CHALMERS

Empirical Evaluation of SpaceCat 6 DOF Input Device

Master of Science Thesis in Interaction Design

AMIR CHAMSAZ

Chalmers University of Technology
University of Gothenburg
Department of Computer Science and Engineering
Göteborg, Sweden, August 2010

The Author grants to Chalmers University of Technology and University of Gothenburg the non-exclusive right to publish the Work electronically and in a non-commercial purpose make it accessible on the Internet.
The Author warrants that he/she is the author to the Work, and warrants that the Work does not contain text, pictures or other material that violates copyright law.

The Author shall, when transferring the rights of the Work to a third party (for example a publisher or a company), acknowledge the third party about this agreement. If the Author has signed a copyright agreement with a third party regarding the Work, the Author warrants hereby that he/she has obtained any necessary permission from this third party to let Chalmers University of Technology and University of Gothenburg store the Work electronically and make it accessible on the Internet.

Empirical Evaluation of SpaceCat 6 DOF Input Device

AMIR CHAMSAZ

© AMIR CHAMSAZ, August 2010.

Examiner: Dr. Morten Fjeld

Chalmers University of Technology
University of Gothenburg
Department of Computer Science and Engineering
SE-412 96 Göteborg
Sweden
Telephone + 46 (0)31-772 1000

Cover: SpaceCat (left) and SpaceNavigator (right)

Department of Computer Science and Engineering
Göteborg, Sweden August 2010

Abstract

As technologies for interacting with state-of-the-art specialised systems become more sophisticated, the need for novel mechanisms to interact with multiple elements simultaneously prompts for the development of various multidimensional input devices. Due to the wide range of input devices it is necessary to identify optimum tools for interaction in various environments.

In this study we empirically evaluate SpaceCat softly elastic 6 DOF input device and examine its quality of use compared to SpaceNavigator's stiffly elastic suspension. Such an experiment would give further clues on the usability of a universal softly elastic multidimensional input device compared to specialized input devices for navigation. Using a within-subjects design, evaluation of performance was done according to subjects' progress in Forsaken 3D game. Usability properties of the input devices were further analyzed by assessing their particular aspects of physical operation, mental effort, accuracy and speed, fatigue and comfort, and overall usability through ISO 9241-9 standard questionnaire.

Results from the analysis did not show a significant difference in performance scores for SpaceCat and SpaceNavigator. Consequently, there is no reason to conclude that for rate control, SpaceCat or SpaceNavigator outperform one another. Results from the analysis of subjective rankings were significant for the force required for navigation, smoothness of navigation, and navigation across Z-axis for moving forward in the game. According to our findings, SpaceCat is smoother to operate by demanding less force for performing navigational tasks and provides easier operation functions on Z-axis. As a result, we believe that SpaceCat can perform navigational tasks equally well compared to SpaceNavigator, while it provides smoother navigational experience.

Table of Contents

1. Introduction 5

1.1 Background.......... 6

1.2 Purpose.......... 7

1.3 Method 8

2. Related work 9

3. Hypotheses 12

4. Methodology 13

4.1 Apparatus.......... 13

4.1.1 Framework 13

4.1.4 Cursor Control 17

4.1.2 Screen Capture.......... 18

4.1.3 Videotaping.......... 19

4.2 Experiment.......... 19

4.2.1 Design 20

4.2.2 Participants.......... 20

4.2.3 Procedure.......... 21

5. Experiment Results 23

5.1 Demographics.......... 25

5.2 Performance.......... 27

5.3 Subjective Ratings.......... 28

5.4 Qualitative Video Analysis.......... 31

6. Discussion 33

7. Future Work 35

8. Conclusion 36

References 37

Appendix A: Demographics Questionnaire 38

Appendix B: Subjective Ratings Questionnaire 39

1. Introduction

Universal input devices have evolved over the time from conventional computer mice with their two degrees of freedom to input devices with three or more DOFs for 3D environments. High-end computer applications such as computer-aided design systems (CAD), scientific data visualisations, medical applications, telerobotics, virtual reality programs, and 3D video games require the user to continuously control multiple degrees of freedom in order to interact with the system. The sophisticated nature of these applications promotes the development of various multidimensional input devices.

The SpaceCat, developed by Axiglaze AB, is a 6 DOF input device with elastic suspension to provide rich sensory feedback and a range of transitional and rotational motions adapted to finger manipulation to allow for actions such as pan, zoom and rotate in a CAD-System, as well as navigation through virtual worlds such as moving the viewpoint in walk through and fly through applications [8].

The SpaceNavigator is a stiffly elastic 6 DOF input device. It is used in conjunction with a computer mouse for navigating 3D applications, such as Google Earth and SketchUp. In a basic workflow, one hand holds SpaceNavigator to position the objects or navigate the environment while the other hand simultaneously uses the traditional mouse for pointing tasks.

In the next sections, a brief description of various multidimensional input devices and their characteristics is provided. In addition, the purpose of this study is described and our method for conducting this research is explained.

1.1 Background

Input devices provide the possibility to interact with computer interfaces. Various input devices have been developed in order to provide interactions with computer applications in multiple degrees of freedom. Some examples of these devices include data gloves, motion tracking sensors, and 3D mice such as the SpaceBall by Labtech and 3DConnexion's SpaceNavigator.

In professional applications, the demand for input devices that provide multiple degrees of freedom lies in their capabilities to favor interaction with 3D data. Applications of multiple degrees of freedom input devices are apparent in the context of navigation and manipulation in visualization and animation tools for the development of computer games, in industrial design, and visualization of products in CAD systems. This includes tasks that require functions such as modeling, motion capture, object positioning, object trajectory definition, camera positioning, camera path definition, light source placement and animations, and part assembly, which all can be efficiently supported by a 6 DOF input device [8].

Despite the variety of tasks, Slater and Davison [9] suggest five task categories for multidimensional input devices in 3D environments. They are *Navigation,* including change of viewpoint position and/or orientation, *Global Selection,* including selection of an object in the scene, *Local Selection,* including selection of a part of an object such as a set of specific points, polygons, or patches, *Rigid Body Transformation,* implying the change of an object's position and/or orientation while its local geometry remains unchanged, and *Deformation,* expressing the change of an object's local geometry such as manipulation of control points of polygons or patches. In this study, navigational tasks are examined in order to evaluate SpaceCat in a 3D environment [8].

Multidimensional input devices are classified based on their *device stiffness* and their *transfer function*. Device stiffness is associated with the relation between the device handle position and the force applied to the handle. Zhai [10] suggests three categories for device stiffness. They are *isotonic inputs* with zero stiffness that measure the deflection and are activated by a constant and often very low force, *elastic inputs* with some stiffness that allow for some deflection and provide a counterforce feedback which increases with deflection distance, and finally *isometric inputs* with infinite stiffness that measure force or torque, but do not allow for deflection [8, 11]. SpaceCat and SpaceNavigator are both elastic devices.

