Talking Games: An Empirical Study of Speech-based Cursor Control Mechanisms

Except where reference is made to the work of others, the work described in this dissertation is my own or was done in collaboration with my advisory committee. This dissertation does not include proprietary or classified information.

David Thornton

Certificate of Approval:

N. Hari Narayanan
Professor
Department of Computer Science and Software Engineering

Juan E. Gilbert, Chair
Associate Professor
Department of Computer Science and Software Engineering

Cheryl D. Seals
Assistant Professor
Department of Computer Science and Software Engineering

George T. Flowers
Interim Dean
Graduate School

Talking Games: An Empirical Study of Speech-based Cursor Control Mechanisms

David Thornton

A Dissertation

Submitted to

the Graduate Faculty of

Auburn University

in Partial Fulfillment of the

Requirements for the

Degree of

Doctor of Philosophy

Auburn, Alabama
December 19, 2008

Talking Games: An Empirical Study of Speech-based Cursor Control Mechanisms

David Thornton

_____

Signature of Author

_____

Date of Graduation

Dissertation Abstract

Talking Games: An Empirical Study of Speech-based Cursor Control Mechanisms

David Thornton

Doctor of Philosophy, December 19, 2008
(M.S., Jacksonville State University, 2003)
(B.S., Jacksonville State University, 2001)

86 Typed Pages

Directed by Juan Gilbert

This document describes a study of speech-based cursor control mechanisms along with a new proposed approach called NameTags. This research is intended to provide empirical user data to inform the design of future systems where one or more of the following conditions are present: real-time demands, very small targets, and moving targets. One such application of this research is in the area of video games, where subjects are often required to make quick selections on numerous small, moving objects. These findings also have implications for physically impaired subjects whose primary or only control modality is speech.

Acknowledgments

It is a pleasure to thank the many people who made this thesis possible.

Firstly, I wish to thank my thesis advisor, Juan Gilbert, for his support and encouragement, for helping to define the scope of the project, and for assisting in publications and travel aid. I also wish to thank the rest of my thesis committee, Hari Narayanan and Cheryl Seals, for their advice in refining the experiment. My thanks to my fellow HCCL lab members Shaun Gittens, Yolanda McMillan, Kenneth Rouse, Idongesit MkPong-Ruffin, and Vincent Cross for their help in reviewing the proposal and thesis.

I am indebted to many of my Jacksonville State colleagues for providing student volunteers, including Jan Case, David Dempsey, Aaron Garrett, Karen Myers, Daniel Smith, Monica Trifas, and Audria White.

Lastly, my heartfelt thanks to my fiancé Chelsea, whose loving support kept me burning the midnight oil.

Table of Contents

List of Figures

Chapter 1

Introduction and Background

1.1 Introduction

Cursor control is a fundamental aspect of modern computer interfaces, especially those whose interaction style is direction manipulation. While the mouse has remained the prevailing cursor control device since its commercial introduction in the early 1980s, it is not always accessible or optimal. Many physically impaired users may prefer or require speech-based interfaces for cursor control. Even users who are not physically impaired may find speech-based cursor control useful or preferable in certain situations. Still others may prefer a multimodal approach, utilizing speech and the mouse (or other pointing device) together to realize cursor control.

Once the target object is selected, a user may manipulate that object. When utilizing cursor control, a user may drag the object or point to a destination location (such as the Windows Recycle Bin or a subfolder). Improving speech-based cursor control for object selection and spatial navigation is the focus of this research. The following section provides a background of the area pertinent to this research, Speech-based Cursor Control.

1.2 Background

Speech as an interface modality is nothing new. However, comprehensive PC control via speech is a greater challenge. Work by Oviatt [25] found that text entry and cursor control were two key elements that must be supported by such a speech-enabled system. The first is easier than the second, since recent advances in speech recognition systems have pushed

recognition accuracy to 98% under controlled environments, making them powerful tools for dictation [13]. Cursor control, on the other hand, is not as naturally mapped to speech. In fact, if such a system is poorly designed, the inherent delay of speech and recognition error rates can make pointing at an object on-screen a tedious task. Work by Sears [31] found that users who employed dictation software spent a third of their time navigating to the target location. Therefore, any improvement of cursor control performance will have a significant impact on total task completion time.

Some research in this area has focused on "target-based" pointing; that is, selecting a predefined labeled point on-screen [6, 17, 18, 19]. Others employ a "direction-based" approach, wherein users give directional commands relative to the current cursor position [14, 16, 24, 20, 21, 22].

Direction-based cursor control can be subdivided into two types: discrete and continuous. Discrete cursor control allows a user to say commands like "Move left 2 inches". Research shows that this approach becomes less effective when the cursor is further from the goal location [20]. Continuous cursor control requires the user to specify a direction (and perhaps a speed) which "drives" the cursor to the desired location. Once the cursor reaches the goal, the user says "stop" or a similar command to stop the cursor's movement. This approach leads to 3 types of delay, however:

- Delay associated with the user perceiving that the cursor has reached the destination
- Delay associated with the user uttering the stop command
- Delay associated with the speech recognition engine's processing of the user's command

As a result, users must predict where the cursor will stop and issue the "stop" command before the cursor reaches the goal. Karimullah [20] proposed an approach which displays a "ghost cursor" along with the actual cursor which indicates where the cursor will end up if the user says "stop" at the given moment.

A variation on this theme is the use of "non-speech", or verbal sounds (such as humming) to move the cursor [14, 33, 16]. Harada's Vocal Joystick allows users to make vowel sounds to choose direction ("ee" for left, "ahh" for right, for example) and volume to control the cursor speed (louder for faster movement). Mihara [24] proposed a hybrid approach called the Migratory Cursor which employed non-speech and a mix of both direction-based and target-based cursor control.

Feng [7, 8] studied speech-based navigation in the context of dictation systems. Feng's work sought to improve error prevention and recovery. Studies found that while the failure rates are not significantly different between direction-based and target-based approaches, direction-based tasks were more likely to fail because they required a longer sequence of commands to reach a target location.

This research focuses on target-based cursor control, which will now be discussed in more detail.

Target-based cursor control "jumps" the cursor directly to a target location. Kamel and Landay [17, 18, 19], for instance, developed speech-based drawing tools for the blind, employing a 3x3 grid overlay. Dai et al. [6] built on this work, applying it to sighted users. The aforementioned grid overlay is a target-based solution, with each cell of the grid assigned a number from 1 to 9. The user utters the desired number to refer to a particular point on-screen. A smaller grid is then shown in place of the chosen cell. The user recursively

issues commands until the target object (and only the target object) is contained within the highlighted area, then says "click" or a similar command to simulate a mouse click on that spot. This grid-based approach allows users to specify any point on-screen, though very small objects may require up to 7 commands to isolate on a 1024 x 768 resolution screen. A screenshot of the grid cursor is shown in Figure 1.1.

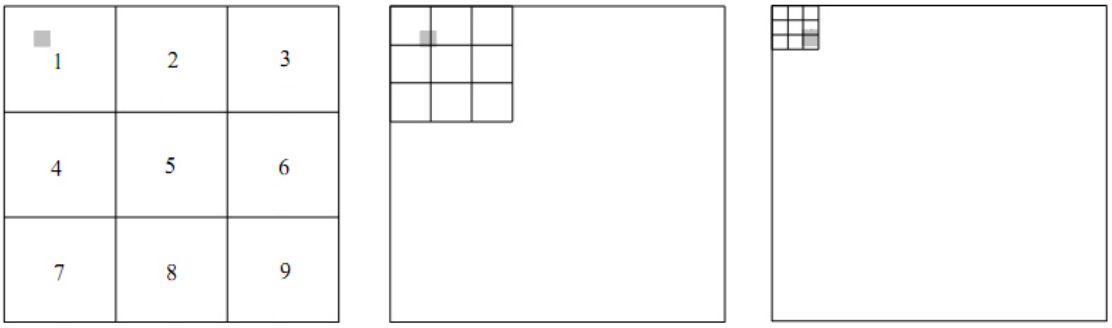

Figure 1.1: Screenshot of grid cursor control mechanism. In this example, the user wants to select the gray object. The user says "one", and the grid shrinks to that cell location. Again, the user says "one", followed by "select nine".

Christian et al. [3] investigated using target-based cursor control to navigate a web browser, both by having the user speak the link's text, and by labeling each link with a number. There was no significant difference between the two approaches in completion time or subjective satisfaction, though users anecdotally preferred to speak the actual link's text, rather than the numbered label.

Speech systems enable individuals with physical impairments to obtain employment in fields that would otherwise be unattainable. In fact, systems like Jeffrey Gray's SpeechClipse [12], a speech-enabled version of the Eclipse programming environment, provide shortcuts that may improve performance and even subjective satisfaction for users without physical

impairments. While this research is conducted with subjects without any documented disabilities, its findings may provide interesting comparative data.

Multimodal systems have become the focus of increased research in the last few years, though they have been studied as far back as the 1980's, when Hauptmann [15] found encouraging results about users' readiness to mix speech and gesture. Multimodal research seeks to incorporate new and innovative control mechanisms with the standard joystick or keyboard-and-mouse control paradigms. Research projects such as [34] and [32] have explored the use of eye gaze to direct the cursor, while the Nintendo Wii, which features motion-based controls mixed with classic control mechanisms, has sold over 9 million units [29]. Perakakis's [28] studies with multimodal input modes on PDAs found that users tend to focus on the most efficient input mode. This research, however, has a special interest in the domain of video games, where users may employ an input modality that is more entertaining over one that is marginally more efficient.

The next section describes the current problem followed by the proposed approach.

1.3 Problem Description

The aforementioned research has been primarily concerned with applications for drawing shapes, navigating web pages, or dictation. Even in those studies which have implemented their experiments in the form of a game [2, 35], where the user is not constrained by time and the target location is stationary. In many systems, however, the target object may not be stationary, but may instead be moving in a predictable pattern (or even randomly). Further, the object to be selected may be only a few pixels wide and/or long, making grid-based solutions less amenable to real-time situations. Also, many programs

display selectable objects which have no textual label. In this case, the numbering system proposed by Christian [3] might be appropriate for small numbers of selectable objects on-screen. As the number of objects increase, the recognition error rate with integers will increase as labels such as "fourteen" and "forty" sound too much alike.

In such cases, the NameTags control mechanism provides a text "handle" for these objects. The next section describes the use of NameTags.

1.4 NameTags

NameTags is a label-based cursor control mechanism that employs common names for a label bank. Users would be given the ability to toggle a "name tag" option that would label each selectable unit on the screen, giving players a verbal "handle" by which to refer to them (Figure 1.2). The use of this control mechanism for selecting objects is straightforward. A user could say, for instance, "Select Bob, David, and Susan". This solution balances the need to distinguish between objects without overloading the screen with details (since it can be toggled).

Using this mechanism for moving a unit to a given point on-screen is similarly straightforward. A user might issue a command like "Move to Kevin", or "Follow Susan", using other objects as "waypoints" or "leaders". This mechanism does not, however, provide a means of pointing to an arbitrary point on-screen (a point where there is no selectable object) like the Grid Cursor mechanism does. Therefore, NameTags would have to be coupled with another mechanism to enable comprehensive PC control.

