedback freely given by participants, unprompted (documented in written notes).

- **Video recordings** of all sessions (from two simultaneous camcorders in two different perspectives in order to simultaneously capture the participant’s facial expressions, arm motions, general gaze away from/towards the screen, circumstances of breaks in attention, and any concurrent events in the game on the computer monitor) for verification purposes and additional observations as overall support for findings from questionnaire and robot data.

**Data Collected for Research Question #1 (Time):**

1) Average **time on task** (recorded by robot during gameplay) was the time spent playing a game. Participants were asked to play each game for a minimum 10 minutes, however they were given the choice to stop at any time, even before the 10 minutes. Suggesting a minimum gameplay time to participants was to provide adequate play experience to provide feedback, but also to dispel any feeling of obligation or pressure to play for extended periods.

   This study obtained the average amount of time spent playing each game. A meta-analysis found additional therapy time to have a small but positive effect on instrumental ADLs, with clinically-relevant individual effect sizes (Kwakkel et al., 2004).

**Data Collected for Research Question #2 (Motion Reps):**

1) Average **number of motion repetitions** (recorded by robot during gameplay).
2) Average **number of hits** for successfully-reached targets (recorded by robot during gameplay).

3) Average **number of misses** (recorded by robot during gameplay).

4) Average **percent of hits** (calculated from the hits and misses).

5) Average **ratios of ROM usage** in-game compared to initial active ROM setup (recorded by robot during gameplay) in four cardinal directions ("pull", "push", "left", "right").

6) Average **rate of repetitions per minute** (recorded by robot during gameplay).

7) Average **rate of hits per minute** (recorded by robot during gameplay).

One of this study’s goals was to identify the game design that caused stroke survivors to perform the greatest number of and rate of motion repetitions, as well as successful ones, i.e., to perform reps that were greater in number, were faster, had a higher success rate, and were larger (taking up more of the patient’s ROM). Several measures were used for repetitions because they each provide only partial information, where one measure can be used to inform where another is limited. These are the aspects of repetitions covered by different measures: number, speed, success, rate, and size. For details on the rationale behind including each of the different measures for repetitions listed above, see Appendix C - “Rationale for the Measures for Repetitions”.

**Data Collected for Research Question #3 (Preference):**

1) Responses to **questionnaires** for “implicit preference” (see Appendix D - “Sources and Groupings of Questionnaire Questions”) using questions from the Intrinsic Motivation Inventory (IMI), questions from several other source papers, and questions created ad hoc for the purposes of this study.

2) Responses to an additional question at the third and final session for each participant: indication of “explicit” **game preference** and the top reasons why.

The questionnaire used in this study had questions inspired by or taken directly from several sources: user advisory meetings, the IMI, and other literature. A few questions were added that were inspired by the Flow Condition Questionnaire (Human Factors International Inc.,
2013), to target conditions of flow (mentioned earlier under Game Development). One question was adapted from McKay & Maki (2010) to have participants self-report how often they would be willing to play the games, because it was thought it could be useful to ask for specific frequencies such as “daily” or “weekly”. There is apparently no existing “standard” in industry for video game testing, and that each company does it differently in the specifics. Furthermore, even if there were a standardized set of questions or questionnaires, they would not get published (personal communication, 2015, May 14). The full questionnaire can be seen in Appendix E - “Custom Questionnaire for Stroke Survivor Participants”.

This study took various measurements of game time on task, motion repetitions, and preference, in order to infer a “better” game design (or at least beneficial aspects for future designs) for stroke survivors. A mix of both qualitative and quantitative measures was used because getting qualitative input from end users can help interpret the “why” behind the numerical data, however end users—even subject-matter experts—do not always self-report consistently with how they actually perform tasks, and a combination of both user-reported data and observation data is necessary to ensure that designs fit actual use and not simply how users think they would use such designs (Morita and Cafazzo, 2016).

This study had three specific research questions, each of which can help to build a complete picture. Measuring game preference and motivation helps us know whether users will want to play a game at all in the first place. Measuring time on task helps us know how long users will keep playing a game once they start (independently of how highly they self-report enjoyment). Measuring motion repetition number, rate, and size (i.e. how many, how fast, and how big the repetitions are) helps us know how well they make use of the time they spend playing a game. The qualitative data will help with interpretation of the quantitative findings. Combining both preference and performance measurements will help determine whether there are any discrepancies between subjective self-reports and objective measures.

4.6. Procedures

Participant Consent
Before a potential participant had their sessions scheduled, the researcher screened them over the phone based on all of the inclusion and exclusion criteria for this study (with a preliminary check of CMSA stage) to avoid making stroke survivors make unnecessary trips only to find out they were not eligible. The researcher used a list of steps typed out as a guideline to ensure all criteria and follow-up steps were followed reasonably. If the researcher was uncertain in any individual potential participant, the supervisor was consulted. The day of the potential participant’s first scheduled session, the therapist in the team would screen the participant using the CMSA. If the participant’s arm movements were within CMSA levels three to five (inclusive), then by that point all the inclusion and exclusion criteria had been passed. Before a participant began their first session, they also had to provide written informed consent. The researcher reviewed with each participant: the consent form, the research goals, what they were expected to do, prototype features, anticipated risks and benefits, and the data to be collected (including camcorder audio and video recordings). Finally, following the consenting process, the researcher asked the participant questions from the demographics form, before getting the participant and experiment set up for gameplay.

Sessions and Game Type Played

Each participant had three sessions, and played one of the games for each session. Each participant was assigned one of the orders of presentation of the games to play (i.e. one of the six possible permutations of the order of presenting the three games). There was at least one week between sessions for each participant, to serve as a “washout period” to reduce possible carryover effects between sessions (i.e. influence on results due to having played one game first instead of another).

This study used counterbalanced measures. The data was also analyzed to find which game had the highest values for measurements compared both between and within subjects. Each participant being presented with all three conditions (“Basic”, “Unmasked”, and “Masked”) yielding six possible permutations and hence covering all possible orders of presentation of the games (Shuttleworth, Retrieved 2015). To avoid ordering and learning effects (MacKenzie, 2013), all possible orderings for presenting the games were used and assigned to participants by the order the participants were registered into the study session schedule, with washout periods
(i.e. gaps between sessions) scheduled to be about one week between sessions for each participant. Participants had one session with each condition. See Appendix F - “Game Assignment Permutations” for a table of assigned permutations.

Note that this process was to be cycled for every six participants who enter this study, since there are six permutations to cover. Hence, participant #7 in our example has the exact same order of presentation as participant #1. The order of the permutations was decided before recruitment started, and participants were assigned permutations based on order of recruitment into this study.

Participants were told that they would be playing three different games, one on each session, but not which game they were playing (unless they asked). The other two people present were the researcher, as well as a therapist or a trained research assistant there for the first session and then as needed for the second and third sessions to help get the participant set up and comfortable (including asking if they needed to take a break or to stop), and to watch for signs of fatigue during gameplay.

For each session, the participant received reminders/demonstrations for familiarizing them on how to use the robot in order to be able to play the game, before they play the game for data collection purposes. The general procedure for each day was standardized and listed in general guideline steps.

Session Duration

Each session was expected to take about 90 minutes in total, which included demonstrating/reminding of system use, setting up before gameplay, responding to the demographics form, playing the games for a recommended minimum of 10 minutes and up to one hour, and responding to the questionnaire.

Session Data

During gameplay, the robot recorded data, and the research team asked participants informal questions related to the context of each specific participant's interaction with a game. These questions would be used to gain insight into the specific participant's gameplay experience that would be otherwise difficult to anticipate.
After the participant decided to stop playing, the investigator administered the custom questionnaire (Appendix E - “Custom Questionnaire for Stroke Survivor Participants”).

The investigator provided a print-out of a seven-point scale with “anchors” marked (see Table 4 - Anchors for 7-Point Scale), for the participant to hold on to while they respond. Any other feedback that participants freely provided during the session without being prompted to do so (i.e. beyond the questionnaire), was noted by the investigator. Free-form responses, whether prompted or un-prompted, would help contribute to insight into the beneficial aspects of the designs of the games.

### Table 4 - Anchors for 7-Point Scale

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Not At All True</td>
<td></td>
<td></td>
<td>Somewhat True</td>
<td></td>
<td></td>
<td>Very True</td>
</tr>
</tbody>
</table>

After a participant had played all three games, they would answer the last question in the questionnaire, which asked them to directly compare the three games and provide their top three reasons for their selection. Clips or images were provided as reminders of the three games.

All sessions were video recorded to ensure accurate documentation and to supplement findings of engagement from the questionnaire data and robot data.

Honorariums of $60 were provided as a token of appreciation for participants’ time. If any participants dropped out before their three sessions were completed, the amount was prorated for the number of sessions in which they participated.

### 4.7. Analysis

Although this study used a mixed methods design, it did not include a formal qualitative analysis of all the verbal feedback in the qualitative data. However, this study used the qualitative data to support the quantitative data. The data from testing each of the three game designs were compared, and also beneficial aspects from all three designs for future development were identified.
What follows are descriptions of how the data collected from participants were analyzed. There were analyses of data for each of the three specific research questions, but also analyses of the more general data (such as patterns in the demographics and free-form feedback) in relation to the data for the specific research questions. Further details on the analysis of the questionnaire follows after.

General Data:

- **Demographics** data were summarized in a table to describe the participant sample. Sample population homogeneity was analyzed upon inspection and with coefficient of variation to make use of a rule of thumb to explicitly describe population homogeneity.

- **Game type** played was matched with the corresponding session’s measures for each participant respectively, to make comparisons between the highest values of the indicators for time on task, number of motion repetitions, and preference. Such comparisons were made both within and between subjects. Other supplementary observations were made when making these comparisons, such as whether the game that participants say they prefer was the same game that they spent more time playing, or was the same game in which they performed more therapeutic movement repetitions.

- **Verbal feedback**, responses to open-ended questions, feedback freely given by participants, observation notes, and transcriptions were reviewed by the researcher to find qualitative data that could explain the findings in the quantitative data.

- Though not originally planned, there were **additional statistical analyses** performed. Statistically significant differences (both ANOVA and multiple t-tests) and effect sizes (Cohen’s d) were found that could be used as preliminary indicators for future work.

Data Collected for Research Question #1 (Time):

- **Time on task** (calculated as averages and standard means) to each game were compared to find the highest value as one of the indicators of the most beneficial game design(s).

Data Collected for Research Question #2 (Motion Reps) - 7 Measures:
• Average values and standard deviations were calculated for the following measures and used as indicators of the most beneficial game design(s): the **number of motion repetitions**, the **number of hits** for successfully-reached targets, the **number of misses**, the **percent of hits**, the **ratios of ROM usage** in-game compared to initial active ROM setup in four cardinal directions ("pull", "push", "left", "right"), the **rate of repetitions per minute**, and the **rate of hits per minute**. The number of motion repetitions are to be compared to the recommended number of 100 or more reps per day (Kimberley et al., 2010).

Data Collected for Research Question #3 (Preference) - 2 Measures:

• Responses to **questionnaires** were used to infer “implicit” preference, mainly using the IMI constructs of “Perceived Usefulness”, and “Enjoyment” (calculated as average values). Higher IMI scores are to be used to indirectly determine game preference, and are to be compared to which game was favoured by the “votes” for game preference to check for inconsistencies.

• Indication of **game preference** was used for “explicit” self-reported preference (counted as “votes”). The top reasons for game preference were also analyzed for reasons for and against the different game designs.

What follows are further details on the calculation of scores for questions in the questionnaire, especially the IMI questions.