Transfer function accounts for the relation between the devices handle position and the movement of an object in an application. Three types of transfer functions are commonly in use. They include *position control,* where object's position and orientation are proportional to the deflection of device handle, meaning that the transfer function from device handle to object movement is a constant. *Rate control* (or *velocity control*) means that object's translation and rotation velocity are proportional to the deflection of the device handle. In other words, the device maps the handle input to the velocity of the object movement. The third type is called *acceleration control.* Acceleration controls are inputs where object's translation and rotation acceleration are proportional to the deflection of the device handle [8]. SpaceCat and SpaceNavigator conform to elastic type of input devices while they maintain distinctive stiffness properties. SpaceCat has a stiffness of approximately 1.3N/cm and a grip stiffness of approximately 1.2N/cm, while SpaceNavigator holds a stiffness of 20N/cm, 24N/cm, and 36N/cm for dX/dY translation, dZ translation, and rotation about the Z-axis, respectively [8, 12]. Due to their varied range of motion and particular stiffness, SpaceCat is considered as a *softly elastic* and SpaceNavigator as a *stiffly elastic* input device. For the purpose of this study, SpaceCat and SpaceNavigator, a softly elastic and a stiffly elastic input device, were compared and evaluated for rate control.

1.2 Purpose

This study aims at empirically evaluating SpaceCat 6 DOF input device and examining its quality of use. In this study, we intend to compare two elastic inputs for rate control, with the SpaceCat being a softly elastic input device and the SpaceNavigator, a stiffly elastic input device. We are further planning to evaluate usability properties of the input devices by assessing their particular aspects of physical operation, mental effort, accuracy and speed, fatigue and comfort, and overall usability through a subjective questionnaire. Results from this study will show whether the SpaceCat's softly elastic transition provides superior performance outcomes and will explain usability characteristics of both devices in comparison to each other under a standard framework.

1.3 Method

A within-group subject design was employed to compare SpaceCat with SpaceNavigator according to their performance in a 3D game environment. Twelve participant took part in the experiment and played Forsaken 3D game with both devices. Subjects were scored based on their progress in the game. A subjective ratings questionnaire was completed after each trial. Subjects' scores were analyzed using the paired-samples t-test and the data from the questionnaire were analyzed using the non-parametric Wilcoxon signed-rank in SPSS.

Figure 1. SpaceCat

Figure 2. SpaceNavigator

2. Related work

The bulk of literature on input devices with multiple degrees of freedom is continuously expanding with new types of multidimensional devices, as well as different interaction techniques. Various performance measures have been put through by researchers in pursuance of evaluating 6 DOF input devices and to analyze how users' performance relates to explicit design attributes for 6 DOF interfaces. A brief overview of the related researches is provided in this chapter.

An empirical within-group experiment was conducted by Sundin and Fjeld [8] to comparatively evaluate SpaceCat and a commercially available 6-DOF input device by Techlab called SpaceBall. The input devices, SpaceCat and SpaceBall, were compared in a 3D docking task. They found out that in terms of completion time SpaceCat being a softly elastic position control outperforms SpaceBall, which is a stiffly elastic rate control. In terms of learning, they realized that inexperienced users significantly preferred SpaceCat's position control to SpaceBall's rate control for solving a CAD task, while for the experienced users, the control order or the kind of input device did not play a major role. They concluded that SpaceCat's softly elastic suspension creates an advantage for achieving both position and rate control compared to SpaceBall's stiffly elastic suspension.

Zhai [10] compared a magnetic tracker with isotonic position control and SpaceBall 2003 with isometric rate control together in a 6 DOF positioning task experiment. The isotonic position control proved superior over isometric rate control with its shorter learning time and shorter task completion time. However, the magnetic tracker accounted for higher fatigue due to its inherent characteristic of operating freely in the air. As a result, the isotonic device provided faster interaction and was easier to learn, while the isometric device showed less fatiguing, as it allowed for the arm to rest on the desktop during operation.

In another experiment, Zhai [10] compared the “elastic general-purpose grip” with elastic rate control and the SpaceBall 2003 with isometric rate control together in a 6 DOF positioning task. There were no significant differences observed between the two devices regarding task completion time, however, the “elastic general-purpose grip" showed shorter learning time. The experiment was repeated for a 6 DOF tracking task experiment and demonstrated the same results as for the 6 DOF positioning task. In their study, Zhai describes that since elastic rate controls provide force feedback through their elastic elements, they allow for greater proprioception and hence, account for shorter learning time compared to isometric rate controls. Zhai concluded that the combinations of isotonic devices with position control and the combination of elastic and isometric devices with rate control are superior over other combinations.

Li [13] conducted an experiment to find out the optimal control mode for pointing with a universal input device using 6 DOF softly elastic input. She used quCat to compare softly elastic position control and softly elastic rate control in a pointing task experiment. QuCat is a 6 DOF input device developed by 3hird Dimension. It is based on the same technology as SpaceCat, but with stiffer springs and smaller range of motion. The experiment included 96 tasks for both position control and velocity control. Results showed that in most tasks position control outperformed velocity control in terms of task completion time, while for long distance tasks the difference was smaller.

In another experiment, Li [13] compared a conventional mouse with isotonic position control and quCat with softly elastic position control in a pointing task experiment to assess the potential of a universal input device that is suitable for 3D tasks as well as 2D pointing. In this experiment, Li replaced the softly elastic rate control with a conventional (isotonic) mouse that used standard Microsoft Windows 2000 settings and conducted the same experiment again. The data was then compared to the data already collected in the previous experiment with softly elastic position control. Results showed that task completion time was 28% quicker for the conventional mouse compared to quCat, with larger difference for higher precision and longer movements. The faster response from conventional mouse, however, should be considered with regards to participants' lack of experience with softly elastic input devices and their extensive experience with conventional mouses. Although using a universal input device for 2D and 3D results in slower response compared to a conventional mouse, it eliminates the switching time between devices for 2D and 3D tasks. As a result, softly elastic inputs maintain their motivation for 2D pointing even if they do not exhibit shorter task completion times.

Froehlich et al. [11] conducted a study to examine the combination of isotonic and elastic inputs where translational and rotational inputs are separated. They compared GlobeFish and GlobeMouse with SpaceMouse, a commercially available 6 DOF input device from 3DConnexion. GlobeFish and GlobeMouse incorporate a 3 DOF + 3 DOF design where the 3 DOF elastic translation allows uniform input for all three axes and a 3 DOF isotonic trackball provides a natural mapping for rotations. The GlobeFish and the GlobeMouse differ in the sense that for the GlobeFish, the trackball is accessible from the top and bottom and can be moved slightly in all spatial directions, while for the GlobeMouse, the trackball is placed on top of a movable base, which requires to change the grip on the device to switch between rotating the trackball and moving the base. They performed a 3D docking task to determine their general performance and separately study rotation and translation performances. Results from their experiment showed a significant performance advantage of over 20% for the two combinational devices compared to SpaceMouse. The concluded that GlobeFish and GlobeMouse performance advantage is mainly due to their efficient isotonic trackball rotation that provides a clear separation between input modalities.

According to the established works in evaluation of multidimensional input devices, little information is available concerning the performance and usability of a universal softly elastic 6 DOF input device compared to specialized input devices for rate control in 3D environments. Figure 3 illustrates the previous combinations of device stiffness and transfer functions for positioning tasks that were provided in this section. We put forward three hypotheses in order to evaluate the performance and usability qualities of SpaceCat compared to SpaceNavigator for rate control.

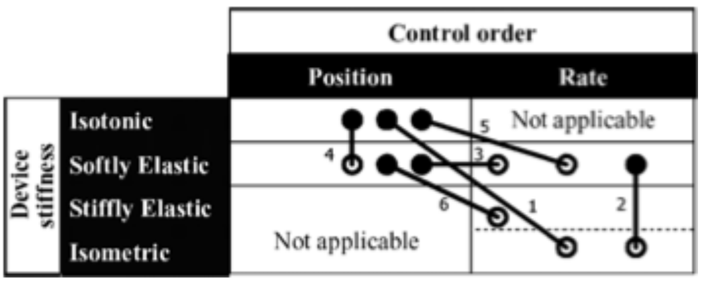

Figure 3. Device stiffness and control order for positioning tasks in previous works [8]
Note. The “double ball bars” refer to previous positioning experiments that compared different kinds of interaction. Solid circles indicate more efficient kinds of interaction.