The use of names in this approach is not simply arbitrary. Labeling units with numbers or letters may lead to a greater chance of recognition errors, since several numbers and

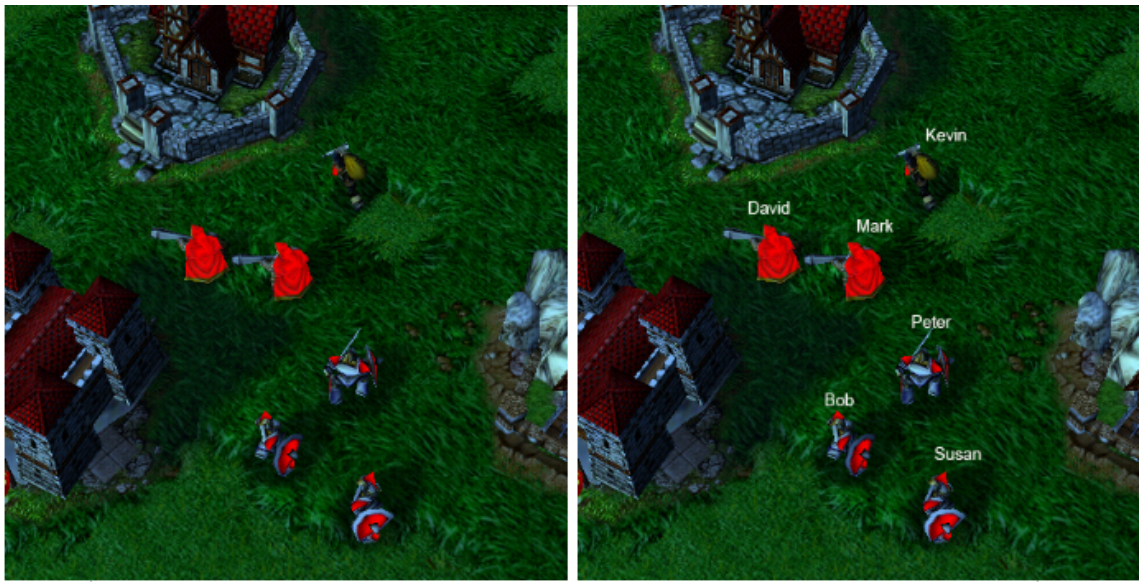

Figure 1.2: Left, a screenshot from Warcraft 3, displaying several units of different types. Right, the same image but with units "tagged" with names.

letters sound similar (e.g. letters like "B", "V", and "D" and numbers like "fourteen" and "forty"), while naming units with labels like "Kevin" and "Mark" may provide more distinct phonemes. Further, if the selectable units are meant to be represented as anthropomorphic (as in this research), using names may be preferable for the player. Naturally, as the number of objects on-screen increases, the recognition error rate is expected to increase. It is possible that a combination of words and numbers yields the optimal balance of aural distinctiveness and brevity, but that is a subject for future work.

1.5 Contributions

This research is concerned with improving speech-based cursor control by comparing three control mechanisms: the joystick, grid-based cursor control, and a proposed label-based mechanism called NameTags. In each experiment, subjects employed these mechanisms (in isolation and/or combined) to select and move objects on-screen. Subject performance and preference (both reported and actual) were measured and analyzed. This research is intended to provide empirical user data as well as interpretations of that data in order to inform the design of future systems where one or more of the following conditions are present: real-time demands, very small targets, and moving targets.

Chapter 2

Joystick vs. Grid Cursor Experiment

This chapter details an experiment comparing two different mechanisms for cursor control - the standard joystick (a.k.a. "gamepad") and the speech-based Grid Cursor. This was the first of three experiments intended to compare the performance and usability of speech-based cursor control methods with the current prevailing modality.

As the joystick is the predominant cursor control device in video games, this experiment sought to find a model for predicting joystick performance versus a speech-based mechanism. Fitts' law was chosen because it is a proven model for pointing devices.

Fitts' Law [9] is a model of psychomotor behavior developed by Fitts in 1964. It describes the time it takes for a human user to acquire a target using a manual input device. Although it was originally formulated for only one-dimensional movement, researchers have found it quite robust even with two-dimensional tasks.

2.1 Experimental Design

2.1.1 Hypotheses

This experiment intended to answer the following questions:

- How does the performance of these two mechanisms compare?
- Do these mechanisms adhere to Fitts' Law for movement time?
- Which mechanism do subjects prefer?

The following hypotheses were tested:

- The joystick will outperform the grid cursor in completion time.
- The joystick will adhere tightly to Fitts' Law (greater than .90 Pearson's correlation).
- The grid cursor will adhere tightly to Fitts' Law (greater than .80 Pearson's correlation) with modifications to the formula.

2.1.2 Subjects

There were 40 subjects for this experiment, 17 females and 23 males, and their average age was 25.3. These subjects were selected from the student body and faculty of Jacksonville State University. The subjects needed no specialized knowledge or experience to participate in this research. Subjects' computer use and game play per week are shown in Figure 2.1. As shown, no subjects reported "None" for computer use. Each category for game play was also well represented.

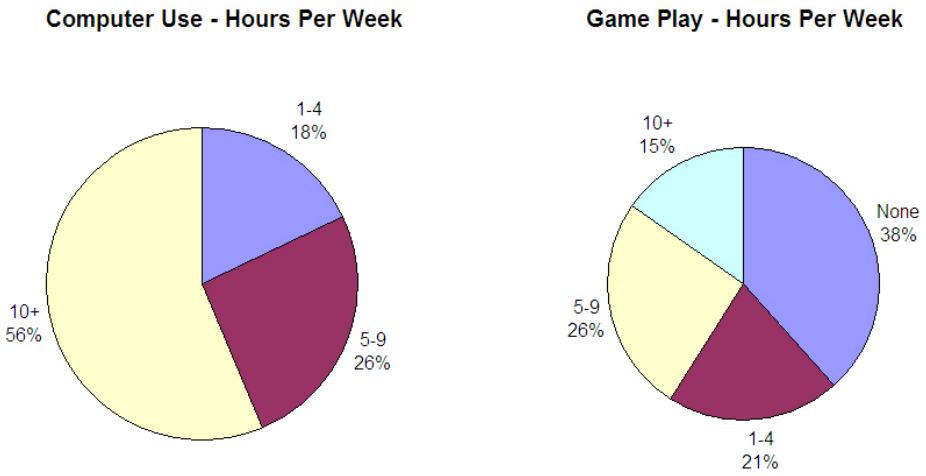

Figure 2.1: Subjects' computer use and game play in hours per week.

2.1.3 Setup

Subjects were asked to play a simple game in which their only goal was to select the stationary object on-screen as quickly as possible. In order to test both mechanisms, a within-subjects design was chosen such that half of the subjects employed the grid cursor first, followed by the joystick, while the other half did the reverse. Each subject performed 50 selection tasks with each mechanism. Subjects were given basic instructions of how to play the game, then were allowed to practice the given mechanism for 3 trial tasks before beginning the game proper.

The speech control component of the game was implemented in CloudGarden [4], a version of the Java Speech API. The game itself was programmed in Game Maker [11], a simple two-dimensional game engine. Communication between the speech component and the game was facilitated by keystroke messages generated by the the Java Robot class.

Figure 2.2 depicts two screenshots of the game. The joystick version of the game is shown on the left, while the Grid Cursor is shown on the right.

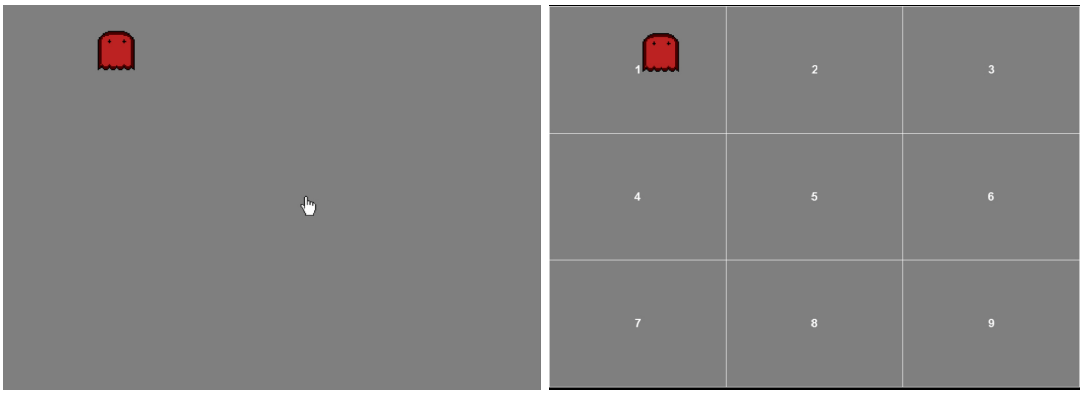

Figure 2.2: Joystick version (left), and Grid Cursor version (right).

In order to mitigate speech recognition errors, subjects completed the introductory session of the Microsoft Speech Recognition Training Wizard before using the grid cursor.

When using the grid cursor, subjects could utter the following commands:

- To shrink grid: "< 1 - 9 >" Example command: "four"
- To select objects: "Select < 1 - 9 >" Example command: "select two"
- To go back: "go back" or "back"

Selecting an object would occur as shown in Figure 2.3.

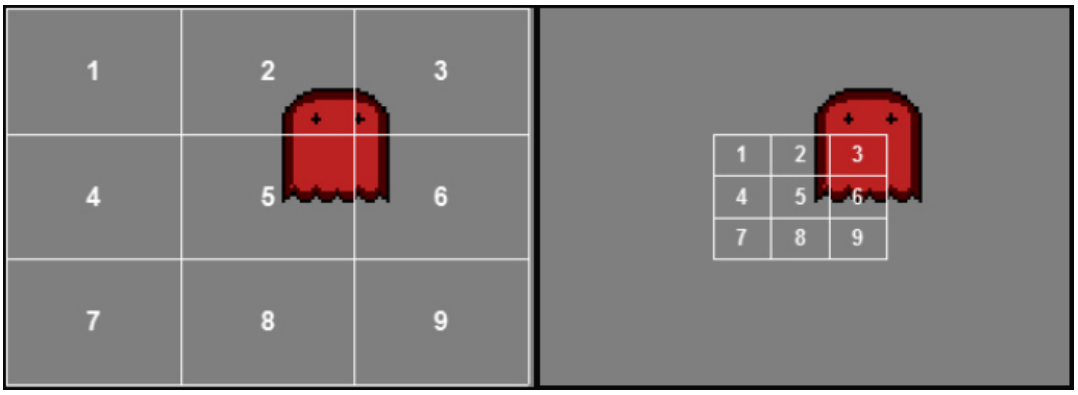

Figure 2.3: The subject sees that the object is mostly inside of cell 5 and utters the command "five". The grid then shrinks to the size and location of cell 5. Now that one of the cells is entirely inside of the object to be selected, the subject can issue the command "select three".

When using the joystick, the following controls were available:

- To move the cursor: Move the directional pad
- To select objects: Press button "A"

After pilot testing, the movement speed for the joystick was set at 210 pixels/second. The joystick cursor was reset to the center of the screen at the beginning of each game level.

2.1.4 Object Size and Placement

The objects to be selected were randomly placed on-screen and were randomly sized between 20 pixels and 100 pixels square on a 1024x768, 17" screen. The size ranges were derived from a popular turn-based strategy game, Galactic Civilizations [10]. This game genre was chosen because selectable objects remain stationary, which is necessary to test the Fitts' Law adherence. Testing with stationary objects also provides a baseline to compare with for future experimentation with moving objects.

2.1.5 Questionnaire

After the game was completed, subjects were asked to fill out a questionnaire which asked subjects for the following information:

- Age
- Gender
- How much time subject uses a computer per week
- How much time subject plays video games per week

Subjects were also asked to rate their subjective impressions on each control mechanism. An excerpt is shown in Figure 2.4.

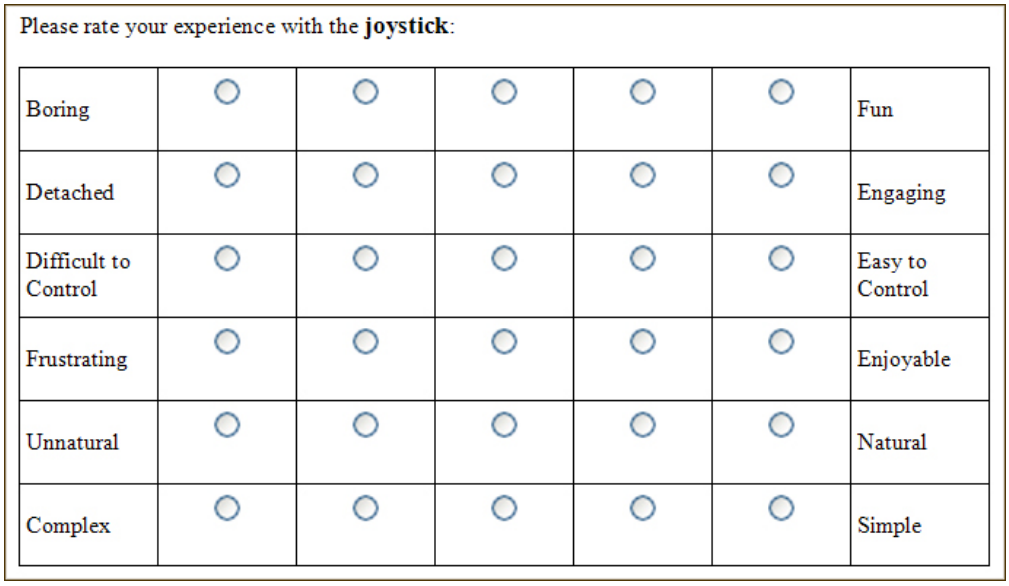

Figure 2.4: An excerpt from the post-experiment questionnaire.