**Questionnaire - Scoring Questions**

Scores for the IMI questions were used as part of the analysis to determine aspects (such as “masking”) that contribute to user game preference. Overall IMI scores were calculated for each construct, for each game, across all participants. For each participant’s questionnaire, IMI scores were calculated for each grouping of questions, i.e. for each IMI construct (see Appendix D - “Sources and Groupings of Questionnaire Questions”). For each construct, the average of the scores of all participants for that construct were grouped across the type of game they played for the session that their questionnaires were filled (“Basic”, “Masked”, and “Unmasked” games). The three final numbers for each construct (i.e. overall scores for each of the three constructs for each of the three games) were compared for further analysis.
Score Calculation: IMI scores were calculated for Perceived Usefulness and Enjoyment from the questions from the IMI (questions 2, 3, 5, 6, 7, 9, 11 as listed in Appendix D - “Sources and Groupings of Questionnaire Questions”). The method of scoring for individual questions was the same for all IMI questions except question 11, which would be reversed = 8 - score. The score for each construct was calculated as the average across all questions for each construct. For example, the score for Perceived Usefulness = (question#3 + question#5 + question#6 + question#9) \div 3. Questions #3, #5, #6, and #9 all IMI-style seven-point scales, so if the scores for questions #3, #5, #6, and #9 were respectively 6, 3, 5, and 6, then the score for Perceived Usefulness would be = (6 + 3 + 5 + 6) \div 4 = 5 and are further analyzed with standard deviation and mean values for the participant group.

Question 12 was calculated on its own, but still used to gain insight into motivation and willingness to adhere to activities. Scores for the other custom questions based on user advisory meetings and from other sources (#1, #4, #8, #10) were calculated separately for further analysis.

Questions 13, 14 and 15 were used for qualitative feedback. For these questions, participants were encouraged to freely answer whatever they want for these open-ended questions. Question 16 is for the participants to explicitly give their game preference, and would only be asked if the participant already played both games, as in a second session. Questions 2b, 3b, 4b, 8b, and 9b were for asking participants to provide reasons for their responses to the respective questions 2, 3, 4, 8, and 9.
5. Results

5.1. Demographics Data

Table 5 below shows totals, averages, and standard deviations for the demographic data of the participants.

Participants were already doing at least some kind of exercise for the upper limb outside of this study, sometimes with much variety. Some self-reported doing a mix of individual exercises at home or group activities where services and facilities were provided. This included specific stretching or muscle exercises (such as using a ball or tensors), or more recreational-style exercises (such as pool therapy, mall walks, or a Wii system provided at a centre). Although for not specifically for the arm, one participant noted playing a few iPad apps with their son in order to work on their fingers (for fine motor control).

The sample population for this study exhibited population homogeneity with similar participant ages. Although the standard deviation can give you a sense of the spread of variation, the coefficient of variation comes with a rule of thumb that explicitly answers what is considered population homogeneity. The coefficient of variation for age was calculated to be 0.193. This value is less than one, and hence is considered to indicate homogeneity in the population, at least in that one measure in the available demographics data. This ignores motivation towards therapy, gender, affected side, and handedness, but computing the coefficient of variation for these other data may be misleading because they are not measured on a ratio scale, and the coefficient of variation should only be used on ratio scale data (Abdi, 2010). There were three times more males than females, and all participants were right handed except for one who self-reported as being ambidextrous. All except one participant self-reported high levels of motivation towards therapy when asked in the demographics questionnaire (the unique response was a self-report of uncertainty but settling on “High/Medium”). One participant explicitly commented “that’s why I’m here”. One participant said that their son was the reason for their high level of motivation towards therapy. Sample population homogeneity will be touched upon later in the discussion section.
Table 5 - Participant Demographics: (n = 8). CMSA = Chedoke-McMaster Stroke Assessment (stage 1 = most impairment, stage 7 = no impairment, based on being able to perform a pre-determined list of motions for each stage). The question for “Current Motivation towards Therapy” was given with three different word answers, and is coded here as 3 = High, 2 = Medium, 1 = Low (for this question, the options were presented to participants as word answers, not as number codes).

<table>
<thead>
<tr>
<th>Demographics:</th>
<th>Total</th>
<th>Average</th>
<th>Standard Deviation</th>
<th>Coefficient of Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>55.6</td>
<td>10.7</td>
<td>0.193</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affected Side</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMSA (Arm)</td>
<td>3.4</td>
<td>0.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Handedness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left Handed</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right Handed</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambidextrous</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current Level of Motivation towards Therapy</td>
<td>2.9</td>
<td>0.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.2. Research Question #1: Time

Research question #1 aimed to determine which game design has users wanting to keep playing for the longest period of time. To infer this, the time on task for each of the games played was recorded (see Figure 5). According to Figure 5, the “Masked” game had the highest average time on task (24.2 minutes), but also the highest standard deviation (11.6 minutes).

**Time on task (minutes):** As shown in Figure 5, the “Masked” game had the highest average time on task (24.2 minutes), but also the highest standard deviation (±11.6 minutes). The gameplay time of any of the games ranged from 4.1 minutes to 46.5 minutes (note the suggested minimum of 10 minutes and the required maximum of 60 minutes). One participant played the “Basic” game for less than the recommended time of 10 min (they requested the researcher to tell them when 4 minutes of gameplay had passed so that they could stop, meanwhile indicating an
apparent strong dislike with comments like “two minutes seems like an hour”, and “eyes falling asleep, I need to be challenged”). Another participant’s gameplay time for the “Unmasked” game was smaller than it could have been due to the researcher’s misunderstanding of what was said by the participant with communication difficulties (expressive aphasia) when asked if they would like to stop. However, even if an extra 10 or 20 minutes were added to that session’s time on task, the average time on task for the Unmasked game would still only rise to about 19 or 21 minutes, which would not change the relative ranking of the three games in terms of average values. Almost all participants wanted to finish the Masked and Unmasked games up to level 11 (because of the text saying how many levels out of 11 they had completed), and then some would stop, sometimes because of and sometimes despite knowing that the game kept going infinitely past level 11.

By itself, time on task may not be a strong indicator, but combined with the other research questions and qualitative results, a fuller picture may be put together. See Appendix I - “Data per Subject” for extra figures showing measurements by participant. Additional statistical analyses were also performed on the data.

![Figure 5 - Time on Task (minutes)](image)

5.3. Research Question #2: Repetitions

Research question #2 aimed to determine which game design contributes to the greatest number, rate, and size (i.e. Range of Motion used in workspace) of motion repetitions. To infer this, the following seven groups of measurements were used: the average number of motion repetitions (see Figure 6), the average score, i.e. the number of successful target hits (Figure 7), the
average rate of repetitions per minute (Figure 8), the average number of misses (Figure 9), the average percent of hits out of targets presented (Figure 10), the average rate of scoring or the number of successful target hits per minute (Figure 11), and the ratio of the maximum reach in the four cardinal directions during game play versus during initial active ROM calibration (Figure 12).

The number of reps averaged over the Masked and Unmasked games was 1012.5 reps, and the number of “hits” averaged over the Masked and Unmasked games was 600 “hits”. Measuring 600 "hits" in an individual participant (and not in an average) would mean that at least 300 motion reps were made by that participant, because to be able to hit two targets, at least one motion rep must be made. The motion reps can be thought of as relative paths made, and the “hits” as positions reached (i.e. one motion rep = pick up and drop off item, ideally within one linear motion; pick up item = one target hit; drop off item = another target hit). Each “hit” corresponds to a target, but two targets are the minimum for inferring a rep from the number of “hits”. Individually, the average number of reps was 356.5 for the Basic game, 1191 for the Masked game, and 750 for the Unmasked game, which were all above the minimum of 100 reps.

If we consider the approximately 1.6 times more reps and “hits” on average for the Masked game than the Unmasked game during the volunteer “equivalency” pre-testing (ignoring the lack of explicit rules on time given to volunteers), then we could roughly adjust the average number of motion repetitions and the average number of hits by dividing their averages by 1.6. This results in the following average values for the Masked game: 744.4 reps and 411.3 hits. These values would be lower than the Unmasked game’s values, potentially affecting the recommended game design. These alternate values are reported in the respective figure descriptions for Number of Motion Repetitions (Figure 6) and Number of Hits (Figure 7). However, using the approximately 1.6 times higher difference itself in the first place is also difficult to interpret because of the informal nature and lack of explicit timeline set for the volunteer pre-testing, and because the focus was on testing the “equivalency” of the Masked and Unmasked games based on the qualitative feel of the features, and not based on performance measures. Adjusting later analysis by dividing the average values for reps and hits by 1.6 downgraded the Cohen’s d effect sizes for the Masked vs. Basic game in Reps (#) and Reps/Min from medium effect size to small effect size, and out of the other effect sizes, only noticeably changed the one for Masked vs. Unmasked in terms of Reps (#) from a medium effect size to a negative effect size.
See Appendix I - “Data per Subject” for extra figures showing measurements by participant.

**Figure 6 - Number of Motion Repetitions.** The most motion repetitions recorded on average were performed in the Masked game (1191 reps), but also with the highest standard deviation (±778 reps). (Alternatively, the Masked game average changes to 744.4 reps if adjusted by the “approximately 1.6 times higher difference” from informal volunteer testing, and would be lower than the average reps for the Unmasked game.) * = Statistical significance.

**Figure 7 - Number of Hits.** Game player scores were counted as the number of successfully reached motion targets (the number of Hits). The Masked game had the highest average of recorded hits performed (658 hits). (Alternatively, the Masked game average changes to 411.3 hits if adjusted by the “approximately 1.6 times higher difference” from informal volunteer testing, and would be the lowest average of all three games.)
Figure 8 - Repetitions per Minute. The highest rate of reps per minute was on average performed in the Masked game (48.2 reps/min). On visual inspection, both the Masked and Unmasked games had higher values than the Basic game and the respective standard deviation error bars to not overlap with the Basic game’s. * = Statistical significance.

Misses (#): For the Masked game, the “aliens” (non-player characters) were intended to reach items within a fixed time, however during use it appeared that the time it took for the “aliens” was not moving consistently to take the same amount of time to “steal” items from players (due to an error in the speed-distance algorithm), and were often faster at “stealing” items than the Unmasked game’s timer at removing items. It was noticed that the Masked game would consistently have one out of the ten target waypoints stolen by the “aliens” (the one nearest to the player's body, which is nearest to where the “aliens” would be returning to home positions in later levels). The implication of this observation is that the data can be “cleaned” by removing one tenth of the misses to attempt to correct for the number of “false misses” of targets due to the “consistently inconsistent” speeds of the “aliens”. However, even if you try to correct the average number of misses for the Masked game (which then becomes 162) and you ignore the large standard deviation for the Masked game, the average number of misses for the Masked game is
visually still much higher than the averages for the two other games.

Figure 9 - Number of Misses: * = Statistical significance.

Figure 10 - Hits (percent of targets presented). After removing one tenth of the misses (because of bias seen from “equivalency” test results) to correct the average number of misses for the Masked game, the Masked game’s value goes to 80.24%, which was still lower than the means for the other two games. * = Statistical significance.
Figure 11 - Hits per Minute. The Basic game had the highest average hits per minute (35.3 hits/min).

The repetitions sizes may have had additional noise in the data due to participants’ movement compensation during in-game reaching actions. Compensation was not encouraged, but neither was there full enforcement of removing compensation throughout this study’s sessions. This limits the interpretation that larger reps were performed in the Masked game than in the other games, as there could be extra noise in the data for all three games.

Figure 12 - Repetition Sizes (percent of axis out from centre of ROM); “PULL” (%), “PUSH” (%), “LEFT” (%), “RIGHT” (%): It should be noted that at certain times participants had major compensation during gameplay in order to reach some targets. * = Statistical significance.
5.4. Research Question #3: Preferences

Research question #3 aimed to determine which game design has the most favourable stroke survivor self-reported preference scores (perceived usefulness, enjoyment, explicit indication of game preference).

Table 6 below tallies the number of participants who preferred each game over the other two, and also shows the average score for IMI categories (i.e. IMI subscales). Each participant was asked which game they preferred out of the three. One participant mentioned seeing use in all three games, but for different stages of recovery, based on what they perceived in the different games as different complexities of visuals and levels of challenge in both ROM and speed/timing. One participant saw both the Masked and Unmasked games as the same but with different graphics, so in the “Explicit Choice” column, one count was added to both the Unmasked row and the Masked row. If that response is ignored, the Masked game still has the most votes, but the Unmasked game has fewer votes than the Basic game.