3. Hypotheses

H1. For rate control, SpaceCat outperforms SpaceNavigator on overall score

Overall performance of each device is assessed by setting interval levels in Forsaken game and observing how far the participants can progress during 10 minutes of playing with each device. This measure was designed in order to determine whether there is a significant difference in using SpaceCat or SpaceNavigator for playing the game and if any of the devices would allow users to achieve higher scores.

H2. SpaceCat requires lower force in navigation compared to SpaceNavigator

The physical force that each input device requires in order to perform navigational tasks was tested with this hypothesis. The data was collected using a questionnaire where participants ranked their experience in playing the game for 10 minutes with each device.

H3. SpaceCat provides smoother navigation compared to SpaceNavigator

SpaceCat is a softly elastic input device, opposed to SpaceNavigator, which is stiffly elastic. We expect that SpaceCat is smoother to operate in navigation compare to SpaceNavigator. We also test the smoothness across X, Y, and Z axes by assessing which device was easier to operate in each direction using the following sub-hypotheses:

H3.1 Moving forward (across Z axis) is easier using SpaceCat compared to SpaceNavigator

H3.2 Turning left and right (across X axis) is easier using SpaceCat compared to SpaceNavigator

H3.3 Turning up and down (across Y axis) is easier using SpaceCat compared to SpaceNavigator

4. Methodology

A within-subject design was used to test the presented hypotheses in this study. Forsaken 3D game was selected as the testing environment to support the requirements of our hypotheses for 3D navigation with rate control. In the course of the research in hand, our methodology for evaluation of the input devices went through a number of modifications and was repeatedly revised until it reached a state of integrity to base our experiments on. We tested Fitts's Law and ISO 9241-part 9 as theoretical backgrounds to find out if they can be suitable for the case of our evaluation and experienced with various software frameworks to use in order compare the input devices. The computer screen was captured during the games in order to make it possible to review back subject's progress at a later time. Participants were also videotaped during their tests for further review and analysis of their behavior with both devices.

4.1 Apparatus

In the following section, our software framework for evaluation of input devices is discussed and the reasons for our approach is explained. In addition, the requirements for configuring the devices in the game environment is noted and a brief overview of various screen capturing software, as well as our videotaping procedure is provided.

4.1.1 Framework

Fitts's law is a well-established model for predicting the time required for moving to a target and clicking on it as a function of the distance to and the size of the target. Fitts's law is mainly used to compare and evaluate new input devices. It calculates an index of performance IP (also called throughput TP) by dividing the index of difficulty ID (in bits) averaged over a block of trials, by the average movement time MT (in seconds). This index of performance is obtained through linear regression and can be used across different input devices in order to compare their performance.

ISO 9241 part 9, requirements for non-keyboard input devices, is a standard developed by the International Standards Organization for evaluation of non-keyboard input devices. It is a methodology designed to evaluate performance and comfort of non-keyboard input devices. The performance is evaluated according to one of the six tasks propose by the method and is calculated in terms of throughput which is based on Fitts's index of performance. ISO 9241-9 suggests using standard deviation of distance over a block of trials instead of the normal Fitts's index of difficulty. Consequently, ISO 9241-9 is considered to be more effective in prediction performance compared to Fitts's law as it accounts for the spatial variability observed in responses [1].

Our investigation showed that Fitts's law and ISO 9241-9 standard would not be suitable for the purpose of evaluating performance in our study. We realized that both theories are based on their implications for assessing the positioning control characteristics of input devices rather than being suited for rate control observations.

We studied several usability assessment methods that were based on subjective ratings of devices. They included NASA-TLX workload assessment, SUMI usability assessment, SUS usability scale, and ISO 9241-part 9 assessment of comfort independent rating scale. The NASA Task Load index is a subjective rating procedure that evaluates workload based on six sub-scales. They are mental demands, physical demands, temporal demands, own performance, effort, and frustration [16]. SUMI is a product usability assessment tool that measures user satisfaction based on five sub-scales and a global scale. The sub-scales include efficiency, affect, helpfulness, control, and learnability [17]. Finally, SUS is a usability scale that can be used for global assessments of systems usability. It is consisted of ten usability questions that are rated according to a five-scale rating. SUS was developed as a freely available usability assessment tool and it correlates well with other subjective measures of usability such as SUMI [18].

ISO 9241-part 9 evaluates comfort and usability by asking subjects to rate their experience with input devices independently and comparatively. The questionnaire is designed according to the devices with the highest score representing those most preferred. The standard questionnaire consists of twelve questions to assess various aspects of input devices including attributes of physical operation, accuracy and speed, fatigue and comfort, and overall usability. Comparative evaluations are then assessed by comparing significant differences between the devices for each rated item [15]. ISO 9241-part 9 independent rating scale was employed for the assessment of usability and comfort in our study. Although the alternative usability scales are applicable for the evaluation of input devices, ISO 9241-part 9 measures are particularly developed for the assessment of non-keyboard input devices and provide more appropriate scales. They assess specific physical aspects of operation with input devices that other scales do not provide the means to evaluate such as the force required for operation, smoothness of the device, and operation speed, as well as finger and wrist fatigue. ISO 9241-part 9 questionnaire was best consistent with our study requirements and was modified in order to accommodate our hypotheses.

Several 3D environments were tested in order to find an environment that is compatible across the two devices and enables for the evaluation of input devices in a consistent manner. A Sword-handling application developed for SpaceCat at t2i lab [2] was examined first. This application is a first person sword wielding simulation that makes use of all 6 DOFs provided by the device. Since the software was primarily developed to work with SpaceCat, new filters were added to the application to make it compatible across the two devices.

Despite all the time and effort put in adapting the application to suit the experiment, it did not turn out to be robust enough to support our experiment. Firstly, the application was developed to accommodate position control instead of velocity control. Users had to navigate through the game using the keyboard while they could manipulate the objects using SpaceCat. Secondly, the application slowed down considerably after manipulating objects in a few number of trials, which made it unreliable to use. To address the first problem, Total Game Control from Digital Transforms was used to map keyboard events to the controls on spaceCat and SpaceNavigator. Total Game Control provides the option to map specific keys on the keyboard to specific axes of an input device. As a result, it enabled us to assign related keyboard keys for navigation in relevant axes of SpaceCat for moving through the game environment. However, Total Game Control was not compatible with SpaceNavigator and it was not feasible to set up the Navigator in the same way as SpaceCat. The second problem could not be solved either, as it was related to a memory leak in the software and required major recoding of the sword-handling application.

Google Earth was the second environment that was tested for the purpose of this study. Using 6 DOf capability of input devices it is possible to navigate through 2D maps and 3D environments in Google Earth more effectively as they provide finer control over the navigation and enable users to perform multiple actions such as zooming and tilting the view at the same time.