2.2 Results

2.2.1 Performance Comparison

The joystick outperformed the grid cursor for completion time substantially, with a mean completion time of 2.178 seconds, versus the grid cursor with 7.634. Further detail is shown in Figure 2.5.

The joystick is the clear winner in regard to completion time. It is important to note, however, that this experiment enforced a requirement on the grid cursor which, if relaxed, would result in faster average performance. This requirement is illustrated in Figure 2.6.

This requirement was enforced in order to simulate the pointing time necessary for the worst-case scenario: a screen filled with numerous, tightly packed, selectable objects. In such a situation, the subject would be forced to employ the grid cursor as they did in the

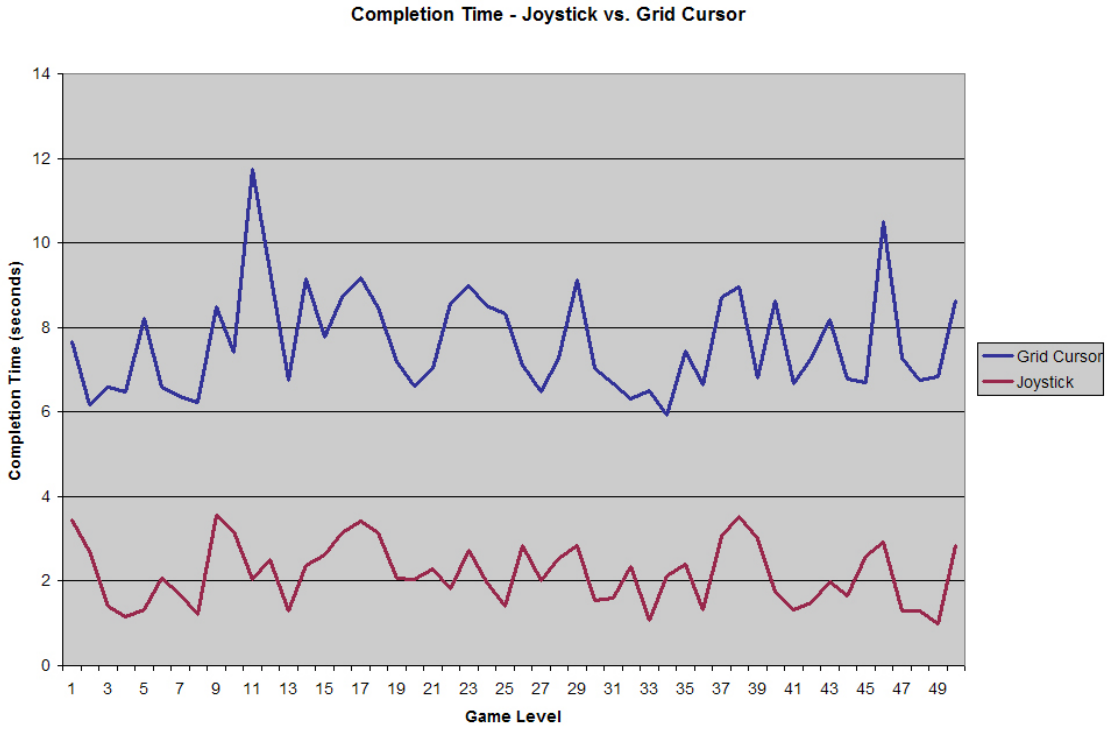

Figure 2.5: Completion times for the grid cursor (blue) and joystick (red).

experiment. In order to match as closely as possible the default Fitts' Law experimental setup, however, only the target selectable object was displayed on-screen.

One-way ANOVA tests on performance data found a wide, statistically significant gap in performance between males and females (alpha=0.05, F=10.30, P=0.003). Male subjects had a mean selection time of 1.903 versus female subjects with 2.533. This may be attributable to hours spent playing video games, because joystick performance was also highly dependent on the time subjects spent playing games. Details of this are shown in Figure 2.7. A comparison of game play per week by gender is shown in Figure 2.8.

By looking at both of these factors together, it seems apparent that gender has little or no significance in joystick performance. Instead, the males in our sample simply spend

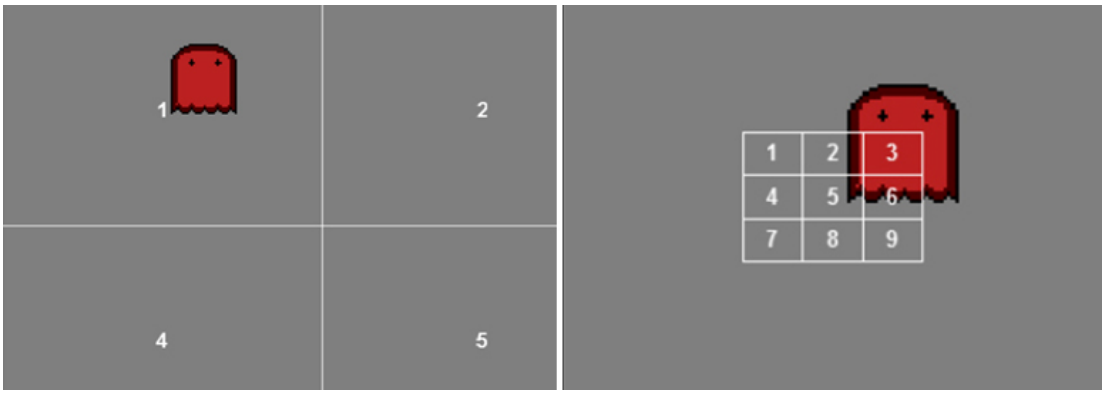

Figure 2.6: Since the object to select in the left picture is the only object inside cell 1, a subject arguably ought to be able to say "select one" to select that object. In this experiment, however, subjects were required to shrink the grid until a cell was completely within the object, as shown on the right.

more time playing games per week than the females. The interaction of these two factors is shown in Figure 2.9.

Speech performance, on the other hand, was highly dependent on time spent using the computer (F=3.61, P=0.037), but not on time spent playing video games.

2.2.2 Adherence to Fitts' Law

This experiment employed the Shannon Formulation of Fitts' Law, as shown.

$$T = a + b \log_2(\frac{D}{W} + 1)$$

The Shannon formulation of Fitts' Law was chosen because the predicted movement time is always nonnegative. In this formula, T refers to movement time, D is the distance from the cursor to the center of the object, and W is the width of the object. The constants a and b can be determined via linear regression, and they represent the start/stop time of the device and the device's inherent movement speed. Since Fitts' Law was originally

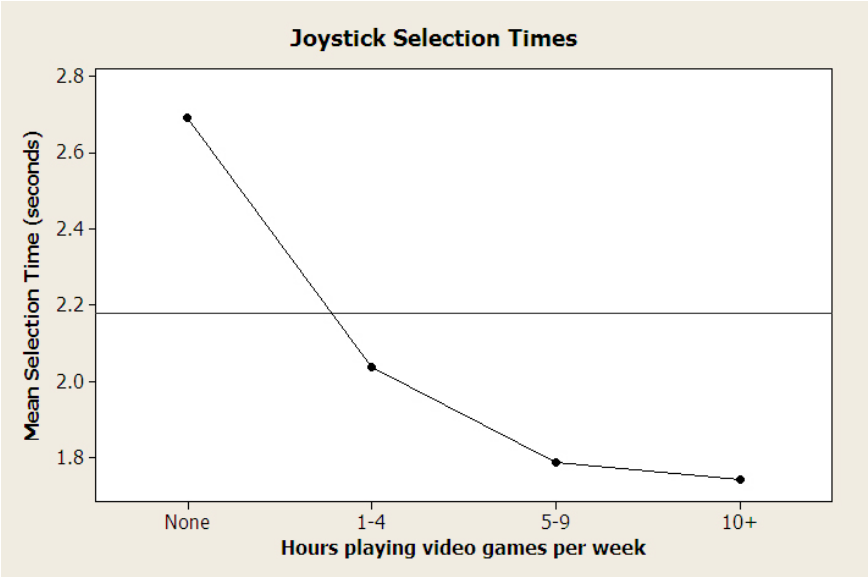

Figure 2.7: Main effects plot for joystick selection time and hours spent playing video games per week.

intended for one-dimensional movement, some modifications must be made. Mackenzie [23] compared several variations of the Fitts' Law formula for two-dimensional tasks, and this experiment employs the following:

- Distance is determined by the Euclidean distance from the cursor to the object's center.
- Width is determined by the greater of the height or the width (this is actually moot in our experiment since the objects to be selected were square).

The joystick adhered tightly to Fitts' law with a Pearson's correlation of 0.912, as shown in Figure 2.10.

The regression equation is Selection Time $= 0.239 + 0.794 \log_2(\frac{D}{W} + 1)$.

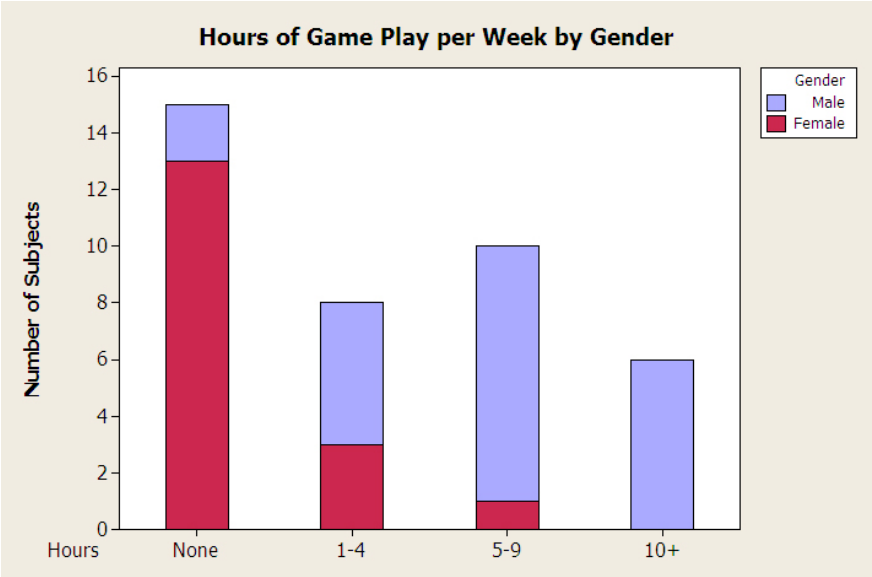

Figure 2.8: Subjects' reported hours spent playing video games per week.

When testing the adherence of the grid cursor to Fitts' Law, it seemed intuitive that distance was not a useful parameter because the grid cursor covers the entire screen. Size seemed to be the primary parameter for determining selection time. To test this, the original Shannon formulation and two variations were tested against observed selection time. They are shown.

- Selection Time $= a + b \log_2(\frac{D}{W} + 1)$ Original Formulation
- Selection Time $= a + b \log_2(D + 1)$ Distance Alone
- Selection Time $= a + b \log_2(\frac{1}{W} + 1)$ Size Alone

The fit of each of these formulations is shown in tabular and graphical format in Table 2.1 and Figures 2.11, 2.12, and 2.13. The tightest fit was found when considering size as the only factor, followed by the original Shannon formulation and the "distance only" formulation.