**Table 6 - Participant Game Preferences and Intrinsic Motivation Inventory (IMI) Scores.** IMI = Intrinsic Motivation Inventory; scores out of 7, with 7 meaning “very true” to statements listed under the IMI categories of questions; *= Unmasked gets 1 and Masked gets 4 if you ignore the response from one participant who said both Masked and Unmasked games (“tied”). This would result in Masked having 4, Basic having 2, and Unmasked having 1 “vote”.

<table>
<thead>
<tr>
<th>Preferred Game:</th>
<th>Explicit Choice (“votes”):</th>
<th>Average Score of IMI Categories:</th>
<th>IMI Standard Deviations:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Game:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unmasked</td>
<td>2*</td>
<td>5.0</td>
<td>0.3</td>
</tr>
<tr>
<td>Masked</td>
<td>5*</td>
<td>5.3</td>
<td>0.6</td>
</tr>
<tr>
<td>Basic</td>
<td>2</td>
<td>5.0</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Table 7 - Questionnaire, 7-Point Scale Questions Only (All questions were out of 7 = very true, except for Question 11, which was reverse-coded because of the negation in the wording, and Question 12, which was coded with 3 = Daily, 2 = 1-3x/week, 1 = Never, 0 = Unsure.)

<table>
<thead>
<tr>
<th>Question Wording</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) I felt like I was doing exercise.</td>
</tr>
<tr>
<td>2) This activity was fun to do.</td>
</tr>
<tr>
<td>3) I think doing this activity could help me to keep up with my exercises.</td>
</tr>
<tr>
<td>4) It was clear what to do next.</td>
</tr>
<tr>
<td>5) I would be willing to do this again because it has some value to me.</td>
</tr>
<tr>
<td>6) I think this is an important activity.</td>
</tr>
<tr>
<td>7) While I was doing this activity, I was thinking about how much I enjoyed it.</td>
</tr>
<tr>
<td>8) It was clear how well I was doing.</td>
</tr>
<tr>
<td>9) I believe doing this activity could be beneficial to me.</td>
</tr>
<tr>
<td>10) I felt I could do the tasks.</td>
</tr>
<tr>
<td>11) I thought this was a boring activity. (reverse-coded)</td>
</tr>
<tr>
<td>12) I would be willing to do this activity: Daily, 1-3x/week, Never, Unsure. (coded out of 3)</td>
</tr>
</tbody>
</table>

Figure 13 below contains the average scores for the IMI questions in the custom questionnaire (7 out of 24 questions, including open-ended questions and questions only for the final
session to compare all three games). Questions 2, 3, 5, 6, 7, 9, and 11 from the custom questionnaire were questions from the IMI. The IMI questions do not make up all of the scaled questions. For all of the scaled questions in the questionnaire, see Table 7 above.

![Intrinsic Motivation Inventory Scores](image)

**Figure 13 - IMI Sub-Scale Scores.** (Perceived Usefulness = questions 3, 5, 6, and 9; Enjoyment = questions 2, 7, and 11).

Table 8, “Comparing IMI Subscale Scores - by Participant”, below shows the average scoring each participant gave for each game, with breakdowns for Perceived Usefulness and Enjoyment, as well as the average IMI scores for each game (Basic, Masked, Unmasked). Table 9, “Comparing Explicit Game Preferences with Other Measures - by Participant”, takes the game each individual participant self-reported as their preferred game, and compares it with the game with the highest average IMI score from Table 8, as well as with other potential indicators of game preference for each participant respectively.

Looking at the Table 9, five out of the eight stroke survivors’ highest IMI ratings of the games did not match their explicit game preferences, three played for the longest time on games they did not report as their preferred game, and six performed the highest numbers of reps on games they did not select as their preferred game. At least six out of the eight participants’ performance did not match with their preferences. Participant 1 thought that different games among the three could be used at different points of recovery. Participant 3 had a minor cognitive impairment and minor vision problem, which may explain the inconsistency of their explicit game preference with their time on task and motion repetitions performed, while on the other hand their consistency with their preference inferred from self-reported IMI scores.
### Table 8 - Comparing IMI Subscale Scores - by Participant

<table>
<thead>
<tr>
<th>Participant:</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>#6</th>
<th>#7</th>
<th>#8</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Usefulness</td>
<td>6.8</td>
<td>6.0</td>
<td>3.5</td>
<td>7.0</td>
<td>7.0</td>
<td>4.0</td>
<td>4.8</td>
<td>3.7</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>4.7</td>
<td>4.3</td>
<td>3.3</td>
<td>7.0</td>
<td>7.0</td>
<td>2.7</td>
<td>1.0</td>
<td>4.3</td>
</tr>
<tr>
<td>Average</td>
<td>5.7</td>
<td>5.2</td>
<td>3.4</td>
<td>7.0</td>
<td>7.0</td>
<td>3.3</td>
<td>2.9</td>
<td>4.0</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.5</td>
<td>1.2</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.9</td>
<td>2.7</td>
<td>0.5</td>
</tr>
<tr>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Usefulness</td>
<td>6.5</td>
<td>6.8</td>
<td>3.0</td>
<td>7.0</td>
<td>7.0</td>
<td>6.0</td>
<td>5.7</td>
<td>2.7</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>4.0</td>
<td>4.0</td>
<td>2.7</td>
<td>6.0</td>
<td>7.0</td>
<td>5.7</td>
<td>3.7</td>
<td>2.0</td>
</tr>
<tr>
<td>Average</td>
<td>5.3</td>
<td>5.4</td>
<td>2.8</td>
<td>6.5</td>
<td>7.0</td>
<td>5.8</td>
<td>4.7</td>
<td>2.3</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.8</td>
<td>1.9</td>
<td>0.2</td>
<td>0.7</td>
<td>0.0</td>
<td>0.2</td>
<td>1.4</td>
<td>0.5</td>
</tr>
<tr>
<td>U</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Usefulness</td>
<td>6.8</td>
<td>6.3</td>
<td>1.5</td>
<td>6.9</td>
<td>7.0</td>
<td>5.0</td>
<td>6.0</td>
<td>2.7</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>5.0</td>
<td>4.3</td>
<td>3.3</td>
<td>6.7</td>
<td>7.0</td>
<td>4.0</td>
<td>4.0</td>
<td>2.7</td>
</tr>
<tr>
<td>Average</td>
<td>5.9</td>
<td>5.3</td>
<td>2.4</td>
<td>6.8</td>
<td>7.0</td>
<td>4.5</td>
<td>5.0</td>
<td>2.7</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.2</td>
<td>1.4</td>
<td>1.3</td>
<td>0.1</td>
<td>0.0</td>
<td>0.7</td>
<td>1.4</td>
<td>0.0</td>
</tr>
</tbody>
</table>

### Table 9 - Comparing Explicit Game Preference with Other Measures - by Participant

<table>
<thead>
<tr>
<th>Participant:</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>#6</th>
<th>#7</th>
<th>#8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explicit Game Preference</td>
<td>M</td>
<td>M</td>
<td>B</td>
<td>M</td>
<td>B</td>
<td>M</td>
<td>U/M</td>
<td>U</td>
</tr>
<tr>
<td>Highest IMI Score</td>
<td>U*</td>
<td>M</td>
<td>B</td>
<td>B</td>
<td>(all)</td>
<td>M</td>
<td>U</td>
<td>B</td>
</tr>
<tr>
<td>Game Preferred Matches Rating?</td>
<td>No*</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Highest Time on Task</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>B</td>
<td>M</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>Game Preferred Matches Time on Task?</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Highest Reps</td>
<td>M/U**</td>
<td>M</td>
<td>M</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>Game Preferred Matches Reps***?</td>
<td>No**</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>------</td>
<td>-----</td>
<td>----</td>
<td>----</td>
<td>-----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
</tbody>
</table>

IMI = Intrinsic Motivation Inventory; B = Basic; M = Masked; U = Unmasked.

* = ("Almost" matches. Participant #1: 4.9, 5.0, 5.1, with average scores' standard deviation equal to 0.2)

** = ("Almost" matches, because the highest number of reps is a tie between the Unmasked and Masked games.)

*** = (Repetitions. Had to be counted visually for the Basic game.)

5.4.1. Preferences - Non-IMI Question Scores

Figure 14 - Question 1 from the questionnaire

Figure 15 - Question 4 from the questionnaire
Figure 16 - Question 8 from the questionnaire

Question 8: “It was clear how well I was doing.”

Based on guidelines and ideas from the advisory panel.

Figure 17 - Question 10 from the questionnaire. Based on guidelines and ideas from the advisory panel.

Question 10: “I felt I could do the tasks.”

Figure 18 - Question 12 from the questionnaire. Based on (McKay and Maki, 2010). These choices of response were not presented to participants as numbers on a scale, but for the sake of plotting, 3 = “Daily”; 2 = “1-3x/week”; 1 = “Never”; 0 = “Unsure”.

Question 12: “I would be willing to do this activity: Daily, 1-3x/week, Never, Unsure”
5.5. Additional Data - Statistical Analysis

There were three additional statistical analyses performed due to noticeable differences in some measurements: 3-way t-tests, Cohen’s d effect sizes, and ANOVA (Analysis Of Variance). The multiple t-tests were used to determine whether there were statistically significant differences for each measure, and to determine which game had the higher comparative value. The Cohen’s d effect sizes were used to determine whether the statistically significant differences were of significant effect sizes. ANOVA was performed as an extra confirmation of the existence of statistical significance found in the multiple t-tests. The t-test is not as conservative as ANOVA, however, ANOVA cannot tell you which of the three conditions has the significantly different value, but the t-test can, and can also tell you which one is a higher value.

The following table shows the statistically significant differences in several of the measures in this study, based on a t-test. Statistical significance was found in seven instances for Masked versus Basic, and four instances for Unmasked versus Basic, and two instances when comparing Masked and Unmasked games; namely, time on task, the number of reps, the number of reps per minute, the number of misses, the percent of hits, and a couple of the percentages of ROM usage. Note that, as expected of different measures, the number of reps had statistically significant differences, while the number of hits (successful reps) did not. It should also be noted that for the questionnaire custom questions and IMI questions obtaining “implicit” preferences, no statistically significant differences were found between games, and that more than half of the objective measurements (such as number of reps) had statistically significant differences.

Table 10 - T-Tests for Significant Difference Between Games: * = Statistically significant difference. Removing one tenth of Misses for the Masked game did not affect statistical significance.

<table>
<thead>
<tr>
<th>2-tail t-test p-values</th>
<th>Masked &gt; Basic</th>
<th>Masked &gt; Unmasked</th>
<th>Unmasked &gt; Basic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time on task:</td>
<td>0.01711*</td>
<td>0.10716</td>
<td>0.46022</td>
</tr>
<tr>
<td>Reps (#):</td>
<td>0.01448*</td>
<td>0.23322</td>
<td>0.00848*</td>
</tr>
<tr>
<td>Reps/Min:</td>
<td>0.00003*</td>
<td>0.49302</td>
<td>0.00002*</td>
</tr>
<tr>
<td>Hits/min:</td>
<td>0.30720</td>
<td>0.88550</td>
<td>0.40938</td>
</tr>
</tbody>
</table>

Removing one tenth of Misses for the Masked game did not affect statistical significance.
The following table shows the Cohen’s d effect sizes for the measures that had statistical significance. Of those, the ones with medium or large effect sizes were the time on task, number of reps, reps/min, number of misses, and the two measures of ROM usage (“pull” and “right”).

**Table 11 - Tests for Effect Size of Differences Between Games:** * = tentatively medium effect size (greater than 0.5 but less than 0.8); ** = tentatively large effect size (greater than 0.8); *** = tentatively small effect size (between 0.2 and 0.5) after adjusting for 1.6 more repetitions and hits for Masked game, but would have been medium effect size at a value of 0.61692 (between 0.5 and 0.8).