While Google Earth worked smoothly with SpaceNavigator it did not perform accurately with SpaceCat. Different calibration adjustments were made to adapt SpaceCat with Google Earth, but the end result was not satisfactory to accommodate our experiment. Designing a solid task to examine input devices in Google Earth environment was also challenging. Google Earth uses KML file format to display geographic data and makes it possible to develop customized codes for navigation in Earth browsers such as Google Earth, Google Maps, and Google Maps for mobile. A KML code was planned to be developed in order to set the viewer at constant distant from particular map objects and calculate the time-on-target on a set of trials. This idea was eventually abandoned due to the incompatibility of SpaceCat with Google Earth environment. Other tasks that were investigated were to test Google Earth in Street View and a helicopter simulator layout. In both cases the environment was not functional by either 6 DOF input devices.

The next framework to examine was Fitts's framework developed by A. De Sena and D. Moschini [3] at Verona University. Fitts's framework project has been developed to allow researchers to test Fitts's law by studying and collecting data on pointer devices. It displays a set of circular targets that lie on a circumference. At each time only two circles are visible and participants must hit the targets while trying to be fast and accurate. The test is run across a set of trials where for each trial 26 targets should be hit. The framework records all the clicks made by participants for further analysis. The main argument for not employing Fitts's framework in this study was its design objective to test positioning control. We realized that the framework is not suited for navigational intends and thus it is not applicable to our experiment.

Two other simulator environments that were compatible with 6 DOF input devices and enabled navigation were tested. FlightGear developed by FlightGear project, which is an open-source flight simulator development project and Space Station Simulator, a Shuttle Docking Simulation developed by NASA, were examined, but none provided a solid framework to base our experiment upon.

Lastly, Forsaken 3D game was reviewed and based on its suitable application it was chosen for our experiment. Forsaken was compatible with both devices and provided reliable support for both 6 DOF input devices. In addition, Forsaken allowed for rate control navigation and supported our requirements for testing the hypotheses. Forsaken is a 3D first person shooter game with 6 DOF gameplay design and allows unlimited 360-degree movements. The primary objective of the game is to destroy the enemies while navigating towards the end of each level within a time limit. Our experiment was carried out in the God mode that enabled invulnerability, full weapons, and unlimited ammo.

Figure 4. Forsaken 3D Game Environment

4.1.4 Cursor Control

A requirement of the experiment in 3D environments including sword-handling application and Fitts's framework, as well as simulators was to enable the 6 DOF device in cursor control mode. By default, SpaceCat and SpaceNavigator do not perform as a mouse to control the cursor on the screen and thus, a driver is required to configure them to work in software environments that natively do not support them.

JMouse from Phelios inc. was the first application that was tested. JMouse is a freeware that enables users to configure a joystick and use it instead of a mouse. It allows users to assign the buttons of the joystick to specific actions in the game. Joystick 2 Mouse is a similar freeware application that provides extended functionalities compared to JMouse and allows the user to map the joystick's axes, buttons, and POV and enables users to assign an action for every button. 3D Mouse from Claro software was the next application that was examined. 3D Mouse is commercial application that allows 3D Mouse controllers including 6 DOF input devices to function as a standard mouse controller. 3D Mouse was the most advanced cursor control application that we came across, but despite of its functionality it was replaced with RBC9, which provided sufficient support for our experiment.

RBC9 is a free driver developed for SpaceNavigator that enables it to function as a Human Interface Device and natively be supported by applications that are compatible with Windows HID devices. RBC9 allows users to define customized layouts to function in different applications and automatically switch between them. It enables users to bind an axe on SpaceNavigator to one axe or multiple axes on a Joystick and configure its sensitivity and delay time.

4.1.2 Screen Capture

Capturing the screen while subjects are performing a test enables the researcher to review subjects' progress for a number of times proceeding the test and provides them with the means to carefully observe differences in the gameplay among various subjects.

A number of screen capturing software were explored in order to find a suitable application to record the gameplay. Camtasia Studio 7.0 turned out to considerably slow down the system and therefore, was dismissed. My Screen Recorder 2.48 generated small screen recordings that were not clear to follow and was eliminated, too. HyperCam 2.23.01 recorded the screen as frame-by-frame screen shots that appeared like progressive images after one another and as a result it was also dismissed. Fraps 3.2.2 was the application that we found effective and useful for recording the screen in our experiment.

Fraps is a benchmarking, screen capture, and real-time video capture application for games. It provides support for assessing computer's performance with games, as well as recording gaming footage. Since Fraps is developed to work with games, it never slowed down the system during the game. It was capable of providing up to 2560x1600 resolutions for screen recording and costume specified frame rates from 10 to 120 frames per second.

4.1.3 Videotaping

Observing the subjects while performing a test is advantageous in explaining the reasons behind certain results and finding bases for test achievements. While qualitative interpretations of experimental procedures help in connecting the findings to motivations behind them, they supplement the experiment by adding meaning to the data.

For our experiment, subjects were videotaped during their experiment with both devices using a HD Camcorder. The Camcorder was positioned in a 45 degrees angle towards the subjects and captured their body posture, their arm, and their hand position while they were performing the experiments. Recordings provided the possibility to monitor subjects' behavior during the experiment and review them for future analysis. It further enhanced the validity of the results by enabling repeated inspection of the data.

4.2 Experiment

In the following section, our experimental design is described in detail. Moreover, the participants and their demographic information is explained. Finally, an overview of the test procedures is provided.

4.2.1 Design

The experiment employed a within-subjects design. All participants were tested using both devices. The order of the input device was counterbalance so that half of the participants experimented with SpaceCat in their first trial and another half used SpaceNavigator for their first trial.

The path throughout the game was divided into thirty-three interval points and subjects could score from zero to thirty-three according to their progress in the game. They were required to kill all enemies and collect certain objects along the game. Each trial would start from the beginning of the game for 10 minutes. The dependent variable for the experiment was performance, which was measured based on subject's score (progress) in the game. The three axes of input devices were assigned identically to actions in the game. X-axis was assigned to turning to the sides, Y-axis to turning up and down, and Z-axis to moving forward along the game.

The subjective rankings questionnaire was based on standard questionnaire from ISO 9241-9, requirements for non-keyboard input devices. In total, there are twelve questions in ISO 9241-9 from which three where omitted and replaced with another three questions. The omitted questions included arm, shoulder, and neck fatigue. Three additional questions where added to the standard format that assessed the ease of navigation in moving forward, turning left and right, and turning up and down. These questions where designed in order to investigate differences in operation across the three axis of input devices compared to each other. In total, twelve questions were employed, each with a rating from 1 to 5 where lower rates represented more favorable effects across all questions. A copy of the questionnaire is provided in appendix B.

4.2.2 Participants

Fourteen participants joined the experiment out of which results from twelve were accepted. In two cases the game failed due to a bug that trapped subjects under a rock during the middle section of the game. Of the twelve participants eight were male and four were women with ages ranging from 23 to 30 (*mean* = 26.7). All participants were right handed, although this was not by design.

All participants were assessed using a demographics questionnaire at the begging of the experiment. Participants experience with games was determined with a five-scale question inquiring how often they play video games. Six participants seldom played video games and one played games very often. The other five positioned in the middle range.

Participants were also asked about their previous experience with 6 DOF input devices according to a five-scale category, which ranged from 1, with no experience to 5, experienced user. Eleven did not have any prior experience with 6 DOF input devices and one had some experience. Although our participant pool did not reveal much diversity in terms of their experience with 6 DOF devices, we believe that our participants' varied experience allows for generalization of our findings to a larger population.