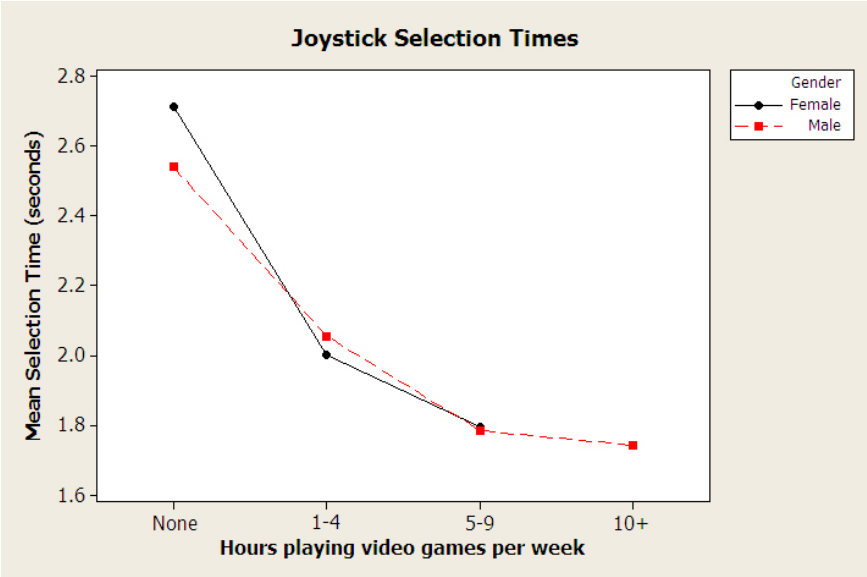

Figure 2.9: Interaction between gender and time spent playing video games for joystick selection time.

Table 2.1: Comparison of 3 variations of Shannon formulation.

Model	Correlation	Regression Equation
Distance Alone	.240	Selection Time = 4.54 + 0.391 ID Distance
Original	.588	Selection Time = 5.58 + 0.841 ID Both
Size Alone	.696	Selection Time = 5.87 + 64.2 ID Size

2.2.3 Subject Preference

As portrayed in the excerpt of the post-experiment questionnaire, 40 subjects were asked to rate their experience with the joystick and the grid cursor based on the following 6 bipolar semantic categories:

- Boring or Fun
- Detached or Engaging
- Difficult to Control or Easy to Control

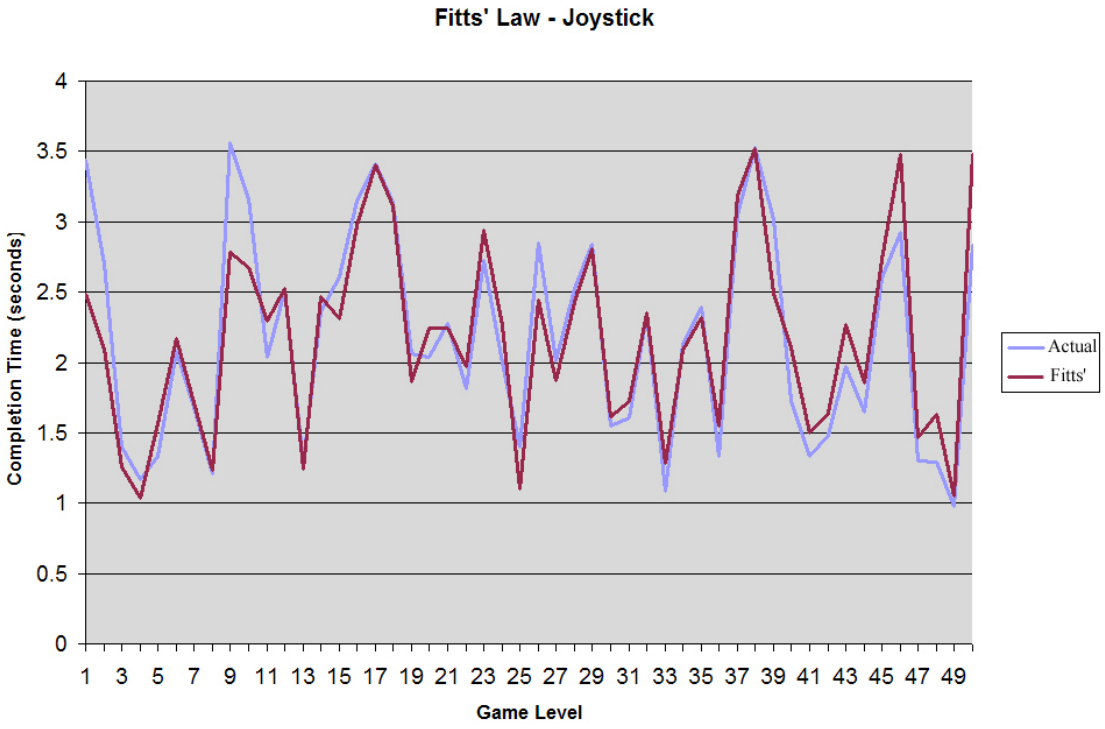

Figure 2.10: The actual selection time (in blue) and the Fitts' Law predicted selection time (in red).

- Frustrating or Enjoyable
- Unnatural or Natural
- Complex or Simple

The grid cursor received a higher mean score than the joystick for 4 out of the 6 categories. Of the 6 categories tested, however, only 2 were statistically significant with an alpha value of .05. Subjects rated the joystick simpler (P=.000), while the grid cursor was rated as more engaging (P=.001). The results are shown in Figure 2.14.

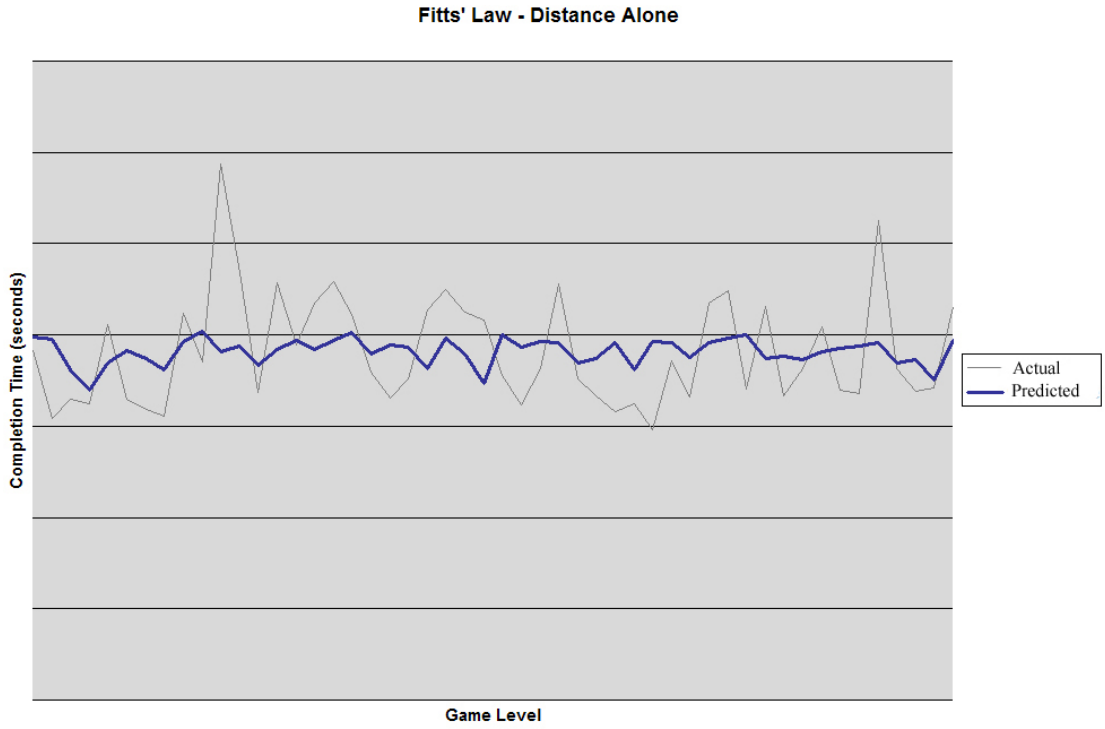

Figure 2.11: Using distance as the only parameter for selection time provided very poor prediction (correlation of 0.240). Observed time is in gray, while predicted time is in blue.

Female subjects rated the joystick as more engaging (Mean=3.71, Std. Dev.=0.772) than male subjects did (Mean=2.81, Std. Dev.=1.09) in one-way ANOVA tests (alpha=0.05, F=8.03, P=0.007). Grid cursor ratings were reversed(F=4.31, P=0.045), with males rating the grid cursor as more engaging (Mean=4.23, Std. Dev.=0.685) than female subjects did (Mean=3.65, Std. Dev.=1.06).

2.2.4 Conclusion

This section provides a brief summary of the results of this experiment.

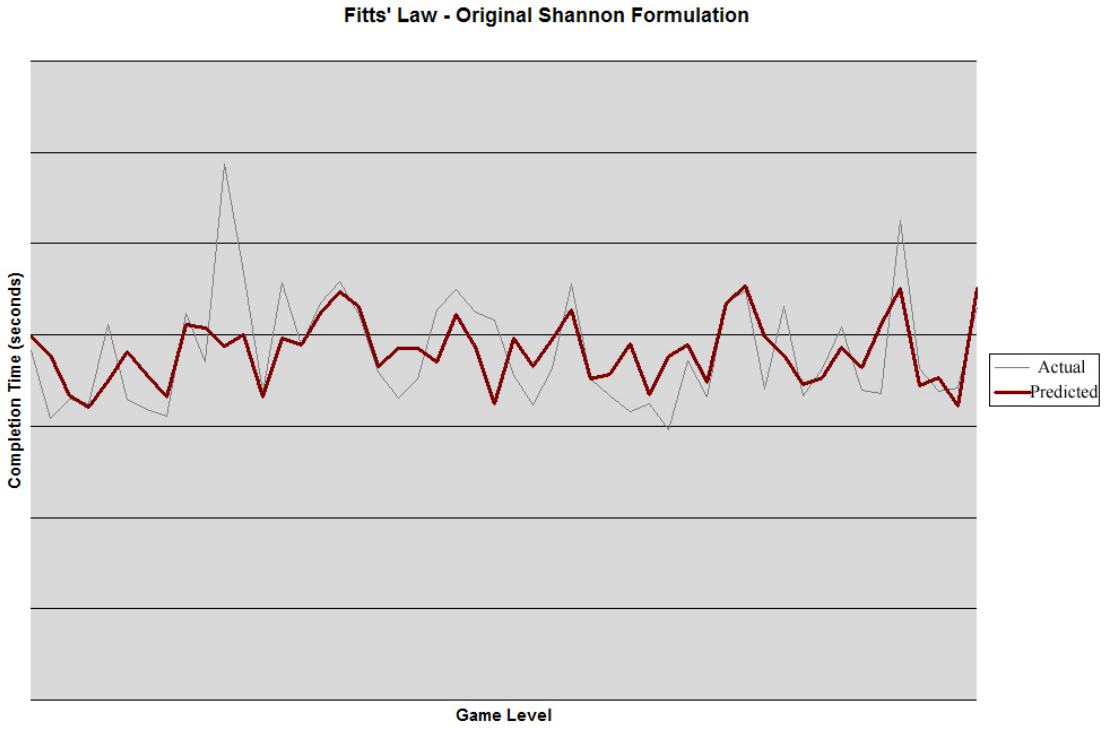

Figure 2.12: Using both distance and size to predict selection time performed much better (correlation of 0.588). Observed time is in gray, while predicted time is in red.

The joystick outperformed the grid cursor in mean selection time by a factor of 3.5, though it should be noted that this is a worst-case performance. Even under optimal circumstances, however, the joystick is the clear winner for performance. Therefore, the grid cursor would be useful only when the joystick is not accessible or when time is not a factor.

Unsurprisingly, joystick performance was highly dependent on the amount of time subjects played video games per week. Speech performance was highly dependent on the amount of time subjects spent using the computer.