<table>
<thead>
<tr>
<th></th>
<th>Masked &gt; Basic</th>
<th>Masked &gt; Unmasked</th>
<th>Unmasked &gt; Basic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time on task:</strong></td>
<td>1.54584**</td>
<td>1.19870**</td>
<td>0.42880</td>
</tr>
<tr>
<td><strong>Reps (#):</strong></td>
<td>0.62119*</td>
<td>-0.09003***</td>
<td>1.53014**</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Masked &gt; Basic</th>
<th>Masked &gt; Unmasked</th>
<th>Unmasked &gt; Basic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hits (#):</td>
<td>0.64428</td>
<td>0.62049</td>
<td>0.94100</td>
</tr>
<tr>
<td>Misses (#):</td>
<td>0.04765*</td>
<td>0.04268*</td>
<td>0.27193</td>
</tr>
<tr>
<td>Hits (%):</td>
<td>0.01724*</td>
<td>0.01677*</td>
<td>0.84725</td>
</tr>
<tr>
<td>PULL (ROM %):</td>
<td>0.00902*</td>
<td>0.48606</td>
<td>0.00057*</td>
</tr>
<tr>
<td>PUSH (ROM %):</td>
<td>0.89258</td>
<td>0.36840</td>
<td>0.22769</td>
</tr>
<tr>
<td>LEFT (ROM %):</td>
<td>0.46099</td>
<td>0.75671</td>
<td>0.58695</td>
</tr>
<tr>
<td>RIGHT (ROM %):</td>
<td>0.02863*</td>
<td>0.30823</td>
<td>0.00318*</td>
</tr>
<tr>
<td>IMI (Perceived Usefulness)</td>
<td>0.77497</td>
<td>0.74625</td>
<td>0.93295</td>
</tr>
<tr>
<td>IMI (Enjoyment)</td>
<td>0.93080</td>
<td>0.76251</td>
<td>0.71715</td>
</tr>
<tr>
<td>Q 1</td>
<td>0.73752</td>
<td>0.95544</td>
<td>0.77964</td>
</tr>
<tr>
<td>Q 4</td>
<td>0.77103</td>
<td>0.73235</td>
<td>0.48132</td>
</tr>
<tr>
<td>Q 8</td>
<td>0.40438</td>
<td>0.48765</td>
<td>0.92301</td>
</tr>
<tr>
<td>Q 10</td>
<td>0.43835</td>
<td>1.00000</td>
<td>0.43835</td>
</tr>
<tr>
<td>Q 12</td>
<td>0.14890</td>
<td>0.51763</td>
<td>0.31896</td>
</tr>
</tbody>
</table>
### Cohen’s d Effect Sizes:

<table>
<thead>
<tr>
<th></th>
<th>Masked &gt; Basic</th>
<th>Masked &gt; Unmasked</th>
<th>Unmasked &gt; Basic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reps/Min:</td>
<td>0.57523*</td>
<td>-1.85794</td>
<td>3.16402**</td>
</tr>
<tr>
<td>Misses (#):</td>
<td>1.08230**</td>
<td>1.11135**</td>
<td>-0.57184</td>
</tr>
<tr>
<td>Hits (%):</td>
<td>-4.03105</td>
<td>-4.02194</td>
<td>0.09811</td>
</tr>
<tr>
<td>PULL (ROM %):</td>
<td>1.49325**</td>
<td>0.35528</td>
<td>2.21709**</td>
</tr>
<tr>
<td>RIGHT (ROM %):</td>
<td>1.19820**</td>
<td>0.52841*</td>
<td>1.47614**</td>
</tr>
</tbody>
</table>

The following table shows the ANOVA F-test values. The statistically significant values were found for all the same measures as in the t-tests, except for the one for time on task. This at least confirms that statistically significant differences should be found in these six measures: Reps (#), Reps/Min (#/m), Misses (#), Hits (%), PULL (%), and RIGHT (%).

**Table 12 - ANOVA Tests for Significant Difference Between Games:** critical F-value = 3.47

<table>
<thead>
<tr>
<th>ANOVA F-Test:</th>
<th>F-value:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time on task:</td>
<td>1.09</td>
</tr>
<tr>
<td>Reps (#):</td>
<td>4.82</td>
</tr>
<tr>
<td>Reps/Min (#/m):</td>
<td>21.61</td>
</tr>
<tr>
<td>Hits/Min (#/m):</td>
<td>0.57</td>
</tr>
<tr>
<td>Hits (#):</td>
<td>0.17</td>
</tr>
<tr>
<td>Misses (#):</td>
<td>4.81</td>
</tr>
<tr>
<td>Hits (%):</td>
<td>7.18</td>
</tr>
<tr>
<td>PULL (%):</td>
<td>6.37</td>
</tr>
<tr>
<td>PUSH (%):</td>
<td>0.56</td>
</tr>
<tr>
<td>LEFT (%):</td>
<td>0.37</td>
</tr>
<tr>
<td>RIGHT (%):</td>
<td>4.27</td>
</tr>
<tr>
<td>IMI (Perceived Usefulness):</td>
<td>0.07</td>
</tr>
<tr>
<td>IMI (Enjoyment):</td>
<td>0.08</td>
</tr>
<tr>
<td>Q 1:</td>
<td>0.06</td>
</tr>
</tbody>
</table>
5.6. Observation Logs – Verbal Feedback & Observations:

The researcher attempted to record all comments and verbal responses made by participants. There are also notes on observed reactions, such as verbal cues as to effort and concentration (like verbalization, or lack thereof during gameplay), and whether the participants appeared to mis-/not understand certain aspects of the games, as well as notes on observed physical reactions, such as trajectories deviating from the expected trajectories for the waypoints, or noticeable physical exertion.

While the time constraints for this mixed-methods exploratory study did not allow for a full qualitative analysis or thematic analysis of the comments and observations (and questionnaire questions), there were specific comments worth including here that were mainly related to future game design recommendations:

- Participants would positively react to images/animations and pay attention to text instructions, sometimes reading them out loud.

- Some participants did not notice the timer in the Unmasked game, which may have affected their perception of variation/novelty, and hence motivation to continue with the game.

- Some participants needed help with the correct directions of movement during ROM calibration.

- Participants in this study needed to be verbally prompted or reminded to do so during “continue? yes/no” screens between levels, to avoid fatigue or over-exertion.
• Sometimes the researcher noticed that participants had paused moving after transitions between animations, calibration, or gameplay, and had to verbally prompt participants to play.

• Almost all participants would want to finish the Masked and Unmasked games up to level 11 (because of the text saying how many levels out of 11 they had completed), and then some would stop, sometimes because of and sometimes despite knowing that the game kept going infinitely past level 11.

• One participant with apparent cognitive deficits had difficulty with the Unmasked game used in this study because they found the game did not meet their expectations of realistic cooking (in terms of ingredients, the steps, how the pot progress bar was filling slower in higher levels when more ingredients are added), or reminded them of real-life negative experiences related to cooking (fear of spilling), which they reported affected their gameplay and they reported should be dealt with.

There were also some recommendations for future game designs that were provided by stroke survivor participants, based on the questions in the questionnaire asking for suggestions. Sometimes suggestions were given during gameplay as participants thought of them, and sometimes even when the researcher did not prompt for suggestions.

• Participants cited including a growing variety of ingredients, increasingly faster speeds of alien enemies, and being able to fight back against enemies, in order to add novelty, variation, and challenge.

• One participant suggested use of all three games, but at different stages of recovery based on complexity of visuals and increasing levels of challenge in both ROM and speed/timing.

• A couple of participants expressed that they wanted to see or know their own previous scores or the scores of other stroke survivor participants.

• One stroke survivor participant explicitly expressed a desire for competition.
5.7. Questionnaire - Reasons for Preferences:

Most stroke survivor participants selected the Masked game. Some other participants preferred the Unmasked or Basic game, and one self-reported a tie in preference for the Masked and Unmasked games because they saw no difference between the two.

For stroke survivors who preferred the Masked game, reasons they gave related to the game design itself were: “it’s the theme—extraterrestrial—outer space”, that they played faster because of the aliens, that the game had “more interactive than the others”, and “space ship [sic: aliens] did the block so it added more challenge”.

For the stroke survivor who only preferred the Unmasked game, the reason they gave related to the game design itself was: “because I can relate to it”. One participant mentioned that the Unmasked game was a close second and said that it had more graphics and a timer, but needed a “block” (like the “aliens” in the Masked game), and so having both a timer and “enemies” would be best.

For stroke survivors who preferred the Basic game, reasons they gave related to the game design itself were that it was “easy to see” and that you can actually challenge yourself by trying to move in the path it tells you.

When asked about how the Masked and Unmasked games could be improved, participants cited including a growing variety of ingredients, increasingly faster speeds of alien enemies, and being able to fight back against enemies.

Some questions did not ask for reasons explaining responses or ratings, but some participants gave reasons anyway, unprompted. For a list of all the questions from the custom questionnaire, see Appendix E - “Custom Questionnaire for Stroke Survivor Participants”.

5.7.1. Questionnaire - Other Open-Ended Questions

Besides the questions on the preferred game questions asking for explanations for responses, questions number 13, 14, 15, and 17 were the arranged open-ended questions in the questionnaire.
For question number 13, participants were asked to give their most favourite parts of the game they had just played in that session. For the Basic game, answers ranged from no favourite part, to simply the feeling of completing tasks, to challenging oneself, but not about the game itself per se. For the Masked game, answers ranged from completion/winning, to the challenge brought by the alien spaceships, to no favourite part, or “all of [it]”. For the Unmasked game, answers ranged from no favourite part, to completion/scoring, the challenge to beat the clock, seeing what kinds of ingredients, or being able to relate to the game.

For question number 14, participants were asked to give their least favourite parts of the game they had just played in that session. For the Basic game, answers ranged from “the rewards” (or lack thereof), to frustration with one’s own movements, lack of gamification, feelings of repetitiveness, or no least favourite part. For the Masked game, answers ranged from none, to anticipation of the next level due to “waiting while it says ready, set, go”, to “in the beginning, when I didn’t have something to ignore—as soon as the aliens starting coming then it made it a game.” For the Unmasked game, answers ranged from components related to waypoint placement, to not being able to relate to the green ingredient, to thinking that the pot may be overflowing, to expressing impatience with “waiting while it says ready, set, go”. One participant with apparent cognitive deficits expressed during gameplay and again self-reported during the questionnaire that they had difficulty with the Unmasked game because they found it broke expectations and reminded of negative memories.

For question number 15, participants were asked to give game features they would like to see added to the game they had just played in that session. For the Basic game, answers ranged from real-time in-game progress charts, to characters or avatars, more exciting sound effects for hits, random waypoint placement, or waypoints placed more often at trouble areas (but would not make it much less boring). For the Masked game, answers ranged from better sound effects that are different for hits and misses, comparative scores, targeting trouble areas, “more cartoon characters” (“good guys” and “bad guys”), faster alien spaceships, or checking for pain. For the Unmasked game, answers ranged from more different ingredients (squash, leafy greens, boiling water), to consulting a chef, to indicating current level and time spent playing.
6. Discussion of Results

6.1. Masked or Unmasked?

If we set aside the question of the “equivalency” between the “Masked” and “Unmasked” games in the first place, ignore the “approximately 1.6 times higher difference”, and only look at the average numbers for the results for each of the three research questions, the “Masked” game outperformed the “Unmasked” and “Basic” games on most of the categories of measures used in this study. Looking individually at each specific research question:

1) Which game design has users wanting to keep playing for the longest period of time?
   The Masked game (but the difference was not statistically significant). By itself, time on task may not be a strong indicator, but combined with the other research questions and qualitative results, a fuller picture may be put together.

2) Which game design contributes to the greatest number, rate, and size (i.e. Range of Motion used in workspace) of motion repetitions? When stroke survivors played either the Unmasked or Masked games, they performed the most, fastest, and largest reps. However, for the Masked game there were a higher number of misses but also of reps. The number of repetitions for all three games in this study met the recommended minimum of 100 reps daily (Kimberley et al., 2010), or a minimum of 100 x 2 = 200 “hits” daily.

3) Which game design has the most favourable stroke survivor self-reported preference scores (perceived usefulness, enjoyment, explicit indication of game preference)? In the explicit choice of game preference, the Masked game got the most “votes” (but its score in the IMI did not have a statistically significant difference).

Based on these three answers, the Masked game out of the three games as used in this study seemed to best fulfill the criteria of highest average time on task, repetitions, and preference. The results found in this study favouring the Masked game may be in part due to the “Masked” aspect of its game design, but also due to other factors.