4.2.3 Procedure

Each experiment took about 45 minutes to complete. A demographic questionnaire was handed to each participant to collect information such as their experience with video games and prior experience with any 6 DOF input devices, as well as their age, gender, and handedness. A copy of the demographic questionnaire is provided in Appendix A. Prior to beginning, participants were briefed about the goal of the experiment and received instructions about the game and the device that they were going to use. They were then asked to relax down and watch a demonstration video of the gameplay, which would acquaint them with the game environment.

The game was played in the God mode that provided invulnerability, full weapons, and unlimited ammo. The God mode enabled players to proceed in spite of whether they were hit or not and eliminated inequalities due to starting from the beginning of the game for each time they were killed.

All participants explored playing with the device for a few minutes before the trial until they feel confidant about how the device works. Each trial started from the beginning of they game when the participant was ready and continued for 10 minutes. The time was kept using a stopwatch. Participants were taped with a HD camcorder and a screen capture software recorded their play for future analysis.

Upon finishing a trial, participants were asked to complete an assessment questionnaire and rate their experience with the device. They could then have a break before proceeding to the second trial. Prior to the second trial, every participant received instructions about the second device and was allowed to play with it for a few minutes until they felt confident with operating the device.

The second trial started from the beginning of the game for another 10 minutes. Afterwards, participants received the second assessment questionnaire to complete and rate their experience. At this point, the experiment was complete and the questionnaires, as well as the HD recordings and screen captures where categorized.

Figure 5. SpaceCat and Forsaken 3D Game

Figure 6. SpaceNavigator and Forsaken 3D Game

5. Experiment Results

Device performance was analyzed using paired-samples t-test in SPSS. T-test is a statistical procedure for comparing two groups of data and examining if the two groups are different by assessing whether their means are statistically different from each other. A paired-sample t-test is used in order to compare two related samples such as samples in a within-subject design. A paired sample t-test takes differences between data in the two group scores, and checks whether the distribution of the differences is too different from the t distribution. If the distribution is significantly different, then the null hypothesis is rejected and effects are concluded. For the purpose of this study, subjects' score with SpaceCat was compared to their score with SpaceNavigator in order to determine whether any of the two devices help in achieving better results.

T-test assumes normal distribution of the sample data, meaning that the data should be collected from a normally distributed population. Shapiro-Wilk test is a statistical procedure for examining whether the data are normally distributed in samples with smaller number of cases. The null hypothesis for Shapiro-Wilk test is that the data are normally distributed. For an alpha at 0.05 level, if the p-value shows less than 0.05 the null hypothesis is rejected. Results from the Shapiro-Wilk test for device performance revealed non-significant and thus, provided support for normality assumption of our t-test. Table 1 presents results of the Shapiro-Wilk test.

Table 1. Shapiro-Wilk test of normality for performance variable

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
SpaceCat	.238	12	.060	.932	12	.397
SpaceNavigator	.131	12	.200*	.935	12	.436

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Wilcoxon signed-rank, which is a non-parametric test, was chosen as the statistical procedure for analyzing the subjective ratings data in this study. Test of normality proved significant for most variables in our sample and as a result, normality assumptions were not met for a parametric t-test. The choice of a non-parametric test was based on implications of this type of analysis to provide valid and reliable test results based on our sample attributes.

Non-parametric procedures are statistical methods that involve estimating or testing the value of parameters such as population means or proportions without reference to specific parameters. They do not rely on the estimation of parameters such as the mean or the standard deviation, as such, non-parametric tests are suitable when sample sizes are small and/or the data is not normally distributed.

Non-parametric methods have particular advantageous over parametric tests. They demand fewer underlying conditions or assumptions to be met in order to produce valid results. For instance, paired sample t-test requires that the data is collected from a normally distributed population. It also assumes that the variance of two samples is same. If these assumptions are not met, the resulting P-values and confidence intervals may not be valid. This situation particularly happens in the case of smaller samples sizes [4].

In addition, non-parametric statistics can produce more reliable results when measurements lack a precise underlying scale that is universally recognized. For example, parametric tests such as analysis of variance, t- tests, and regressions assume that measurements are at precise intervals, meaning that they represent equally spaced intervals on the scale. Consequently, non-parametric tests are more suitable in cases of ordinal data that represent a rank ordering of observations rather than precise measurements [5].

Non-parametric methods, however, are regarded as less sensitive in detecting existing differences between populations. As such, in non-parametric methods the information is preserved in the form of ranks while the actual values are discarded. In addition, it becomes more difficult to make quantitative statements about the actual difference between populations in non-parametric procedures, as there are no parameters to describe [4].

Wilcoxon signed-rank test is a statistical technique that is used to compare median differences of a population for the case of two related samples or repeated measurements on a single sample. The test assumes that samples are derived from a random population, with a symmetric frequency distribution. The symmetric assumption does imply normality of the data. It presumes that the distribution of the differences is symmetric, meaning that the sample includes approximately the same number of values above and below the median [6].

Wilcoxon signed-rank test calculates the differences between the first and second measurements for each pair. It then ranks the differences according to their absolute value by ignoring the signs and setting the values in an ordered list from one to the number of pairs. Next, the ranks of the positive and negative differences are summed up separately. If the null hypothesis is true, it is expected that half of the values stand above the median and thus, the rank sums for positive and negative ranks to be the same [7].

5.1 Demographics

Descriptive analysis of demographics data is provided in the following tables.

Table 2. Demographic Descriptive Analysis

	N	Minimum	Maximum	Mean	Std. Deviation
GameXP	12	1	4	1.83	1.030
DeviceXP	12	1	2	1.08	.289
Age	12	23	30	26.75	2.179
Gender	12	1	2	1.33	.492
Hand	12	1	1	1.00	.000
Valid N (listwise)	12				

Table 3. Game experience Frequency

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	6	50.0	50.0	50.0
	2	3	25.0	25.0	75.0
	3	2	16.7	16.7	91.7
	4	1	8.3	8.3	100.0
	Total	12	100.0	100.0	

Table 4. Device experience Frequency

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	11	91.7	91.7	91.7
	2	1	8.3	8.3	100.0
	Total	12	100.0	100.0	

Table 5. Participants age Frequency

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	23	1	8.3	8.3	8.3
	24	1	8.3	8.3	16.7
	25	2	16.7	16.7	33.3
	26	1	8.3	8.3	41.7
	27	2	16.7	16.7	58.3
	28	2	16.7	16.7	75.0
	29	2	16.7	16.7	91.7
	30	1	8.3	8.3	100.0
	Total	12	100.0	100.0	

Table 6. Gender Frequency

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	male	8	66.7	66.7	66.7
	female	4	33.3	33.3	100.0
	Total	12	100.0	100.0	

Table 7. Handedness Frequency

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	right handed	12	100.0	100.0	100.0

Table 2 shows descriptive statistics for demographic data. It shows the number of cases (N), minimum and maximum scores of each response, mean, and standard deviation for each variable. Tables 3 to 7 present frequency tables for demographic variables. They show the frequency, percentage, valid percentage (without missing values), and cumulative percentage for each score. The cumulative percent for a given score demonstrates the percentage of cases with smaller scores or equal scores to that score.

5.2 Performance

A paired-samples t-test was conducted to compare device performance for SpaceCat and SpaceNavigator. Device performance was the Dependent Variable that was examined according to subject's score with both devices. Results from the analysis using SPSS are presented in the following tables.