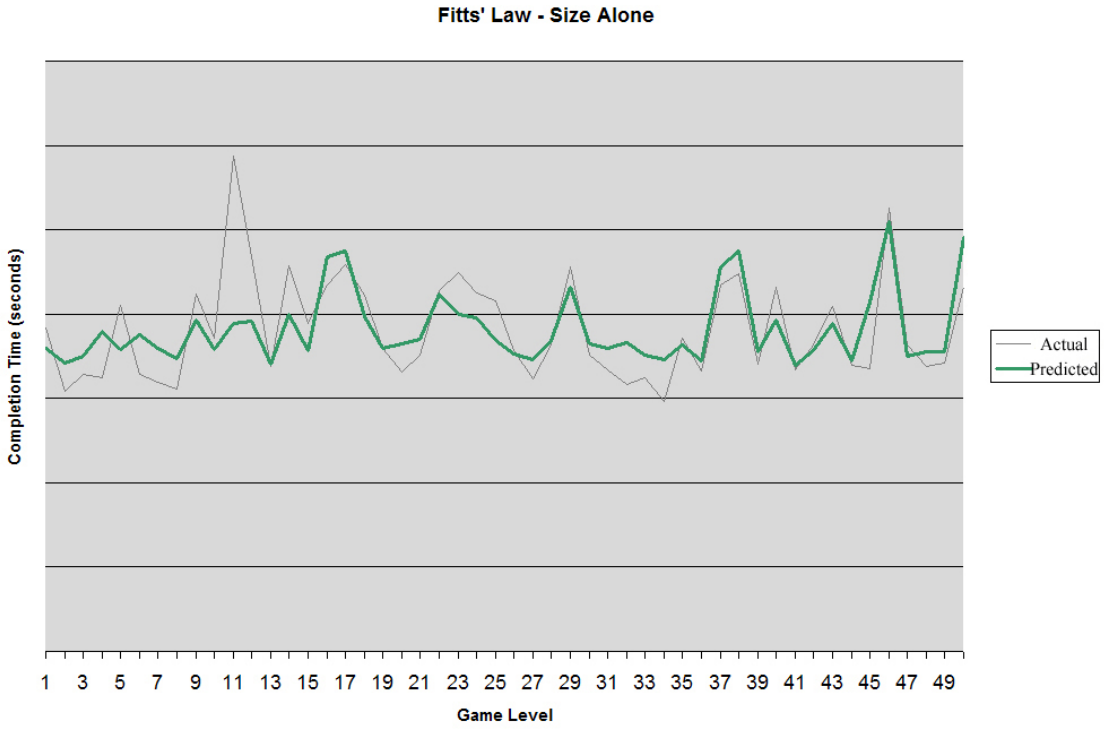

Figure 2.13: Using size as the only parameter for selection time provided the most accurate prediction (correlation of 0.696), though not as accurate as the joystick (correlation of 0.912). Observed time is in gray, while predicted time is in green.

The joystick adhered tightly to Fitts' Law, with a Pearson's correlation of 0.912. The grid cursor did not adhere as tightly as the joystick, though it provided the tightest fit when using size as the only input parameter (correlation of 0.696).

When subject preference was tested, the grid cursor received a higher mean score than the joystick for 4 out of the 6 categories. The joystick was rated simpler, while the grid cursor was rated as more engaging. This suggests that even though the grid cursor is the inferior performer in regard to completion time, subjects may still prefer to use this mechanism over the joystick in some situations.

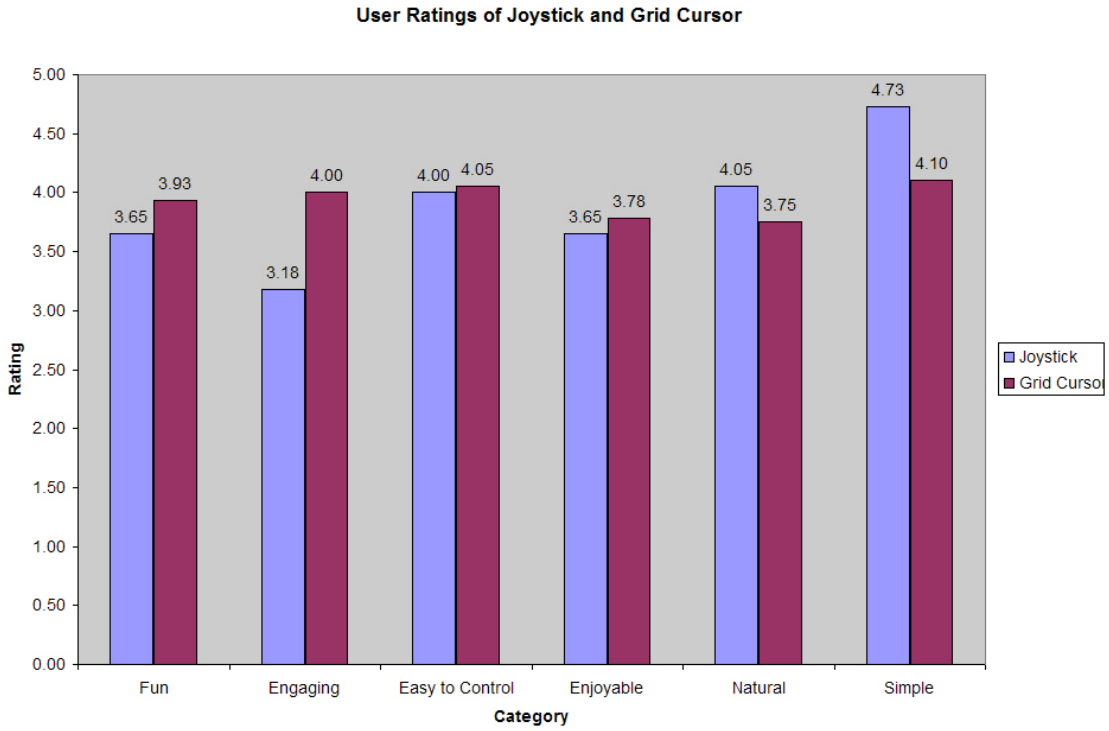

Figure 2.14: Subject ratings for the joystick and grid cursor. Joystick shown in blue, grid cursor in red.

Female subjects rated the joystick more engaging than male subjects, while male subjects rated the grid cursor more engaging than the female subjects did. This may imply that many users are most engaged when they are being challenged, in accordance with psychologist Mihly Cskszentmihlyi's work on "Flow" [5].

Chapter 3

Joystick vs. Speech (Nametags + Grid Cursor) Experiment

This chapter details the results of an experiment comparing two different mechanisms for cursor control - the standard joystick (a.k.a. "gamepad") and a combination of NameTags and the Grid Cursor. This was the second of three experiments intended to compare the performance and usability of speech-based cursor control methods with the current prevailing modality.

Experiment 1 established the joystick's performance superiority over the grid cursor for object selection. Subjects also found the joystick to be simpler than the grid cursor, while they rated the grid cursor as more engaging.

Experiment 2 combined NameTags and the grid cursor in an effort to improve object selection performance. As aforementioned, the grid cursor is necessary when pointing to unlabeled on-screen coordinates.

3.1 Experimental Design

3.1.1 Hypotheses

This experiment intended to answer the following questions:

- Which performs better for object selection, the joystick or NameTags?
- Which performs better for object movement, the joystick or the grid cursor?
- How does object size affect object selection performance for each mechanism?
- Which mechanism do subjects prefer?

The following hypotheses were tested:

- NameTags will outperform the joystick in selection time.
- NameTags performance will increase as object size decreases.
- The speech mechanism (NameTags + grid cursor) will be preferred over the joystick.
- NameTags will be preferred over the grid cursor.

3.1.2 Subjects

There were 40 subjects for this experiment, 16 females and 24 males, and their average age was 21.1. These subjects were selected from the student body and faculty of Jacksonville State University. The subjects needed no specialized knowledge or experience to participate in this research. Subjects' computer use, game play, and speech interface experience are shown in Figure 3.1. Distributions for computer use and game play are quite similar to the previous experiment. Speech system use was tracked in this experiment, and shows that an overwhelming majority of subjects had never used speech prior to using this system.

3.1.3 Setup

Subjects were asked to play a simple game in which their goal was to select the stationary object on-screen and guide it to a goal location as quickly as possible. The object to be selected had a text label when using NameTags, while the goal location did not. In order to test both mechanisms, a within-subjects design was chosen such that half of the subjects employed speech first, followed by the joystick, while the other half did the reverse. Each subject performed 50 tasks with each mechanism. Subjects were given basic instructions of

how to play the game, then were allowed to practice the given mechanism for 3 trial tasks before beginning the game proper.

The speech control component of the game was implemented in CloudGarden [4], a version of the Java Speech API. The game itself was programmed in Game Maker [11], a simple two-dimensional game engine. Communication between the speech component and the game was facilitated by keystroke messages generated by the the Java Robot class.

Figure 3.2 depicts two screenshots of the game. The joystick version is shown at the top, and the speech version is at the bottom.

In order to mitigate speech recognition errors, subjects completed the introductory session of the Microsoft Speech Recognition Training Wizard before using speech control.

When using speech, subjects could utter the following commands:

- To select objects: "Select < Name >"
- To shrink grid: "< 1 - 9 >"
- To move objects: "Go to < 1 - 9 >"
- To go back: "go back" or "back"

When using the joystick, the following controls were available:

- To move the cursor: Move the directional pad
- To select objects: Press button "A"
- To move objects: Press button "B"

As in the previous experiment, the movement speed for the joystick was set at 210 pixels/second. The joystick cursor was reset to the center of the screen at the beginning of each game level.

3.1.4 Object Size and Placement

As in the previous experiment, the objects to be selected were randomly placed on-screen and were randomly sized between 20 pixels and 100 pixels square on a 1024x768, 17" screen. The goal location was fixed at 100 pixels square.

3.1.5 Questionnaire

After the game was completed, subjects were asked to fill out a questionnaire which asked subjects for the following information:

- Age
- Gender
- How much time subject uses a computer per week
- How much time subject plays video games per week
- How much total time subject has used speech for dictation or control (pre-experiment)

Subjects were also asked to rate their subjective impressions on each control mechanism.

3.2 Results

3.2.1 Performance Comparison

This section details the performance of each control mechanism. The first section covers object selection, while the second section covers object movement.

Object Selection Performance

Overall, the joystick outperformed NameTags for stationary object selection by 9.5%. NameTags had much less variability in selection time, with a standard deviation of 0.226, compared to the joystick's 0.758. Results are shown in Figure 3.3.

NameTags, however, outperformed the joystick for selecting small objects. Figure 3.4 arranges the data points shown in Figure 3.3 in order of descending object size. In accordance with Fitts' Law, the joystick's selection times increase as the object to be selected decreases in size. NameTag's selection times, however, have a nearly flat slope. Therefore, one would expect that there is a point at which an object is so small that it takes longer to select with the joystick than with NameTags. In this experiment, that point was reached at approximately 44 pixels. Objects smaller than this took less time, on average, to select with NameTags than with the joystick.

Object Movement Performance

Similar to the previous experiment, the joystick outperformed the grid cursor in movement time by a factor of 2.9. Results are shown in Figure 3.5.

3.2.2 Subject Preference

As in the previous experiment, 40 subjects were asked to rate their experience with the joystick and speech based on the following 6 bipolar semantic categories:

- Boring or Fun
- Detached or Engaging
- Difficult to Control or Easy to Control
- Frustrating or Enjoyable
- Unnatural or Natural
- Complex or Simple

In order to differentiate between NameTags and the grid cursor, subjects were also asked to rate their experience with these control mechanisms. The next two sections detail the results of the questionnaire.

Subject Preference - Joystick vs. Speech

The speech mechanism received a higher mean score than the joystick for 4 out of the 6 categories. Of the 6 categories tested, 3 were statistically significant with an alpha value of .05. Subjects rated speech more fun (P=.002), more engaging (P=.000), and more enjoyable (P=.016). The results are shown in Figure 3.6.

One-way ANOVA tests revealed a substantial difference on the "ease of control" for speech ratings between male and female subjects(F=9.55, P=0.004). Male subjects gave speech a mean rating of 3.95 (std. dev.=0.999), versus female subjects' mean of 4.8 (std.

dev.=0.414). Hours spent playing video games shows an inverse relationship with "easy to control" speech ratings (F=3.39, P=0.029). Details of this relationship are shown in Figure 3.7.

Subjects' rating for speech in the "simple" category were affected by time spent using the computer (F=3.57, P=0.039) as well as time spent playing video games (F=3.83, P=0.018). Details of this relationship are shown in Figures 3.8 and 3.9.

A roughly inverse relationship exists between hours spent playing video games and speech's "natural" category(F=4.54, P=0.009). Details of this relationship are shown in Figure 3.10.

Another inverse relationship exists between hours spent using speech and speech's "engaging" category(F=3.33, P=0.048). Details of this relationship are shown in Figure 3.11.

Female subjects rated speech as more enjoyable (Mean=4.80, Std. Dev.=0.56) than male subjects did (Mean=4.00, Std. Dev.=1.07) in one-way ANOVA (F=7.03, P=0.012).

Subject Preference - NameTags vs. Grid Cursor

Subjects were also asked to compare NameTags to the grid cursor on the same 6 categories. While NameTags received a higher rating than the grid cursor in all 6 categories, none of them were statistically significant. The results are shown in Figure 3.12.

One-way ANOVA tests revealed a substantial difference on the "fun", "enjoyable", and "easy to control" ratings for NameTags between male and female subjects. Details of these relationships are shown in tabular format in Table 3.1.