Upon inspection of patterns in the demographics of the participants, the Basic game was preferred only by the two females in this study, who also happened to have the highest CMSA...
(arm) levels of 4 and 5 (while the other participants were male and had CMSA level 3). However, of these two individuals who reported preferring the Basic game at the end of their sessions, one seemed to dislike the Basic game on the day of their gameplay session and asked to stop at 4 minutes of gameplay. The Unmasked game was preferred by the oldest participant. In both cases, with Basic or Unmasked games being reported as preferred, there were either seemed to be or were self-reported cognitive deficits, however this was also the case in at least one participant for the Masked game too. Although these initial results can be interpreted to mean that cognitive deficits do not guarantee self-reported preference favouring a certain game type, this study only had eight participants, and so this result would need to be further investigated to find any real correlation.

The results for the overall performance and preferences by the stroke survivors as a population visually seem to favour the Masked game over the Unmasked and Basic games, although the Masked game had high standard deviations that overlapped with the standard deviations of the other games. This is suggested by the corresponding greater average values in time on task, number of reps, reps per minute, number of hits, number of misses (even with one tenth removed), and to a lesser extent hits per minute and IMI scores for the Masked game. The rep sizes by percent of active ROM were on average all higher for the Masked game. The number of stroke survivors who preferred the Masked game over the other two games was also the highest. Not so for the percent of hits, which was actually lowest for the Masked game (and also further demonstrates the difference between the respective counts and rates of hits and motion repetitions). However, note that these results favouring the Masked game are limited because, for most measurements, the standard deviation was also the highest for the Masked game design. This was shown with the large standard deviations in the IMI subscale scores for perceived usefulness and enjoyment, with the means being within 0.2 to 0.6, which was smaller than the integer intervals for the 7-Likert scale used.

Based on the statistically significant differences in the number of reps, the number of reps per minute, the percent of hits, the number of misses, and a couple of the percentages of ROM usage, the Masked and Unmasked games would seem to be both more beneficial than the Basic game. However, there were no statistically significant differences between the Masked and Unmasked games for most measures, except that the Unmasked game had a higher ratio of hits to misses and fewer misses (higher success rate). This would seem to favour the Unmasked game
because this could mean that it made users perform higher-quality motions. However, this may not matter if we simply want to have stroke survivors performing higher-quantity (i.e. more) reps in the first place, before getting to higher-quality (“better”) reps. Either way, these interpretations may be limited by the effects of original “equivalency” between these two games discussed in another section. If the table of significant differences is taken at face value, it is recommended that future studies measure both quantitative and qualitative measures, because user preferences between options of tool interfaces do not always match task performance with those tools, known as the “paradox of preference versus performance”, where users sometimes prefer interfaces they unknowingly perform tasks worse with (Morita and Cafazzo, 2016; Bailey, 1993), and this can even apply to decisions made by system designers (Bailey, 1993). Furthermore, none of the “implicit” (and qualitative) measures of preference had statistically significant differences in this study, while some of the objective (and quantitative) measures of performance did. Simply presenting novel but failure-inducing interfaces can drive preference over real performance (Morita and Cafazzo, 2016), so system designers need to take care when testing and comparing options of how to present activities. One way to make preference and performance align better may be to simply implement the system with better performance but explain to users the favourable aspects of the lower-preference option so that their preferences may change (Andre & Wickens, 1995). The system itself could be designed to make it clear to users as to the performance effects of the chosen design (Andre & Wickens, 1995). However, all of this does not mean that stroke survivors’ self-reported preferences should be ignored (Andre & Wickens, 1995), but that their opinions should be taken with caution; the first priority for our purposes is to have stroke survivors willing to participate in the activity in the first place, before finding a way to improve performance. In the end, we want to optimize both preference and performance (Andre & Wickens, 1995). In any case, studies should include corresponding measures for both quantitative and qualitative data in terms of preference and performance, and bias should be mitigated with methods as suggested in human factors literature (Morita and Cafazzo, 2016).

A recent study was similar to the present study in that it compared “coaching” with “gamified” modes of games, and found that game mechanics strongly influence user behaviour and activity levels (Cameirão et al., 2016). It also found that the “coaching” mode (corresponding to this study’s Basic game) generated higher activity levels than the “gamified” mode (corre-
sponding to the Unmasked game), but that the “gamified” mode generated more efficient movement (Cameirão et al., 2016). In comparison, the numerical data from all eight participants in the present study suggests that both the Masked and Unmasked games resulted in higher activity levels (and the Masked game actually produced much less efficient motions with its many more misses). As well, it is still up for debate whether less repetitions of better quality are more efficient than many repetitions (Cameirão et al., 2016). Cameirão et al. concluded that the “coaching” mode should be used for acute care and many repetitions of fine motor skills, while the “gamified” mode should be used for long-term home-based fitness and maintenance (Cameirão et al., 2016). This is similar to how one participant in the present study explicitly said they saw value in all three games and that it should be used at different recovery stages.

A previous study looking into games that could fall under Masked and Unmasked games had some similar results as the present study when it came to the accuracy of movements. Although not framed as a comparison between Masked and Unmasked games, that study had a few games that could be characterized as being more like “Masked” games (such as the Sponge and Twirl games), which were found to have on average longer durations of single movements, longer movement path lengths, and higher movement speeds than the more “Unmasked” games in that study (Steinisch et al., 2012). The results were consistent with Fitts’ law, a model of human movement often applied to human-computer interaction (Steinisch et al., 2012) that shows a roughly positive (though logarithmic) relationship between travel distance and the time it takes to move a pointer to a target (M. Goktürk, n.d.; Fitts & Peterson, 1964). The two above interpretations, namely that “Masked” games can induce faster and larger movements, and that such movements can come at the cost of accuracy (and taking longer to reach targets), are consistent with the present study’s findings that the “Masked” game induced on average faster and larger movement repetitions but also higher numbers of misses than in the “Unmasked” game.

Looking at the comments explicitly provided by stroke survivors themselves, it seems that despite both Masked and Unmasked games having different forms of a timing mechanism (Unmasked: a literal timer countdown before items disappear; Masked: the speed of aliens stealing items), stroke survivors that preferred the Masked game perceived it to the more interactive, or even personified the alien spaceships to be “fighting” the player. However, this may be due to the addition of non-player characters during gameplay, or due to the irregular and faster speed of the aliens in the Masked game when compared to the timer in the Unmasked game. On the other
hand, it could just be the fact that the game was “Masked” that made it have the most “votes” of preference over the other games. One stroke survivor literally said that their reason for preferring the Masked game was “the theme—extraterrestrial—outer space”. The implication this could have for therapy is that Masked games are supported by stroke survivors, so Masked games should continue to be developed and considered for use in practice to help stroke survivors get that initial motivation to follow-through with therapy sessions, at least potentially for a majority of stroke survivors. However, even though the Masked game had more votes than the Unmasked and Basic games individually, together, almost half of the participants preferred one of either the Unmasked or the Basic game over the Masked game. If you combine that fact with the high standard deviations for the Masked game, as well as the adjustments to the data, then the data suggests that games should be chosen to match the individual preferences of stroke survivors. If we combine this with the previously-mentioned idea of informing users in order to encourage preference and performance to align better, then maybe future research should obtain better statistical significance for performance data to then literally incorporate this data into later game designs informing users of performance effects (for example “data shows this game helps people do more motions” for the Masked game or “data shows this game helps people do higher-quality motions” for the Unmasked game) and then seeing the effects on preference and performance. However, such a study and its ethics must be thought out carefully to avoid giving misleading permanent impressions. If Masked games significantly improve performance compared to the Unmasked game, then future designs need to be designed to better mimic a variety of movements for real-life activities and be tested for transfer to real-life activities. If it turns out that instead the Unmasked game does better, then future designs need to be designed to make sure that motivation to continue is sustained in the long-term. Either way, informing both stroke survivors and therapists of the performance data may help with aligning preference with performance. Alternatively, future games could even avoid the whole issue altogether and blend both the Masked and Unmasked games together, with activities from both being incorporated into a more elaborate storyline. For example, framing “Masked” activities as “breaks” between “Unmasked” activities (maybe as rewards), or subtly incorporating “Unmasked” activities as preparation for a larger “Masked” story, such as making a sandwich before an alien attack on your picnic table. A game of this kind would then also naturally incorporate variation and novelty to maintain interest, which is also one of the suggestions for future work.
6.2. Recommendations and Lessons Learned for Future Research

The following are nine game design recommendations for future research, based on things that stood out from observations or comments that were reported to affect gameplay or could have affected gameplay:

The first game design recommendation is based on the measurements made in this study:

1) **Consider using and improving on the Masked game as a starting point.** As-is, the game designated as “Masked” in this study was the game preferred by the most stroke survivors, and had stroke survivors playing for the longest period of time and performing a higher rate of motion repetitions. The first priority is to have stroke survivors motivated to participate in activities in the first place and also performing the most motion reps possible, before improving the quality of performance of such movements and their generalizability. Once better statistical significance and effect sizes in performance parameters are found for whichever game design, this data should be shown or incorporated into the game design itself to help preferences align with actual performance metrics.

The next five game design recommendations are based on at least stroke survivor comments:

2) **Different games for different levels of recovery, or blend game types.** One participant suggested use of all three games, but at different stages of recovery based on complexity of visuals and increasing levels of challenge in both ROM and speed/timing. Possibly a blend of game times incorporating one within the other could be carefully designed, as discussed earlier.

3) **Build in explicit prompts for taking breaks,** as the researcher noticed that participants in this study sometimes needed to be verbally prompted or reminded to take breaks during “continue? yes/no” screens between levels, to avoid fatigue or over-exertion. One stroke survivor noted that prompts for breaks could have been built into the games.

4) **Show levels completed out of a total number, but afterwards reveal more levels** upon completion, as almost all participants would want to finish the Masked and Unmasked games up to level 11 (because of the text saying how many levels out of 11 they had completed), and then some would stop, sometimes because of and sometimes despite
knowing that the game kept going infinitely past level 11. However, this also needs to be complemented with non-repeating levels, and based on comments, this should be done by also slowly adding novelty, variation, and challenge. For example, participants cited including a growing variety of ingredients, increasingly faster speeds of alien enemies, and being able to fight back against enemies. The Masked and Unmasked games may have tended to perform better than the Basic game because the higher levels in those games slowly introduced variation (“you win” screens breaking up periods of gameplay, a slow curve of increasing number of targets per level) and extra features (different items to pick up, then a mechanism to add time pressure for reaching targets). This recommendation is similar to findings that continually offering challenges induces learning (Camerão et al., 2016).

5) Avoid violating expectations or inducing bad memories. A realistic “Unmasked” game should have carefully-designed game mechanics so expectations of how to conduct the ADL are not violated, and so that reminders of negative real-life experiences of task failure are avoided. One participant with apparent cognitive deficits had difficulty with the Unmasked game used in this study because they found it broke some expectations of realistic cooking, or reminded them of real-life negative experiences related to cooking, which they self-reported affected their gameplay. Bad memories can be induced by different things for different people, and specific contexts may be unforeseen, but can be avoided through therapist consultation with individual stroke survivors and past experience.

6) Future research and development should decide whether to include personal progress and/or other participants’ scores, as well as competition (or cooperation) preferably with physically present people. A few stroke survivors in this study indicated a desire for these features. However, caution and customization should be taken into consideration, because most individual stroke survivors may prefer only one of either competition (between human players) or cooperation (between human players against the computer), and dislike the other option (Novak et al., 2014). More research is needed into 2-player games with stroke survivors at different levels of recovery and with different kinds of impairments (Rego et al., 2010). If this multiplayer route is taken, having the other player physically in the same room is better than over an online connection (Novak et al., 2014).
This recommendation is related to the other recommendation of a more “holistic” view of the robotic game system, such as incorporating a social element in games.

The last three game design recommendations are only based on researcher observations:

7) **Combine voice-over with the animations and text instructions** and do not just use text and/or images alone, especially to account for vision deficits. Text and images are clearly helpful as participants would positively react to images/animations and pay attention to text instructions (which were intentionally kept laconic for usability), sometimes reading them aloud. Adding animations or voice-over could have made the timer more noticeable in the Unmasked game (some participants did not notice it). Sometimes the researcher had to verbally prompt participants to play after transitions between animations, calibration, or gameplay.