Table 8. Paired Samples t-test descriptive statistics for performance

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Cat_Score	17.83	12	7.133	2.059
	Nav_Score	18.50	12	5.402	1.559

Table 9. Paired Samples t-test correlations for performance

		N	Correlation	Sig.
Pair 1	Cat_Score & Nav_Score	12	.922	.000

Table 10. Paired Samples t-test results for performance

		Paired Differences				
					95% Confidence Interval of the Difference	
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper
Pair 1	Cat_Score - Nav_Score	-.667	2.995	.865	-2.570	1.236

		t	df	Sig. (2-tailed)
Pair 1	Cat_Score - Nav_Score	-.771	11	.457

Table 8 provides descriptive statistics for performance variable. It shows the mean, number of scores (N), standard deviation, and standard error of the mean for SpaceCat and SpaceNavigator, separately. Table 9 shows the number of pairs (N), the correlation between the two devices, and the significance of the correlation (sig.). Table 10 displays the mean, standard deviation, standard error of the mean, and 95 percent confidence interval of the paired differences, the value of t, the degrees of freedom (df), and the two-tailed significance level.

5.3 Subjective Ratings

The subjective ratings assessment of input devices was based on the analysis of device assessment questionnaire. A Wilcoxon signed-rank test was conducted to compare twelve aspects of user experience for SpaceCat and SpaceNavigator. Results from the analysis are provided in the following tables.

Table 11. Subjective ratings Descriptive Analysis

	N	Mean	Std. Deviation	Minimum	Maximum
1. The force required for navigation	12	1.33	.492	1	2
2. The mental effort required for operating the device	12	2.67	.888	1	4
3. Smoothness during navigation	12	1.83	.835	1	3
4. Moving forward	12	2.00	1.044	1	4
5. Turning left and right	12	2.42	1.084	1	4
6. Turning up and down	12	2.67	1.155	1	5
7. Accurate targeting	12	2.67	1.155	1	4
8. Navigation respond	12	2.08	.996	1	4
9. Finger fatigue	12	1.83	.718	1	3
10. Wrist fatigue	12	2.58	1.165	1	4
11. General comfort	12	2.42	.996	1	4
12. Overall the input device	12	2.58	1.084	1	4
1. The force required for navigation	12	2.83	1.267	1	5
2. The mental effort required for operating the device	12	2.75	.866	2	4
3. Smoothness during navigation	12	3.17	1.030	1	4
4. Moving forward	12	2.50	1.000	1	4
5. Turning left and right	12	3.08	.793	2	4
6. Turning up and down	12	3.17	.937	2	5
7. Accurate targeting	12	3.25	1.357	1	5
8. Navigation respond	12	2.17	.937	1	4
9. Finger fatigue	12	2.33	1.073	1	4
10. Wrist fatigue	12	2.50	1.243	1	4
11. General comfort	12	3.17	1.267	1	5
12. Overall the input device	12	3.00	1.044	2	5

Table 12. Wilcoxon Signed Ranks for subjective ratings

		N	Mean Rank	Sum of Ranks
1. The force required for navigation - 1. The force required for navigation	Negative Ranks	1[a]	3.00	3.00
	Positive Ranks	10[b]	6.30	63.00
	Ties	1[c]		
	Total	12		
2. The mental effort required for operating the device - 2. The mental effort required for operating the device	Negative Ranks	4[d]	5.50	22.00
	Positive Ranks	5[e]	4.60	23.00
	Ties	3[f]		
	Total	12		
3. Smoothness during navigation - 3. Smoothness during navigation	Negative Ranks	1[g]	3.00	3.00
	Positive Ranks	9[h]	5.78	52.00
	Ties	2[i]		
	Total	12		
4. Moving forward - 4. Moving forward	Negative Ranks	1[j]	2.50	2.50
	Positive Ranks	5[k]	3.70	18.50
	Ties	6[l]		
	Total	12		
5. Turning left and right - 5. Turning left and right	Negative Ranks	2[m]	4.75	9.50
	Positive Ranks	7[n]	5.07	35.50
	Ties	3[o]		
	Total	12		
6. Turning up and down - 6. Turning up and down	Negative Ranks	3[p]	3.00	9.00
	Positive Ranks	5[q]	5.40	27.00
	Ties	4[r]		
	Total	12		

a. 1. The force required for navigation < 1. The force required for navigation
b. 1. The force required for navigation > 1. The force required for navigation
c. 1. The force required for navigation = 1. The force required for navigation
d. 2. The mental effort required for operating the device < 2. The mental effort required for operating the device
e. 2. The mental effort required for operating the device > 2. The mental effort required for operating the device
f. 2. The mental effort required for operating the device = 2. The mental effort required for operating the device
g. 3. Smoothness during navigation < 3. Smoothness during navigation
h. 3. Smoothness during navigation > 3. Smoothness during navigation
i. 3. Smoothness during navigation = 3. Smoothness during navigation
j. 4. Moving forward < 4. Moving forward
k. 4. Moving forward > 4. Moving forward
l. 4. Moving forward = 4. Moving forward
m. 5. Turning left and right < 5. Turning left and right
n. 5. Turning left and right > 5. Turning left and right
o. 5. Turning left and right = 5. Turning left and right
p. 6. Turning up and down < 6. Turning up and down
q. 6. Turning up and down > 6. Turning up and down
r. 6. Turning up and down = 6. Turning up and down

		N	Mean Rank	Sum of Ranks
7. Accurate targeting - 7. Accurate targeting	Negative Ranks	5[s]	5.00	25.00
	Positive Ranks	7[t]	7.57	53.00
	Ties	0[u]		
	Total	12		
8. Navigation respond - 8. Navigation respond	Negative Ranks	4[v]	5.50	22.00
	Positive Ranks	5[w]	4.60	23.00
	Ties	3[x]		
	Total	12		
9. Finger fatigue - 9. Finger fatigue	Negative Ranks	2[y]	2.50	5.00
	Positive Ranks	5[z]	4.60	23.00
	Ties	5[aa]		
	Total	12		
10. Wrist fatigue - 10. Wrist fatigue	Negative Ranks	5[ab]	7.50	37.50
	Positive Ranks	6[ac]	4.75	28.50
	Ties	1[ad]		
	Total	12		
11. General comfort - 11. General comfort	Negative Ranks	3[ae]	4.83	14.50
	Positive Ranks	7[af]	5.79	40.50
	Ties	2[ag]		
	Total	12		
12. Overall the input device - 12. Overall the input device	Negative Ranks	5[ah]	5.00	25.00
	Positive Ranks	6[ai]	6.83	41.00
	Ties	1[aj]		
	Total	12		

s. 7. Accurate targeting < 7. Accurate targeting
t. 7. Accurate targeting > 7. Accurate targeting
u. 7. Accurate targeting = 7. Accurate targeting
v. 8. Navigation respond < 8. Navigation respond
w. 8. Navigation respond > 8. Navigation respond
x. 8. Navigation respond = 8. Navigation respond
y. 9. Finger fatigue < 9. Finger fatigue
z. 9. Finger fatigue > 9. Finger fatigue
aa. 9. Finger fatigue = 9. Finger fatigue
ab. 10. Wrist fatigue < 10. Wrist fatigue
ac. 10. Wrist fatigue > 10. Wrist fatigue
ad. 10. Wrist fatigue = 10. Wrist fatigue
ae. 11. General comfort < 11. General comfort
af. 11. General comfort > 11. General comfort
ag. 11. General comfort = 11. General comfort
ah. 12. Overall the input device < 12. Overall the input device
ai. 12. Overall the input device > 12. Overall the input device
aj. 12. Overall the input device = 12. Overall the input device