Table 3.1: One-way ANOVA results show female subjects rate NameTags substantially higher than male subjects did.

	Mean		**Std. Dev.**			
Category	**Male**	**Female**	**Male**	**Female**	**F**	**P**
Fun - NameTags	3.73	4.40	0.827	0.737	6.43	0.016
Enjoyable - NameTags	3.95	4.60	0.722	0.633	7.86	0.008
Easy to Control - NameTags	4.18	4.73	0.907	0.458	4.70	0.037

As with speech in general, there is an inverse relationship between time spent playing video games and NameTag's "ease of control" ratings(F=3.36, P=0.030). Details of this relationship are shown in Figure 3.13.

3.2.3 Conclusion

This section provides a brief summary of the results of this experiment.

Overall, the joystick narrowly outperformed NameTags for selection of stationary objects. However, NameTags outperformed the joystick for objects smaller than 44 pixels square. NameTags selection time also had less variability since it is unaffected by distance from the target object. The joystick again proved more effective than the grid cursor for object movement.

Subjects gave speech a higher rating than the joystick for 4 out of 6 categories, 3 of which were statistically significant. Subjects rated speech more fun, more engaging, and more enjoyable than the joystick.

Female subjects rated speech higher as easier to control and more enjoyable than male subjects did. The more time subjects spent using a computer or playing games per week, the lower they rated speech as "simple". The more time subjects spent playing games, the less natural they considered speech as well.

The general trend here may be that people who spend very little or no time playing games may regard speech control as less complicated, since they have no other control mechanism (such as the joystick) to compare it with. The joystick has the advantage of familiarity, since most gamers use this control mechanism regularly. On the other hand, speech control has the advantage of novelty, since most subjects have very little or no experience using this mechanism. This is partially supported by the inverse relationship found between speech control experience and speech's "engaging" ratings.

Subjects narrowly preferred NameTags to the grid cursor. Female subjects rated NameTags as more fun, enjoyable, and easier to control than male subjects did. As with speech in general, time spent playing games and NameTags' "ease of control" rating were in an inverse relationship.

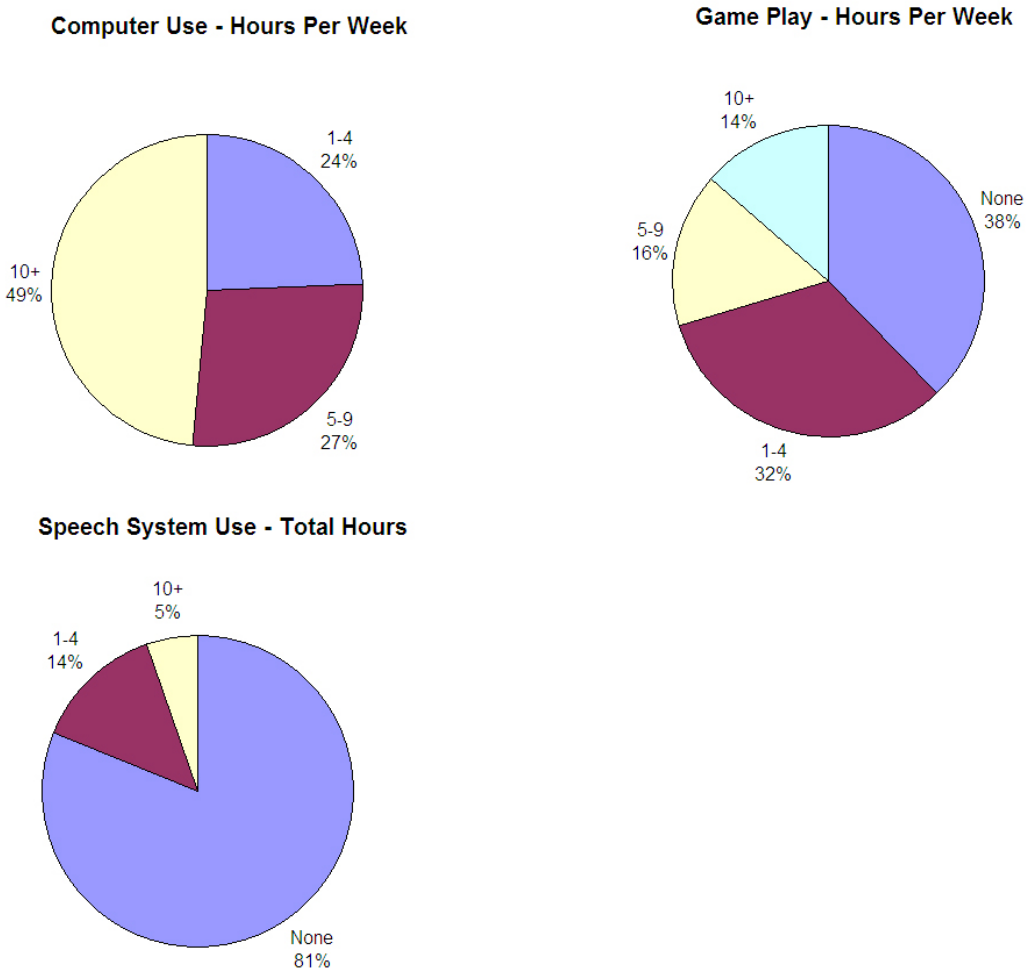

Figure 3.1: Subjects' computer use and game play (hours per week), as well as speech interface experience (total hours).

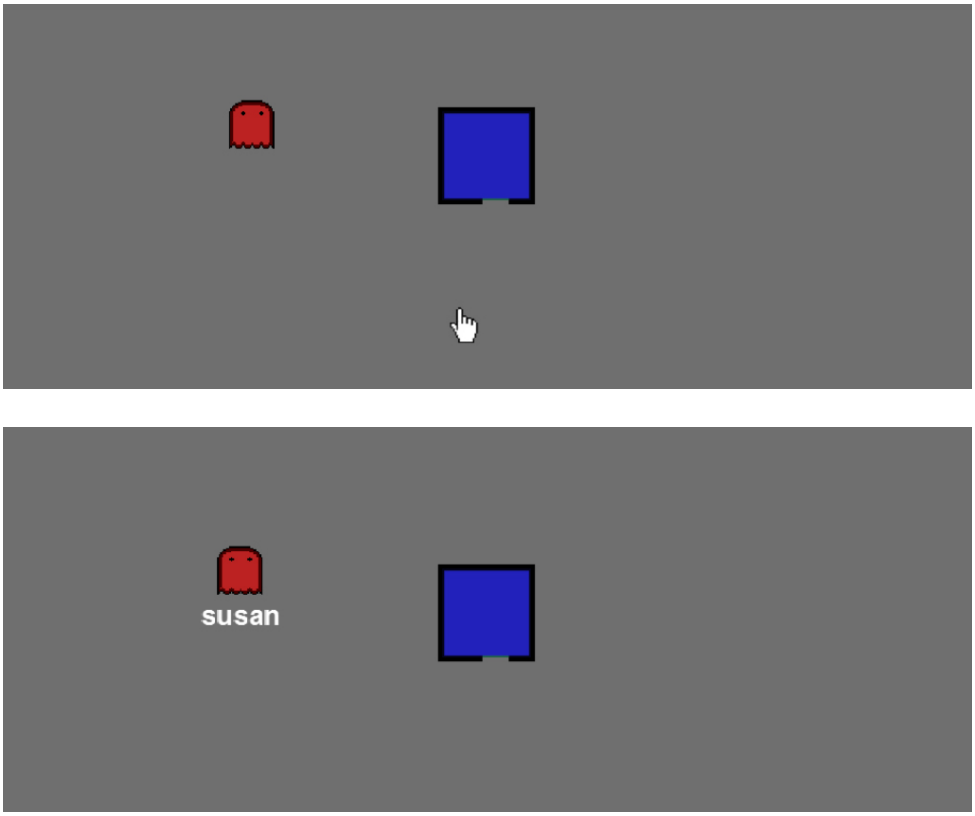

Figure 3.2: Joystick version of the game (top), and the speech version (bottom).

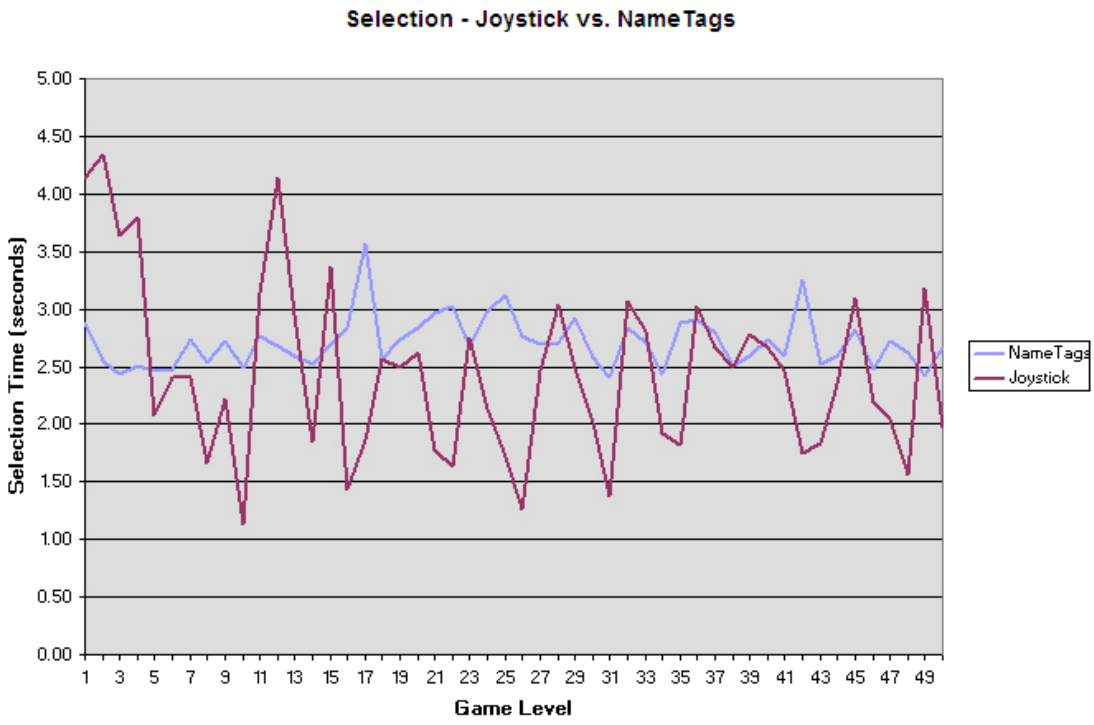

Figure 3.3: Selection times for the joystick and NameTags. NameTags shown in blue, joystick in red.

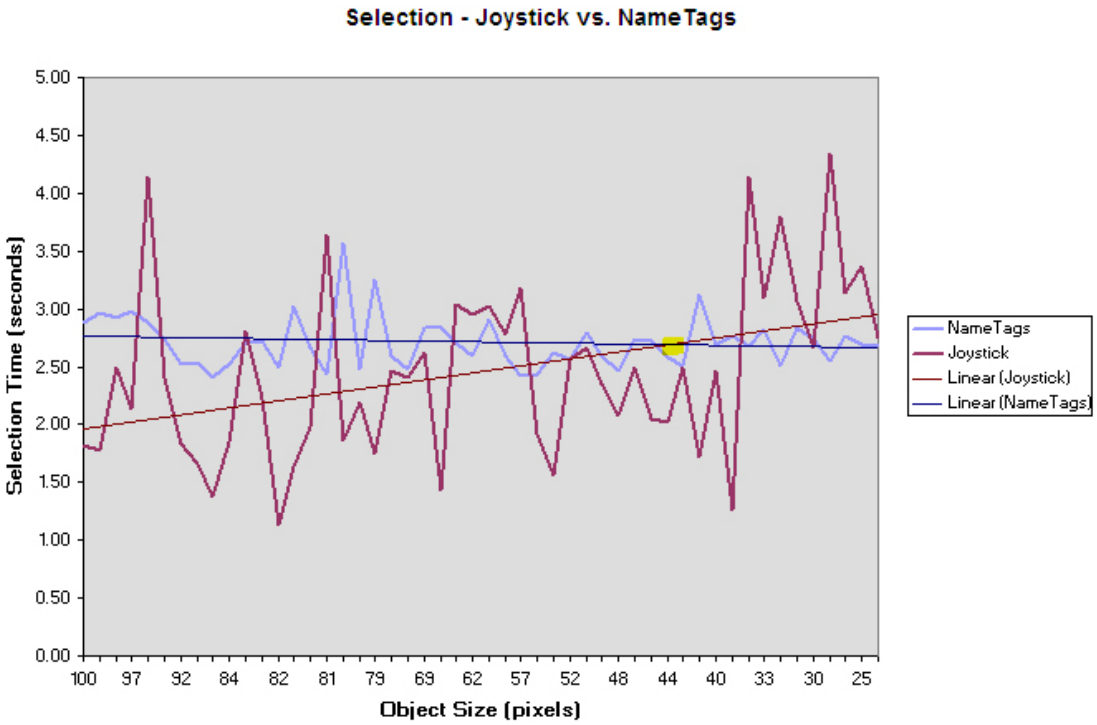

Figure 3.4: Selection times as object size decreases. NameTags shown in blue, joystick in red. A yellow square marks the point where the trendlines meet.