8) Animations should ideally show **real-life video clips of someone demonstrating movements during ROM calibration** to mitigate any ambiguity in text instructions (some participants needed help with the correct direction to move the end-effector with their arm), especially since text instructions need to be kept short for user-friendliness.

9) **Keep the single-coloured trapezoid backgrounds for simple 3D perspective**, which may have helped, as this study (albeit with small sample size) did not have issues heard of in a concurrent study where stroke survivors with cognitive deficits had trouble correctly mapping visuals on the screen (physically displayed vertically) to the plane of the table (physically horizontal).

The following are three more recommendations for future game development processes, based on lessons learned from conducting this study that can be done better in the future:

1) **Check the original “equivalency” between games to isolate variables** better before testing on actual study participants. This is the only recommendation that comes from a game design flaw that could have biased the data.

2) Game preference and motivation may have been biased or affected for the better, but either way **a more “holistic” view of the robotic game system** is recommended. Before participating in this study, participants had differing levels of experience with other
games and activities, including different kinds of rehabilitation therapies, which may have introduced preconceptions but also potential experience with a range of alternatives they have tried for what constitutes “good” games. It could be assumed that participants should have a better idea of which alternative activities work for them, in comparison to other stroke survivors who have had no experience with other alternatives. A more “holistic” view may need be needed for future game designs, such as incorporating a social element with real people as fellow participants in games, such as other stroke survivors or family members, which seemed to be a strong motivator for at least one participant. Recovery and motivation towards therapy cannot be tied solely to the computer games used in any single study, because participants do a variety of activities related to recovery, and target more than just the arm or even physical side of recovery. Hence, a more “holistic” view is recommended.

3) During the initial prototyping phase of this study, there were delays in getting designs finished and tested. However, recently some scholarly and non-scholarly work has been published based on best practices to speed up the process of testing ideas. In order to both rapidly prototype and test designs for future designs, one could consider a “Design Sprint” methodology of best practices, which includes a series of 5-day structured “sprints” for rapid testing to obtain feedback early and adjust designs early (Knapp et al., 2016; GV, 2016; Google Developers, 2015). Design Sprints have been used in a recent student workshop (Wichrowski et al., 2015) and a recent masters thesis (Grahl and Söderling, 2015). One thing that could have improved the initial advisory committee was the Design Sprint’s suggested use of individual thinking time before group discussion (GV, 2016), as opposed to using brainstorming, which can be subject to the effects of group-think if not carefully prepared (Brown et al., 1998).
7. Limitations

7.1. Sessions and Follow-Up

This was a short-term study, with only one session with each game for each stroke survivor parity (i.e., three sessions per participant). A longer study without follow-up over a longer period of time could gain better insight into the long-term motivation as well as actual adherence to the difference games.

Some measurements of time on task may have been biased due to factors external to this study. For example, some stroke survivors may have felt an urgency to leave early to ensure a ride home after certain sessions, depending on their schedules. Other such uncontrolled variables that the researcher did not consider may of course also factor into the observed measurements.

7.2. Statistical Significance

Despite the statistical significance found in some measurements, these results should be interpreted with caution as there was a small sample size and high population homogeneity that limit generalizability, as well as the possible issue with the original “equivalency” between the Masked and Unmasked games, as discussed below. For some of the measures, no significant differences were found between games.

7.3. Equivalency Testing between “Masked” and “Unmasked” Games

Unfortunately, based on additional data reviewed after the fact, there may have been a lack in “equivalency” between the “Masked” and “Unmasked” games in the first place and that may have affected the data collected in this study, but this was at least accounted for by a couple rough adjustments to the data. During the short “equivalency” pre-testing sessions with volunteers, the following measures also happened to be recorded: time on task, motion repetitions, score, and percent of active ROM. For those volunteer pre-tests, the average number of motion repetitions and scores were about 1.6 times higher for the “Masked” game than for the “Unmasked” game. This was not investigated until after another thing was noticed: during stroke survivor sessions, participants playing the “Masked” game performed much more misses than the other two games. The “Masked” game may have been already predisposed to have a bias in that it already encouraged users to do more motion repetitions and get higher scores, as seen earlier
recorded in the volunteer “equivalency” checks. In early tests with the “equivalency” volunteers, this 1.6 times higher difference in performance may have been because of the effect of the “Masking” itself, however during tests with stroke survivors it was observed repeatedly that the “aliens” in the "Masked" game were not consistent in the amount of time they took to “steal” the “fuel” (this was meant to take the same 10-second period that “ingredients” would take to disappear when the “timer” timed-out in the “Unmasked” game), and would consistently happen at one out of the ten set waypoints (the one nearest to the player's body, which is nearest to where the “aliens” would be returning to home positions in later levels). The implication this has is that the data can be “cleaned” by removing one tenth of the misses to attempt to correct for the number of “false misses” of targets due to the “consistently inconsistent” speeds of the “aliens”. The number of motion repetitions and scores (“hits”) can also be divided by 1.6 (as well as other data partially calculated with these values). Although doing this does not fully account for the lack of original “equivalency”, the different measurement results stayed mostly unchanged or were made slightly weaker. The biggest change would have been that the answer for the specific research question #2 in terms of repetitions would have a slightly weaker case (less repetitions, but still faster and larger repetitions) for the Masked game. This would not change the answers for the other two specific research questions (time on task and preference), which would still favour the Masked game.

Noting how motion reps and scores (hits) were around 1.6 times higher for the Masked game during initial design “equivalency” pre-testing with volunteers, it would be expected that the Masked game would already have an advantage and would already not be “equivalent” to the Unmasked game in the first place, and so then performance measures (time on task and repetitions) and preference measures (IMI, explicit preference, and IMI subscales) would also be higher. Although this could be due to the game being more engaging because it is Masked, it could also be more engaging because of the unfair advantage the Masked game had with the faster and variable time for its “aliens” to steal targets from the player. In other words, the Masked game could have been objectively and/or subjectively more difficult to begin with, and hence more engaging, and hence skew results. However, this approximately 1.6 times higher difference itself may not be reliable as the number of reps and hits could have been biased by the informality and lack of explicit time rules given to volunteers, and the fact that such numbers
were recorded as an afterthought and not set out as the basis to test for the subjective “equivalency”. Even if we consider the measurements, the number of reps per minute for the Unmasked game were actually higher for each healthy volunteer during the equivalency testing. This because volunteers spent less time in the Unmasked game, even though the game had a lower absolute value for the number of reps (not the reps per minute). Furthermore, how “equivalent” the Masked and Unmasked games are may still be perceived differently by stroke survivors than our initial volunteers who were not stroke survivors. One stroke survivor said that they thought the two games were exactly the same but just with different visuals.

Combining the overall preference with the general though limited favouring of the Masked game based on performance (time on task and repetitions), the game labelled in this study as a “Masked” game design is the one most preferred and helpful in motivating stroke survivors to keep playing longer and perform more, larger reps, even if they also miss more targets. These results may be due to one of or a combination of the following: the “Masking” in the design, the variable time constraint, or the mere presence of alien graphics being anthropomorphized in the users’ mind as enemies. Determining which is outside the scope of this study and is a limitation that can be addressed in future iterations of gaming studies, although comments do seem to lean towards the presence of the aliens but also the extraterrestrial theme of the game (“Masked”). Either way, as would be suggested by human factors research, sometimes users prefer designs of systems with which they perform less optimally, in what is called the “paradox of preference versus performance” (Morita and Cafazzo, 2016). This is supported by the fact that at least six out of the eight individual stroke survivors in this study gave explicit preference for a game that did not match the game for which they gave the highest average IMI ratings, spent the longest time playing, or performed the most repetitions (see Table 9 - “Comparing Explicit Game Preference with Other Measures - by Participant”). However, overall, there are much more reps being performed in the Masked game, upon visual inspection of the data plots. Now, whether these reps are of therapeutic sizes can be indirectly captured by the number of hits (successfully hit targets), as well as the percent of active ROM used. By visual inspection, for the Masked game, the average number of hits is slightly higher, and the average percents of active ROM usage are all higher, with the Masked game being the only one out of the three games to
have some stroke survivors reaching outside their active ROM in the “push” and “right” directions. This suggests that, although stroke survivors make more misses, they also perform more successful target hits and use more of their active ROM when playing the Masked game.

The interpretation of the rep sizes (via the values for ROM %) may be limited by the fact that compensation was not accounted for (this study prioritizes preference over performance). There is also the possibility that the perceived mechanics of the Masked game may have translated into slightly different motion paths during motion repetitions (because of the possibility of the assumption that aliens can steal from your hand), which could have biased rep sizes and number of reps, because users may think they need to avoid “touching” the aliens in the Masked game. Even though one stroke survivor described the aliens as a kind of “block”, the aliens can only steal the fuel if the user has not already “picked up" the fuel. Another difference between the games was that the Basic game had dotted guide lines that could be interpreted as the “best path” going directly from the previous player position to the current target, while the Masked and Unmasked games did not have such guide lines from player to target.

Another consideration is the “Zeigarnik effect”, in which it was observed that both adults and children may be more likely to remember more items from an incomplete task than from a completed task (Zeigarnik, 1938). This effect may explain this study’s observations of more time on task for the Masked game (due to game feedback indication incomplete target hits), but also of stroke survivors wanting to complete 11 out of 11 levels (even or especially when told that the last level just keeps repeating infinitely) for both the Masked and Unmasked games (due to game feedback indicating progress along those 11 “base” levels). This may also partly explain the higher measurements for the Masked game, in that the higher number of misses may have caused a self-reinforcing loop of slight struggle leading to more rigorous gameplay, in turn leading to lower accuracy, and again more targets missed.

7.4. Limitation Related to Permutations of Order of Presentation of Games

There were 6 possible permutations and 8 participants, meaning that two of the permutations were repeated twice (namely the permutations with shorthands “B/M/U” and “B/U/M”). This could possibly increase bias towards/against the Masked or Unmasked games since the
Basic game is first in these two permutations, although having more data for any one permutation could conceptually move average sample values closer to average population values (i.e. more samples ideally increases the likelihood of approaching the “true” value). This conflict between minimizing sample size and the required number of permutations ideally should be avoided in future three-tiered studies by changing the sample size to 12, if possible within recruitment limitations.

7.5. Limitations Related to Participant Sample

The results of this study should be taken with caution and may not be generalizable to all stroke survivors, because of the small sample size, potential self-selection bias, and population homogeneity. Even though random selection was attempted during recruiting, any of the demographics variables that exhibited homogeneity (see Demographics Data section and Table 5 - “Participant Demographics”) could have skewed the results to apply to only a subset of the overall population. It may be that only participants self-reporting high levels of motivation towards therapy are willing to participate in the first place. It may be informative to do another study looking specifically into the negative aspects of games, i.e., what deters or does not help with motivation to keep playing and doing reps (upon research ethics approval). Future studies should aim for larger sample populations, and maybe even try to recruit stroke survivors with varying levels of motivation towards therapy.
8. Overall Significance

8.1. Significance for Gaming Development

Masked games should continue to be developed. The games used in this study, as well as the lists of recommendations and lessons learned, should be used as a starting point for future work in developing games for robotic stroke rehabilitation. However, as recommended, a blend of game types, a more holistic view of gaming, and improved development processes should be considered to improve the quality of data and to better align both game preference and exercise performance.

8.2. Clinical Implications

Masked should continue to be considered for use in practice, because some stroke survivors find value in them, and Masked games can help stroke survivors to stay motivated to follow through with exercises in the first place. Future work should make use of the lists of recommendations and lessons learned, as well as all the three games used in this study, to compare motor recovery outcomes, as well as transferability to activities of daily living for stroke survivors. To gain further insight, data from this study could be analyzed by grouping the entire set of comments and observations into clusters, based on a full formal qualitative analysis or thematic analysis, following REB (Research Ethics Board) approval. With improved games to frame exercises, there is the potential to improve the recovery of upper limb mobility post stroke. The outcomes of this project may be applied in future larger clinical studies at Toronto Rehabilitation Institute and potentially at other facilities. There is a potential for the stroke robot system to be manufactured and made commercially available to improve treatment outcomes for stroke survivors, increase the availability of upper limb post stroke rehabilitation, and reduce the potential burden on stroke therapists and the health care system.
9. References


10. Appendices

10.1. Appendix A - Advisory Panel Questions and Brainstorming

- Step 1: Find out what activities/exercises (and the motions that comprise them) are commonly given during therapy & are known to be effective, and brainstorm ideas on how to implement them into games.