Table 13. Wilcoxon Signed Ranks test statistics for subjective ratings

Test Statistics[c]

	1. The force required for navigation - 1. The force required for navigation	2. The mental effort required for operating the device - 2. The mental effort required for operating the device	3. Smoothness during navigation - 3. Smoothness during navigation	4. Moving forward - 4. Moving forward	5. Turning left and right - 5. Turning left and right	6. Turning up and down - 6. Turning up and down
Z	-2.705[a]	-.061[a]	-2.539[a]	-1.730[a]	-1.582[a]	-1.294[a]
Asymp. Sig. (2-tailed)	.007*	.951	.011*	.084	.114	.196

a. Based on negative ranks.

c. Wilcoxon Signed Ranks Test

Test Statistics[c]

	7. Accurate targeting - 7. Accurate targeting	8. Navigation respond - 8. Navigation respond	9. Finger fatigue - 9. Finger fatigue	10. Wrist fatigue - 10. Wrist fatigue	11. General comfort - 11. General comfort	12. Overall the input device - 12. Overall the input device
Z	-1.153[a]	-.061[a]	-1.561[a]	-.409[b]	-1.357[a]	-.733[a]
Asymp. Sig. (2-tailed)	.249	.951	.119	.682	.175	.463

a. Based on negative ranks.

b. Based on positive ranks.

c. Wilcoxon Signed Ranks Test

In this analysis, Table 11 displays descriptive statistics of the subjective ratings. It shows the number of scores (N), mean, standard deviation, and minimum and maximum rates for each question. Table 12 shows the number of negative ranks, the number of positive ranks, and the number of ties for each question. Negative ranks are the number of cases where the first variable's rank is less that the second variable's rank, positive ranks are the number of cases where the first variable's rank is greater that the second variable's rank. In the same way, the number of ties is the number of scores with equal ranks. The notes bellow the table demonstrates the direction of these differences.

Table 13 presents the value of Z, which is the standardized normal approximation to the test statistics and the asymptotic two-tailed significance that is estimated based on the normal approximation. The negative sign of Z shows the order in which the devices were compared. This sign is unimportant and does not affect the result, because the standardized normal distribution is symmetrical.

5.4 Qualitative Video Analysis

In light of our user observations, we can conclude a number of facts according to the video analysis of the participants' performance. Most participants preferred to operate the SpaceNavigator with their fingertips, while for the SpaceCat they used their palm to interact with the device. Overall, the participants expressed greater physical control with the SpaceCat due to its larger size compared to SpaceNavigator. On the other hand, SpaceNavigator's dense metal base provided better stability during operation, where the SpaceCat would occasionally slid on the desk.

SpaceNavigator's short controller cap suggests operating the buttons on the sides with the same hand. A number of participants started to operate the SpaceNavigator with one hand, but in all cases they switched to both hands after a while as they found it difficult to perform cognitively demanding tasks including targeting and shooting with fingers of the same hand. SpaceCat's controller handle indicates the need for operation with both hands in order to use the controller buttons that are situated around the device base.

Most participants had difficulties with maintaining skin resistance by holding the top button on SpaceCat's controller handle. In order to eliminate unwanted activation, SpaceCat operates as long as it detects the user's skin resistance. In times of cognitively demanding tasks, participants used to forget to place their palms on the controller handle to maintain the connection with the controller and would become frustrated with unresponsiveness of the device.

According to the participants, SpaceNavigator provided less feedback in comparison with SpaceCat, which made it more difficult to interact with. The stiffly elastic property of the SpaceNavigator allowed for greater activation with less deflection of its controller cap compared to SpaceCat. With regard to targeting, participants found SpaceCat an easier device to comprehend and operate. In addition, the round physical appearance of the SpaceNavigator did not provide enough feedback regarding the correct orientation of the device on the desk. The cable was the main indicator of the device position. As a result, during high demanding tasks it was not recognizable by many of the participants who failed to notice that the device was incorrectly rotated.

6. Discussion

A paired-samples t-test was conducted to compare device performance in Forsaken 3D environment for SpaceCat and SpaceNavigator. Results from the analysis showed that there was not a significant difference in the scores for SpaceCat (M=17.83, SD=7.133) and SpaceNavigator (M=18.50, SD=5.402); t(11)=-.771, p = .457. As a result, the null hypothesis cannot be rejected and no conclusions regarding device performance can be made. A non-significant result does not however mean that the null hypothesis is true. It basically implies that the data are not strong enough to conclude that the null hypothesis is not true. In statistics, type II errors arise while a difference is concluded not to be statistically significant when the null hypothesis is, in fact, false.

Our results show that there is no reason to conclude that for rate control, SpaceCat or SpaceNavigator outperform one another. As a result, the study of device performance in Forsaken game environment is inclusive and it is suggested that additional user studies shall be carried out in order to increase its power.

Results from the analysis of subjective rankings were significant at $p < .05$ level for the force required for navigation (Wilcoxon signed-rank test, $Z = 2.705$, $p = .007$, two-tailed). The medians of SpaceCat and SpaceNavigator were 1.33 and 2.83, respectively. This indicates that SpaceCat requires lower force for operating in navigation compared to SpaceNavigator and supports our H2 hypothesis. A differentiating characteristic of SpaceCat and SpaceNavigator is their elastic property. SpaceCat is a softly elastic input device, while SpaceNavigator is stiffly elastic. A softly elastic device has wider range of movement on its clutch. SpaceCat in particular, is provided with a smooth clutch that enables it to operate using little force. Results from this study show that this characteristic is significant compared to SpaceNavigator.

Smoothness of navigation was also significant with medians of SpaceCat and SpaceNavigator showing 1.83 and 3.17, respectively (Wilcoxon signed-rank test, $Z = 2.705$, $p < .011$, two-tailed). This provides support for our H3 hypothesis that SpaceCat provides smoother navigation compared to SpaceNavigator.

In order to investigate smoothness across different DOFs, data was collected based on ease of navigation for X, Y, and Z-axes of both input devices. Analysis of the results showed that only the navigation across Z-axis for moving forward in the game achieved a significant level (Wilcoxon signed-rank test, $Z = 1.730$, $p < .05$, one-tailed). The medians of SpaceCat and SpaceNavigator were 2.00 and 2.50, respectively. For this variable, the p-value for asymptotic two-tailed test was 0.084. Based on our hypothesis that the SpaceCat is favorably easier to navigate along the Z-axis, the one-tailed p-value would be significant at .05 level ($p = 0.042$). As a result, H3.1 hypothesis was supported by the experiment results.

The rest of variables in the assessment questionnaire did not achieve the significance level to be interpreted in our study. Although the means where favorably supporting SpaceCat on mental effort, moving along X and Y axes, accurate targeting, navigation response, finger fatigue, general comfort, and user preference, they were not significant and thus, no decisions can be made regarding their contribution to the results of this study. Few significant results can be due to the small sample size for our analysis and additional experiments are required in order to increase the power of the analysis and achieve further outcomes.

According findings, SpaceCat is smoother to operate and demands less force for performing navigational tasks, while it provides easier operation functions on Z-axis. According to Sundin and Fjeld [8], a requirement of SpaceCat was to offer elastic suspension for providing rich sensory feedback and results from our study are in favor of that.

7. Future Work

Future research will display additional information regarding the specific aspects of this evaluation that did not achieve a significant level to interpret. Moreover, it will allow for examination of further attributes of the input devices through alternative experimental designs.