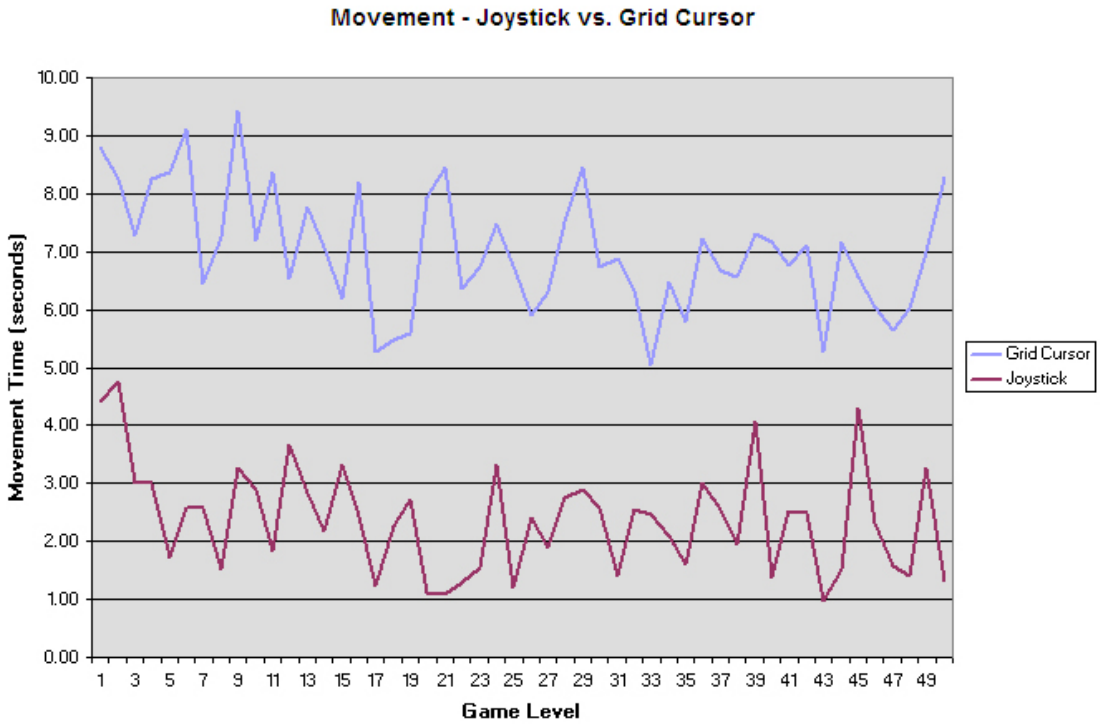

Figure 3.5: Movement times for the joystick and grid cursor. Grid cursor shown in blue, joystick in red.

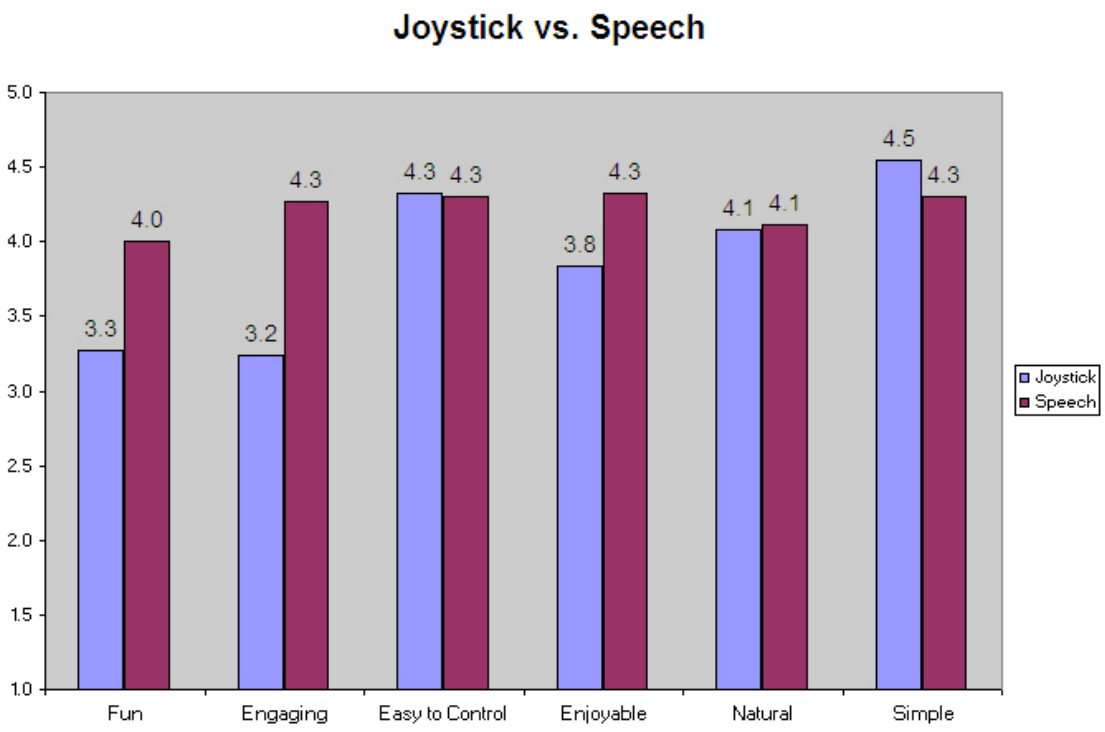

Figure 3.6: Subject ratings for the joystick and speech. Joystick shown in blue, speech in red.

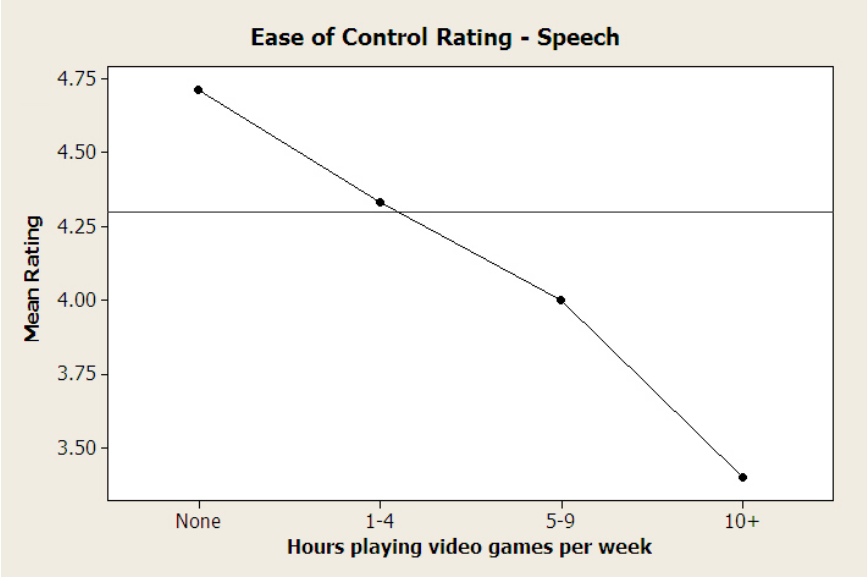

Figure 3.7: Ease of control ratings for speech across time spent playing video games.

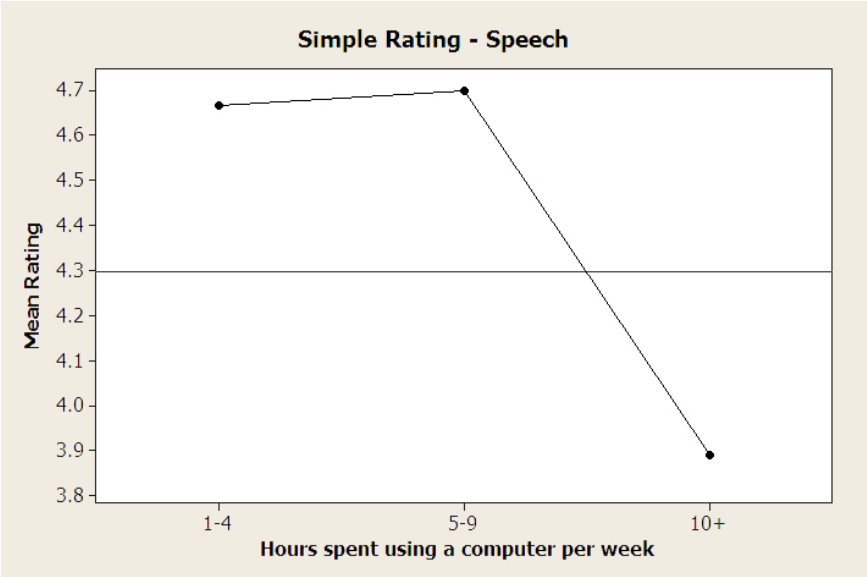

Figure 3.8: "Simple" ratings for speech based on time spent using a computer.

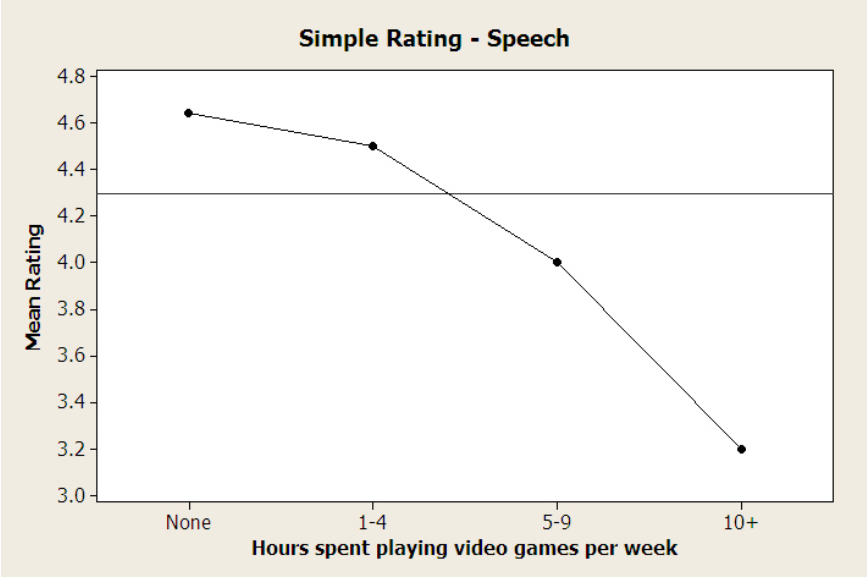

Figure 3.9: "Simple" ratings for speech based on time spent playing video games.