- Step 2: Brainstorm effective and/or common ways of framing these motions & exercises into game goals based on real-life activities such as driving a car to a destination or organizing a shelf.

- Step 3: Learn more about what kinds of daily activities stroke survivors would like to be able to perform again, and to break them down into therapeutic motions that can be put into games.

- Step 4: From these realistic activities, brainstorm ideas for fantasy-themed games that offer the same exercises/use the same motions, albeit presented in a fantastical way.

- Step 5: Learn more about what degrees of motion would be appropriate for different kinds of games; for example, for action games (i.e. games with heavy emphasis on motion and eye-hand coordination, like an air hockey game) how can we design them to be inclusive to stroke survivors with impairments while still being engaging?

- Step 6: Learn more about what kinds of games stroke survivors are more inclined to enjoy (for instance, action vs. puzzle games), and what ranges of difficulty in games would be appropriate (so that games can engage players without being boring, and without being too difficult).

- Step 7: Learn about common physical/cognitive impairments caused by stroke & how games should be designed to accommodate for them.

- Step 8: Gather opinions on what both stroke survivors and therapists expect to be offered by therapeutic treatments, so the games may be designed to include them.
10.2. Appendix B - Measurement Definitions and Details

What follows are further details and definitions of measurements used to answer the three specific research questions.

**Time on task (minutes):** During play, the software recorded the time on task or amount of time the user spends on a game. Time on task helps address the objective of determining game design aspects that contribute to users wanting to keep playing for longer periods of time. Time on task could also be an indirect (though weak) supplementary indicator of game preference.

**Number of Motion Repetitions; Reps (#):** The number of motion repetitions was the number of times that the user makes a certain motion in the game. To facilitate the counting of motion repetitions for this study, the kind of motions to be recorded were to be limited to those with a clear beginning and end, as opposed to circular paths which may be ambiguous or difficult to evaluate reliably with programming. For example, the game’s software can count the number of times that a user moves their arm forwards and then backwards (one rep), or left and then right (one rep). The count of the number of repetitions was part of the gameplay experience, such as performing the full motion to produce some kind of discrete action within the game and some kind of visual feedback. For example, throwing a virtual object. This helps address the objective of determining game design aspects that contribute to greater motion repetitions for each game. This measure is different from the number of hits because it is independent of the score and the number of targets hit or missed. A minimum shake or movement was treated as a tremor and hence ignored. However, the caveat with this measure of reps is that it uses an arbitrarily-chosen threshold for the size of what constitutes a(n intended) repetition, based on manually testing motions with the robot arm to simulate a reasonable tremor magnitude based on experience, and then choosing the threshold value for change in velocity (both magnitude and direction). The threshold magnitude of change in position was compared to the absolute difference between the current coordinates and the running average the previous sets of coordinates from the four last timestamps. Another thing to note is that the number of reps for the Basic game was counted by going through the camcorder video recordings and doing counting “manually”, while the numbers of reps for the Masked and Unmasked games were counted “automatically” by the software.

**Number of Hits (#):** This was the number of successfully-reached targets (within a countdown time, when the countdown game mechanic is present). This was different from the
number of reps due to this measure being independent of which path or number of times that different participants move their arms to reach targets. The motion reps can be thought of as relative paths made, and the “hits” as positions reached (i.e. one motion rep = pick up and drop off item, ideally within one linear motion; pick up item = one target hit; drop off item = another target hit). Each “hit” corresponds to a target, but two targets are the minimum for inferring a rep from the number of “hits”.

**Number of Misses (#):** This was the number of targets that were not reached within a countdown time (for the Unmasked game, there was a timer, but for the Masked game, there were non-player characters that could “steal” items).

**Percent of Hits (%):** In contrast to the “Hits (#)” number of successfully-reached targets, this “Hits (%)” was the *percentage* of successfully-reached targets, out of the total number of waypoint targets that were presented to the player. This is a separate measure from the number of hits and misses because it can signal a rate of error, since how many targets are hit is independent from how many targets are missed, at least theoretically.

**Ratios of ROM Usage; PULL (%), PUSH (%), LEFT (%), RIGHT (%):** These were percent measures of how much of their active ROM a participant uses during a game, specifically measuring in four directions, similar to maximum achieved movement in four directions defined by therapists in (Pirovano et al., 2016). This may be useful in indirectly quantifying how much each game encouraged participants to go beyond their current ROM. Each participant sets their active ROM at the beginning of each game, and during each game the robot keeps track of positions of the controller. The percentages reported here only use the maximum values that gave the widest in-game reach. See Figure 19 below for an illustration. For the Unmasked and Masked games, this was automatically tracked and updated in real time. For the Basic game, this was calculated by the researcher based on the data generated by the robot (namely the maximum positions in-game and in the active ROM data) and was not automatically calculated by the software. It should be noted that the active ROM was calibrated differently in the Unmasked and Masked games than in the Basic game. In the Unmasked and Masked games, the active ROM was calibrated when the participants were explicitly told by the game to move their arm in four cardinal directions as far as they could and hold each position for five seconds, whereas in the Basic
game, they were verbally told by the researcher to try to make a “circle” or “oval” on the screen by moving their arm.

![Diagram of ROM Use Percentages in Four Directions](image)

**Figure 19 - Example Schematic of ROM Use Percentages in Four Directions.** Light grey area = the active ROM, simplified to an oval, with edges marked by the red lines and centre marked by crosshairs, calibrated before the game; Dark grey area = where the player moved their arm during gameplay, with edges marked by blue lines; Red lines = the widest points away from the centre of the active ROM; Blue lines = the widest points away from the centre of the active ROM, but during gameplay. Each of the four percentage distances (“PULL” (%), “PUSH” (%), “LEFT” (%), and “RIGHT” (%)) were measured outward from the centre of the active ROM (red crosshairs) and straight out in one of the four directions. Note that the fictitious example in Figure 19 above has the blue mark in the “LEFT” direction moves further left than the respective red mark for “LEFT”, and hence the “LEFT” (%) would have a value over 100% for ROM percent usage.

**Reps/Min:** This was the number of repetitions per minute. This is a measure that suggests the intensity of repetitions by quantifying the number of repetitions over time. The reason for how this measure may be useful as a separate measure from the number of repetitions can be illustrated with a couple of the possible scenarios: participants may perform reps slowly even if they get a high number of reps, or they may perform reps quickly but only get a small number of reps.

**Hits/Min:** This was similar in rationale to the Reps/Min in terms of its inclusion as a separate measure from the number of Hits.

**Explicit game preference:** This was to explicitly ask the participants which of the games they preferred in the questionnaire, after they have played all three games. This helps address the objective of determining which aspects (such as "masking" and other game features) contribute
to stroke survivors’ self-reported preferences (perceived usefulness, enjoyment) for any of the games presented to them.

IMI constructs: These were groupings of the additional questions within the questionnaire, besides the explicit question for game preference. The measures for these constructs serve as additional indicators (besides time on task and repetitions) of aspects that can contribute to overall game preference and to users wanting to keep playing for longer periods of time, and performing a greater number of motion repetitions. Motivation was defined as “the will to continue”; the reported degree of how much a user wants to continue an activity (Longo et al., 1992; Tinsley et al., 1980), and in our case, to continue spending more time with the same therapy exercise, be it the time within a session, or the number of future sessions. The different constructs found in the IMI were perceived usefulness and enjoyment. This study used the following definitions:

1. Perceived Usefulness: “see as useful for their goals”; how much a person thinks an activity or system will help them reach a specific goal (Dickinger et al., 2008), and increase their performance within certain contexts, for example, within the context of a workplace (Davis et al., 1989), or more generally, within the context of exercises towards their personal rehabilitation objectives.

2. Enjoyment: “fun in its own right, not utility-wise”; how much a user perceives using the system to be enjoyable in itself, independent of its (perceived) potential usefulness or impact (Davis et al., 1992).

Questions from each of the above constructs were included in the questionnaire (see Appendix E - “Custom Questionnaire for Stroke Survivor Participants”). These questions were selected to operationalize the construct measures as indicators, so the majority of the questions were taken from existing questionnaires in literature, especially from the Intrinsic Motivation Inventory (“Intrinsic Motivation Inventory,” n.d.), also known as the IMI.

The above constructs were found in the IMI as “sub-scales” or subdivisions of the whole inventory of questions, with each sub-scale having been tested in previous research (“Intrinsic Motivation Inventory,” n.d.). It is not necessary to use the full IMI list of 45 questions (“Intrinsic
Motivation Inventory,” n.d.), just as previous research cited in the IMI document has used shortened versions of the IMI (Deci et al., 1994; McAuley et al., 1987; Ryan et al., 1990; Ryan et al., 1991). Although questions under each sub-scale may be redundant, it is still recommended that one use multiple questions from each sub-scale, mix questions from different sub-scales together, and randomize the order of questions. One also does not have to use all questions for a sub-scale. One example from the IMI document had a set of ten questions, with only two of those questions coming from some of the sub-scales (“Intrinsic Motivation Inventory,” n.d.). This was helpful because it allowed us to reduce the total number of questions and avoid overwhelming participants with the number of questions they were being asked to answer. Questions can also be re-worded to match the context of the experiment (for example, changing the wording of the activity from “reading” to “playing”). A seven-point scale for questions was used, to keep consistency with the existing IMI scale where possible, and to be able to compare results with other studies that use the IMI.

The IMI subscale of interest or enjoyment is considered the self-report measure of intrinsic motivation, however all subscales have been shown to be stable across a variety of tasks and conditions, with experimenters usually using several but not all subscales, and using multiple subscales consistently outperforms using any single subscale (“Intrinsic Motivation Inventory”, n.d.). Shorter versions of the IMI have been shown to be reliable (“Intrinsic Motivation Inventory”, n.d.), based on a study that used 17 questions from the IMI (Colombo et al., 2007), as well as based on two other studies, one that used 26 questions, and another that assessed the psychometric properties of an 18-item version of the IMI, both studies being cited in (Colombo et al., 2007). This study uses 7 questions from the IMI. This is because other questions were to be included in the questionnaire based on ideas generated from input during the advisory panel and during literature search. Questions were selected if they were relevant to the context of the games. For example, questions were not included if they had to do with interpersonal interactions, since users were involved in individual activities. Questions were also selected to be different from each other if possible, by not having questions that say the same thing with different wordings, but while also limiting the number of questions to limit “question fatigue”.
10.3. Appendix C - Rationale for the Measures for Repetitions

Several measures are used for repetitions because they each provide only partial information, which can be used to inform where another measure is limited.

For example, the number of motion repetitions is not the same as the score (i.e. the number of successful target hits), similar to how a 2016 study treated the number of repetitions as a separate measure from successfully achieved movements or score (Cameirão et al., 2016). The number of repetitions reported in this study only approximates the defined count of elbow flexion and extension sequences (Cameirão et al., 2016) because it is automatically counted by the software, using the end-effector’s velocity magnitude and direction changes as detected by the robot (above a tremor threshold, to mitigate false positives). The number of motion repetitions tends to overestimate the number of actual repetitions because multiple motion repetitions may be made before reaching a target, and motion repetitions detected by the robot may not be of a size considered clinically relevant. In contrast, the number of hits tends to underestimate the number of actual repetitions for some users because, although it captures the overall motions over guaranteed distances of movement by counting successfully-reached targets, it is irrespective of the path(s) taken.