By conducting additional tests under the same framework, it will be possible to increase the power of analysis and obtain further significant results. A larger sample size will lead to accurate estimates for additional variables in this experiment. Power analysis is a statistical procedure for verifying the number of extra experiments that are required to enable statistical judgments and can be used in order to estimate the number of necessary test cases for particular variables of this study. We suggest that five more tests should be conducted under the same framework in order to reveal additional significant results.

Subjects where videotaped during their trials. Additional video analysis of their behavior while operating with each device can uncover supplementary details about the SpaceCat and the SpaceNavigator. Future qualitative observations will expand the range of knowledge presented in this study and will account for the qualities of user experience with both devices.

Additionally, we suggest a supplementary experimental design to assess game expertise associations and learning effects across the two softly elastic and stiffly elastic devices for navigation in 3D environments. In a future research it would be interesting to incorporate novice and experienced Forsaken 3D game players to examine their performance with SpaceCat and SpaceNavigator and evaluate their user experience.

8. Conclusion

In this study we compared two elastic 6 DOF inputs for rate control in a 3D game environment. We evaluated the SpaceCat with softly elastic stiffness compared to the SpaceNavigator with stiffly elastic suspension. We assessed usability properties of the input devices by evaluating their particular aspects of physical operation, mental effort, accuracy and speed, fatigue and comfort, and overall usability through a subjective ratings questionnaire.

Results from our study showed that SpaceNavigator provides smoother navigation compared to SpaceNavigator and it is comparatively more robust on the Z-axis. Our findings proved that SpaceCat with softly elastic suspension requires less force for performing navigational tasks. Much of our endeavors for explaining differences between SpaceCat and SpaceNavigator remained inconclusive due to inability in interpreting insignificant findings. An explanation for the little significant results in this study is the small number of test subjects in the experiment and additional tests may result in supplementary findings.

Furthermore, results did not provide evidence in favor of SpaceNavigator for any of the studied variables. As a result, we believe that SpaceCat can perform navigational tasks equally well compared to SpaceNavigator while it provides smoother navigational experience.

Quantitative analysis of the video recordings revealed that the SpaceCat allowed for greater control as a result of its physical size in comparison with SpaceNavigator, while the SpaceNavigator provided greater stability on the surface due to its heavy base. SpaceCat showed more frustrating to handle cognitively demanding tasks before the participants would get used to maintaining contact with its upper button. On the other hand, while the design of SpaceCat provided clear information about its positioning on the desk, SpaceNavigator lacked sufficient feedback regarding its accurate orientation, which occasionally resulted in incorrect operation across the X and Y axes of the controller.

References

[1] Natapov, D., Castellucci, S. J., and MacKenzie, *ISO 9241-9 evaluation of video game controllers.* Proceedings of Graphics Interface 2009. Toronto: Canadian Information Processing Society, 2009: p. 223-230.

[2] Franjcic Z., A., M., Sundin, M., and Fjeld, M., *Softly Elastic Device for 6DOF Input.* Proceedings of ACM NordiCHI 2008, 2008.

[3] De Sena, A., and Moschini, D., *A Mouse Evaluation Framework for Fitts' Test.* Dipartimento di Informatica, Universita` di Verona, 2006.

[4] Dallal, G.E. *Nonparametric Statistics*. 2008 16 July [cited 2010 15 August]; Available from: http://www.jerrydallal.com/LHSP/npar.htm.

[5] StatSoft. *Nonparametric Statistics*. [Electronic Statistics Textbook] 2010; Available from: http://www.statsoft.com/textbook/nonparametric-statistics/.

[6] Hamilton, M.J. *Non-Parametric Univariate Tests: Wilcoxon Signed Rank Test*. 2004 [cited 2010 15 August]; Available from: http://www.unm.edu/~marcusj/WilcoxonSR.pdf.

[7] Shaw, C., et al., *Wilcoxon Signed Rank Test.* Journal of Clinical Nursing, 2000. **9**: p. 574-584.

[8] Sundin, M., Fjeld, M., *Softly Elastic 6 DOF Input.* International Journal of Human-Computer Interaction, 2009. 25(7).

[9] Slater, M., and Davison, A., *Liberation from Flatland: 3D interaction based on the DesktopBat.* In F. H. Post & W. Barth (Eds.), Eurographics 91, 1991: p. 209–221.

[10] Zhai, S., *Human performance in six degree of freedom input control.* Toronto, Canada: University of Toronto, 1995.

[11] Froehlich, B., Hochstrate, J., Skuk, V., and Huckauf, A., *The globefish and the globemouse: two new six degree of freedom input devices for graphics applications.* In CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems, 2006: p. 191-199.

[12] Eng, K., *Miniaturized Human 3D Motion Input.* In: Asai, Kikuo. Human computer interaction: new developments. Wien, AU, 2008: p. 231-242.

[13] Li, X., *Improved Transfer Function of Spacecat.* (Internal Tech. Rep). Axiglaze AB, 2001.

[15] ISO, *9241-9 Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9 - Requirements for non-keyboard input devices.* International Organisation for Standardisation, 2000.

[16] Field, M., *NASA TLX*. 2 ed. 2003: NASA Ames Research Center.

[17] Porteous, M., Kirakowski, J., and Corbett, M., *SUMI User Handbook*. 2 ed. 1998: Human Factors Research Group, University College Cork, Ireland.

[18] Brooke, J., *SUS: A quick and dirty usability scale*. In P. Jordan, B. Thomas, B. Weerdmeester, & I. McClelland (Eds.). 1996, London: Taylor & Francis.

Appendix A: Demographics Questionnaire

How often do you play video games
Seldom 1…..... 2 …..... 3 …..... 4 …..... 5 …..... Very often

Experience with 6 DOF input devices
No experience 1…..... 2 …..... 3 …..... 4 …..... 5 …..... Experienced

Age
….…

Gender
Male…..... Female ….…

Hand orientation
Right handed …..... Left handed …....

Appendix B: Subjective Ratings Questionnaire

1. The force required for navigation was
Very low 1….... 2 ….... 3 ….... 4 ….... 5 ….... Very high

2. The mental effort required for operating the device was
Very low 1….... 2 ….... 3 ….... 4 ….... 5 ….... Very high

3. Smoothness during navigation was
Very smooth 1….... 2 ….... 3 ….... 4 ….... 5 ….... Very rough

4. Moving forward was
Very easy 1….... 2 ….... 3 ….... 4 ….... 5 ….... Very difficult

5. Turning left and right was
Very easy 1….... 2 ….... 3 ….... 4 ….... 5 ….... Very difficult

6. Turning up and down was
Very easy 1….... 2 ….... 3 ….... 4 ….... 5 ….... Very difficult

7. Accurate targeting was
Very easy 1….... 2 ….... 3 ….... 4 ….... 5 ….... Very difficult

8. Navigation respond was
Responsive (quick) 1….... 2 ….... 3 ….... 4 ….... 5 ….... Poorly responsive

9. Finger fatigue
None 1….... 2 ….... 3 ….... 4 ….... 5 ….... Very high

10. Wrist fatigue
None 1….... 2 ….... 3 ….... 4 ….... 5 ….... Very high

11. General comfort
Comfortable 1….... 2 ….... 3 ….... 4 ….... 5 ….... Very uncomfortable

12. Overall the input device was
Very easy to use 1….... 2 ….... 3 ….... 4 ….... 5 ….... Very difficult to use