Another pair of measures that the reader may wonder about are the average number of target hits and the average number of target misses, which are not completely inversely correlated, and can be thought of as theoretically independent. For example, one may imagine that if each user were given 100 targets, that each target must either be missed or hit (and hence number out of 100), however this incorrectly assumes that each user attempts the same number of targets, and hence this does not apply to these count numbers of hits and misses. On the other hand, this does apply to the percentage of hits and misses, and so only the average percent of successful target hits is included in this study, and not the percent of misses. As for how the number of misses can be interpreted, it could possibly be the number of misses that a player is willing to overlook if they also have a relatively large time on task or have other signs of continued play despite a large number of target misses.

The number of hits and the number of repetitions, versus their respective rates (i.e. average hits per minute and average motion repetitions per minute) are two pairs of measures that capture different information. One can imagine two users completing the same number of 100
targets, but one user completing the targets faster than the other user. One may even further imagine this latter “other” user to have one of the following two situations: they may simply be performing motion repetitions at a slower rate, or they may be making many motion repetitions but are having trouble successfully reaching targets—another reason to include both the hits and the motion repetitions. Again, the number of targets hit/missed does not have a simple dependency on the time on task.

Hence, this study uses measures for more (reps), faster (reps/min and hits/min), bigger (the ROM %’s), and successful (hits, misses, hits %) reps.

10.4. Appendix D - Sources and Groupings of Questionnaire Questions

The questions from the IMI were not the only questions in the questionnaire. Other questions specific to this study with seven-point scales were also used. One additional question was used for the construct of Perceived Usefulness, but was structured to match the question from McKay & Maki, 2010. There were also some open-ended questions in order to obtain more rich feedback. See (Appendix E - “Custom Questionnaire for Stroke Survivor Participants”) for all the questions on the questionnaire.

Table 13 - Sources of the Questions

<table>
<thead>
<tr>
<th>Source:</th>
<th>Questions:</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMI (“Intrinsic Motivation Inventory,” n.d.)</td>
<td>2, 3, 5, 6, 7, 9, 11</td>
</tr>
<tr>
<td>Adapted from (McKay &amp; Maki, 2010)</td>
<td>12</td>
</tr>
<tr>
<td>Inspired from advisory meetings and from the Flow Condition Questionnaire (Human Factors International Inc., 2013)</td>
<td>4, 8</td>
</tr>
<tr>
<td>Inspired from advisory meetings</td>
<td>10</td>
</tr>
<tr>
<td>Ad hoc questions</td>
<td>1, 13, 14, 15, 16; 2b, 3b, 4b, 8b, 9b</td>
</tr>
</tbody>
</table>
IMI Questions for the Constructs:

According to the IMI documentation, the questions taken from the IMI should to be mixed together and randomized (“Intrinsic Motivation Inventory,” n.d.). The groups of questions in the questionnaire that contributed to IMI scoring are regrouped and listed as follows:

1. Perceived Usefulness (4 questions); “see as useful for their goals”

<table>
<thead>
<tr>
<th>Question</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 3</td>
<td>“I think doing this activity could help me to keep up with my exercises.”</td>
</tr>
<tr>
<td>Question 5</td>
<td>“I would be willing to do this again because it has some value to me.”</td>
</tr>
<tr>
<td>Question 6</td>
<td>“I think this is an important activity.”</td>
</tr>
<tr>
<td>Question 9</td>
<td>“I believe doing this activity could be beneficial to me.”</td>
</tr>
</tbody>
</table>

2. Enjoyment (3 questions); “fun in its own right, not utility-wise”

<table>
<thead>
<tr>
<th>Question</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 2</td>
<td>“This activity was fun to do.”</td>
</tr>
<tr>
<td>Question 7</td>
<td>“While I was doing this activity, I was thinking about how much I enjoyed it.”</td>
</tr>
<tr>
<td>Question 11</td>
<td>“I thought this was a boring activity.”</td>
</tr>
</tbody>
</table>
### 10.5. Appendix E - Custom Questionnaire for Stroke Survivor Participants

<table>
<thead>
<tr>
<th>Participant ID:</th>
<th>Researcher:</th>
<th>Date:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Game played this session (circle one):**

1. U
2. M
3. B

**Say this to the participant:** “We’d like to ask you questions about the game you just played.”

**7-point scale for questions 1 to 11:**

1 = not at all true; 4 = somewhat true; 7 = very true.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I felt like I was doing exercise.</td>
</tr>
<tr>
<td>2</td>
<td>This activity was fun to do.</td>
</tr>
<tr>
<td>3</td>
<td>I think doing this activity could help me to keep up with my exercises.</td>
</tr>
<tr>
<td>3.b)</td>
<td>Why do you think that?</td>
</tr>
<tr>
<td>4</td>
<td>It was clear what to do next.</td>
</tr>
<tr>
<td>4.b)</td>
<td>What made it clear / not clear?</td>
</tr>
<tr>
<td>5</td>
<td>I would be willing to do this again because it has some value to me.</td>
</tr>
<tr>
<td>6</td>
<td>I think this is an important activity.</td>
</tr>
<tr>
<td>7</td>
<td>While I was doing this activity, I was thinking about how much I enjoyed it.</td>
</tr>
<tr>
<td>8</td>
<td>It was clear how well I was doing.</td>
</tr>
<tr>
<td>8.b)</td>
<td>What made it clear / unclear?</td>
</tr>
<tr>
<td>9</td>
<td>I believe doing this activity could be beneficial to me.</td>
</tr>
<tr>
<td>9.b)</td>
<td>Why?</td>
</tr>
<tr>
<td>10</td>
<td>I felt I could do the tasks.</td>
</tr>
<tr>
<td>11</td>
<td>I thought this was a boring activity.</td>
</tr>
</tbody>
</table>

**This question is a little different:**

12. I would be willing to do this activity: □ Daily □ 1-3x/week □ Never □ Unsure

**The last few questions are open-ended:**

13. What was your favourite part of the game?

14. What was your least favourite part of the game?

15. Are there any features you would like to see added to the game?

16. [Ask this last, and only if the participant has played all games in the 3 separate sessions:] After playing all 3 games, which do you prefer? And what are your top 3 reasons why? [show clips/ images]

Which game: □ “This game” □ “The last game” □ “The first game” [show clips/images]

Reasons:

1)  
2)  
3)  

Finally, ask how they felt about specific game mechanics/actions/features.
10.6. Appendix F - Game Assignment Permutations

Table 14 - Example assignment of permutations to participants: B = Basic; M = Masked; U = Un-masked.

<table>
<thead>
<tr>
<th>Session</th>
<th>Participant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4 5 6 7 8</td>
</tr>
<tr>
<td>1</td>
<td>B B U U M M B B</td>
</tr>
<tr>
<td>2</td>
<td>M U B M B U M U</td>
</tr>
<tr>
<td>3</td>
<td>U M M B U B U M</td>
</tr>
</tbody>
</table>

10.7. Appendix G - Distilled Notes of Meeting Ideas

What follows is a synthesis of distilled items from meetings with various people as well as the advisory panel. The notes have been regrouped into the following sub-categories for the sake of grouping ideas when performing development:

**It’s Clear What to Do**
- such as simple reminders of instructions if forget
- fantasy level proportional to cognition level
- be careful that fantasy be not too abstract

**It’s Clear How to Do It**
- with controls that are intuitive and obvious to the end user

**It’s Clear How Well You are Doing**
- interface feedback on progress in multiple modes in case one fails to be sensed, but in a simple way to avoiding overstimulation, and also in a non-patronizing way

**There’s a Clear Feeling that You are Able to Do It**
- ease the challenge into bigger ROM use, farther away from body, smaller movements, more alternating movements alternating, faster movements, and timing/count-down
• progress > goals; “more than last time” > “pass/fail/half-way”
• focus on positives, rewards > loss
• “easy --> hard --> easy”:
  • breaks between “bosses”; maybe bonus stages
• basketball hoop analogy
  • adjust both: speed + accuracy
• set up for success
• but challenge enough

**Time**
• integrate regular breaks to avoid injury / fatigue
• vary which muscles are recruited
• vary activities
• don’t worry too much about fatigue
  • easier w. motivation, just think Wii games
  • intense --> benefits
  • they can do 20-30 minutes
  • but can be individual

**Choice**
• tailor to individuals; individualize
  • (gold / tennis / Mario kart with grandkids / driving)
  • [me: can choose tool? can draw tool? can type tool name like Scribblenauts?]

**Robot/Game Features to Account for Cognitive Deficits**
• simple background
• big graphics
• add sound
• sound mute option (in case sound distracting)
• avoid failure sounds
• update ROM (Range of Motion) while playing (or between games)

**Miscellaneous**
• have variability of goals/tools, vs. perceived repetition of same thing
• make it natural to switch between different games
• priority: use ROM the most, but it’s good to target different needs (such as accuracy)
• low-fidelity testing
• humour & fun
• but not childish
• don’t worry too much about “childish”
• new study: Catherine Lang, 2014, speed of movement post-stroke (vs. just movement)
  • “faster --> did better”
• ! keep in mind: translation/generalization to multiple real-life skills!
• CMSA Stages:

**Table 15 - CMSA Stages Notes**

| CMSA Stage 3 | • need to train extension  
|              | • break synergy patterns (but maybe start easier motion for motivation)  
|              | • do not require going back & forth (extend/flex)  
|              | • avoid constant flexion  
|              | • assist to avoid flexion  
|              | • pauses! (breaks for possible fatigue) ☹  
|              | • some resistance good for stage 3 --> touch feedback  

| CMSA Stages 4 & 5 | • more involved motions (but still not “patterned”)  
|                   |  

| Advanced Stages | • save fail states for advanced stages  

**10.8. Appendix H - Game Design Types in Previous Studies**

1) For the MIT-MANUS robot, early games included drawing circles, stars, squares, and diamonds (Reinkensmeyer et al., 2000; Ma et al., 2014), which would seem to fall under this study’s definition of “Masked” games.

2) For the ARMin robot, two “Masked” games were tested: directing a ball through a labyrinth without touching the walls, and moving a virtual handle to catch a ball rolling down a virtual ramp (Staubli et al., 2009).
3) For the UL-EX07, eight games were tested (Kim et al., 2013): four were described as purely assessment tools ("flower", "paint", "joint movement", and "reach"), and four were for testing bimanual/unilateral use ("pong", "circular pong", "pinball", and "hand ball"). All eight of those games made use of 3D graphics (Kim et al., 2013). The former four games involved explicit training of reach and ROM (and hence are "Basic" games), while the latter four games involved activities that may or may not be an ADL for a stroke survivor but are nonetheless framed as realistic sports or games that they may encounter in real life (and hence are "Unmasked" games) (Kim et al., 2013).

4) For the T-WREX, three “Unmasked” games were tested: grocery shopping, cleaning a stovetop, and playing basketball (Housman et al., 2009).

5) For the Trackhold with EEG, five games were tested that would go under “Unmasked” or “Masked” categories: picking up a virtual object and moving it to an avatar sitting at a table, revealing a picture by wiping the screen with a virtual sponge, using a virtual hand holding a swatter to swat a randomly moving bug, making circular arm motions while keeping the cursor within a blue segment on the screen to un-distort an image, and moving a virtual hand straight towards moving butterflies (Steinisch et al., 2012).

6) For the combination of the Cyberglove, CyberGrasp, and Haptic Master, there were two “Unmasked” games (a simulated piano, and hammering a wooden cylinder into the floor) and two “Masked” games (moving a paddle to hit a ball, and catching/releasing a moving hummingbird) that were tested (Merians et al., 2011).

10.9. Appendix I - Data per Subject

Note: this analysis was made after the fact, and was not planned for beforehand.

The Masked and Basic games had maximum values per participant in the table below.

**Table 16 - Games with Highest Values per Subject (M=Masked; B=Basic; U=Unmasked):**

<table>
<thead>
<tr>
<th>Time on Task:</th>
<th>M</th>
<th>M</th>
<th>M</th>
<th>M</th>
<th>B</th>
<th>B</th>
<th>B</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hits (#):</td>
<td>M/U</td>
<td>M</td>
<td>M</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
</tbody>
</table>

The following figures show measurements grouped by game design type, with measures from each game for the same participant connected by lines.
Figure 20 - Time on Task, Within Subjects

Figure 21 - Hits (#), Within Subjects