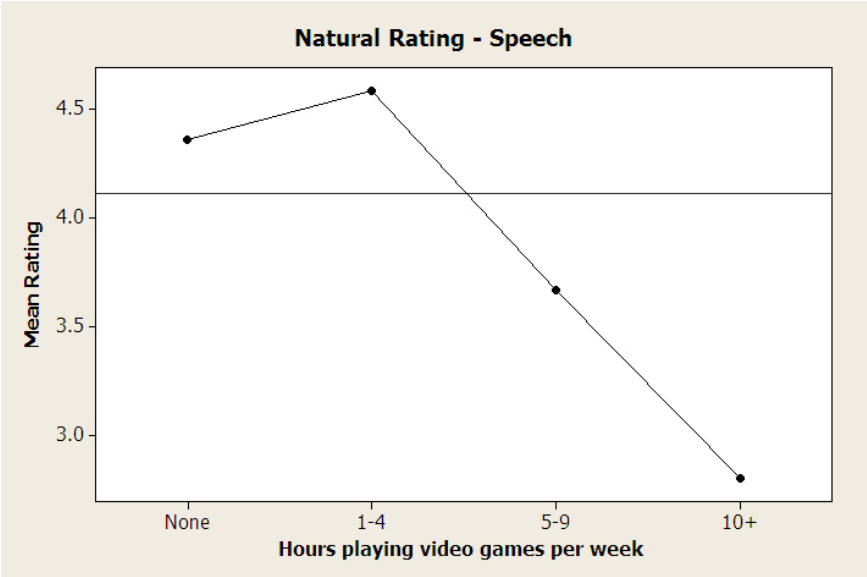

Figure 3.10: Natural ratings for speech based on time spent playing games.

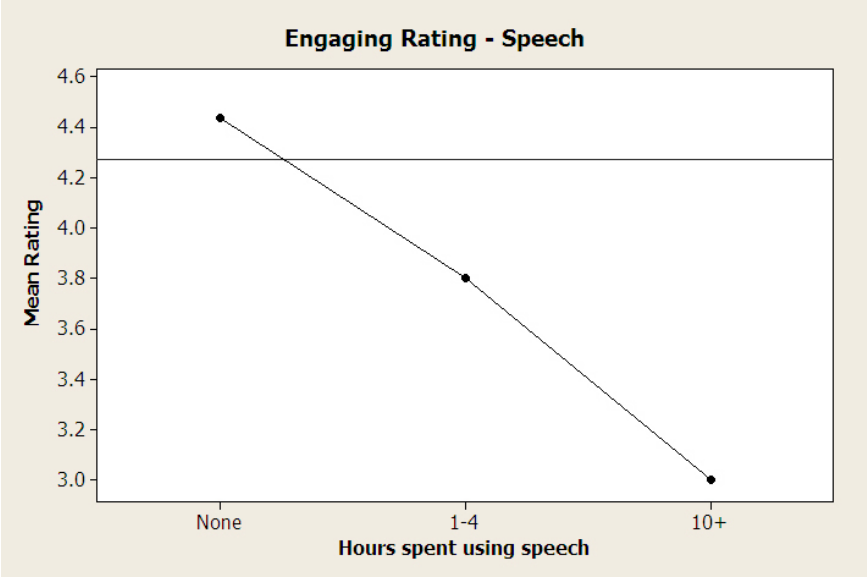

Figure 3.11: Engaging ratings for speech based on time spent using speech.

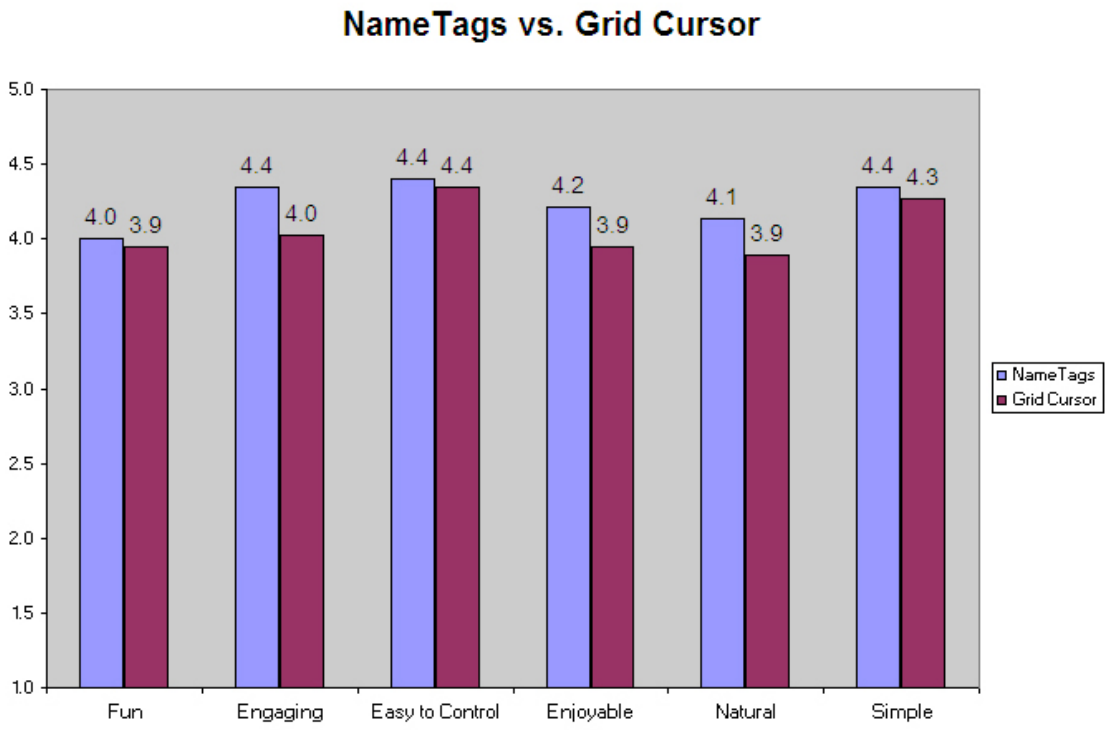

Figure 3.12: Subject ratings for NameTags and grid cursor. NameTags shown in blue, grid cursor in red.

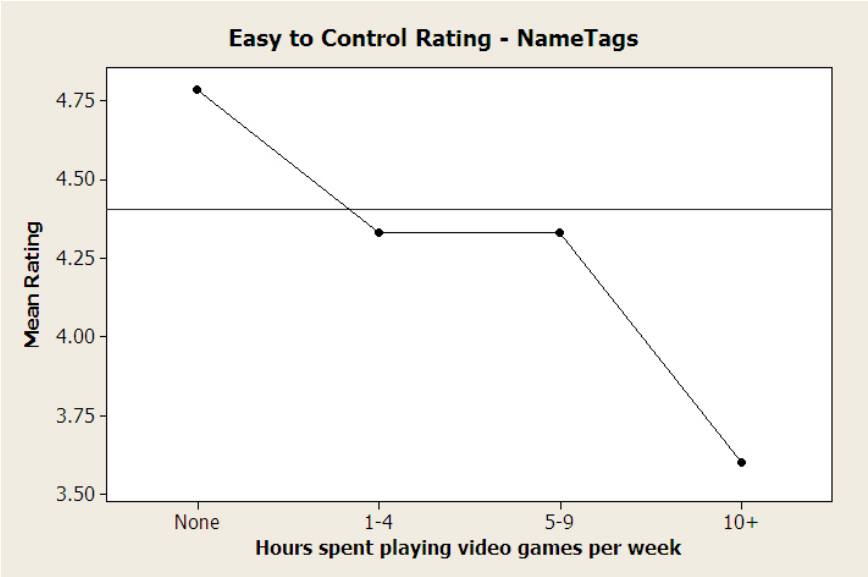

Figure 3.13: "Easy to control" ratings for NameTags based on time spent playing video games.

Chapter 4

Joystick vs. Multimodal (Speech + Joystick) Experiment

This chapter details the results of an experiment comparing two different modes of cursor control - the standard joystick (a.k.a. "gamepad") and a multimodal version (joystick, NameTags and the grid cursor). This was the third of three experiments intended to compare the performance and usability of speech-based cursor control methods with the current prevailing modality.

Experiment 1 established the joystick's performance superiority over the grid cursor for object selection. Subjects also found the joystick to be simpler than the grid cursor, while they rated the grid cursor as more engaging.

Experiment 2 showed that NameTags outperforms the joystick for object selection when objects are sufficiently small. Subjects rated speech more fun, more engaging, and more enjoyable than the joystick.

Experiment 3 sought to discover whether a multimodal interface could outperform either speech or joystick alone in terms of completion time and subjective satisfaction.

4.1 Experimental Design

4.1.1 Hypotheses

This experiment intended to answer the following questions:

- How is object selection affected by moving selectable objects?
- How is object selection affected by selecting multiple objects?

- How does multimodal performance compare with the joystick?
- When given the option to go multimodal, what will subjects prefer (both reported and observed)?

The following hypotheses were tested:

- Multimodal control will outperform the joystick overall in an environment with multiple, moving target objects.
- Subjects will prefer NameTags over the joystick for object selection in an environment with multiple, moving target objects.
- Subjects will prefer multimodal control to the joystick alone in an environment with multiple, moving target objects.

4.1.2 Subjects

There were 40 subjects for this experiment, 21 females and 19 males, and their average age was 25.9. These subjects were selected from the student body and faculty of Jacksonville State University. The subjects needed no specialized knowledge or experience to participate in this research. Subjects' computer use, game play, and speech interface experience are shown in Figure 4.1. Subjects' game play had a larger percentage of "None" than the previous experiments, as did speech system use.

4.1.3 Setup

Subjects were asked to play a simple game in which their goal was to select the multiple moving objects on-screen and guide them to a goal location as quickly as possible. The

objects to be selected had a text label in the multimodal version, while the goal location did not. In order to test both mechanisms, a within-subjects design was chosen such that half of the subjects employed the joystick first, followed by the multimodal version, while the other half did the reverse. Each subject performed at least 71 selection actions and at least 15 movement actions with each version of the game. Subjects were given basic instructions of how to play the game, then were allowed to practice the given mechanism for 3-4 trial tasks before beginning the game proper.

The speech control component of the game was implemented in CloudGarden [4], a version of the Java Speech API. The game itself was programmed in Game Maker [11], a simple two-dimensional game engine. Communication between the speech component and the game was facilitated by keystroke messages generated by the the Java Robot class.

Figure 4.2 depicts two screenshots of the game. The joystick-only version is shown at the top, and the multimodal version is at the bottom.

In order to mitigate speech recognition errors, subjects completed the introductory session of the Microsoft Speech Recognition Training Wizard before using speech control.

When using speech, subjects could utter the following commands:

- To select objects: "Select < Name >"
- To shrink grid: "< 1 - 9 >"
- To move objects: "Go to < 1 - 9 >"
- To go back: "go back" or "back"

When using the joystick, the following controls were available:

- To move the cursor: Move the directional pad

- To select objects: Press button "A"
- To move objects: Press button "B"

As in the previous experiment, the movement speed for the joystick was set at 210 pixels/second. The joystick cursor was reset to the center of the screen at the beginning of each game level.

4.1.4 Object Parameters

As in the previous experiment, the objects to be selected were randomly placed on-screen, and were fixed at 45 pixels square on a 1024x768, 17" screen. The goal location was fixed at 100 pixels square. Objects moved in predictable patterns, tracing one of three randomly chosen shapes: a circle, a square, or a line. Object movement speed was a constant 50 pixels/second. Each level contained either 6, 10, or 14 selectable objects, exactly half of which were the target objects. The others were present in order to simulate a more realistic game experience.

4.1.5 Questionnaire

After the game was completed, subjects were asked to fill out a questionnaire which asked subjects for the following information:

- Age
- Gender
- How much time subject uses a computer per week
- How much time subject plays video games per week

- How much total time subject has used speech for dictation or control (pre-experiment)

Subjects were also asked to rate their subjective impressions on each control mechanism.

4.2 Results

4.2.1 Multimodal Usage Statistics

This section covers usage statistics for each version of the game.

Overall, subjects overwhelmingly preferred NameTags over the joystick for object selection, and they preferred the joystick over the grid cursor for object movement. This is shown in Figure 4.3. The same data is shown as a function of time in Figures 4.4 and 4.5.

Usage histograms reiterate subjects' preference for NameTags over the joystick. Histograms for the joystick versus the grid cursor, on the other hand, show an interesting polarization. These show a marked "love it or hate it" pattern of usage. This is especially apparent in the histogram for the grid cursor. The aforementioned histograms are shown in Figure 4.6.

4.2.2 Performance Comparison

This section details the performance of each version of the game.

Overall, multimodal control outperformed joystick for level completion time by 25.7%. Detailed results are shown in Figure 4.7.

Figure 4.8 arranges the data points shown in Figure 4.7 in ascending order of number of objects to select. This data shows a flatter completion time slope for multimodal control than for the joystick, which indicates that as the number of objects increases past 7, multimodal control would likely continue to widen its performance gap with the joystick.

Table 4.1: Subjects' reported hours spent playing video games per week.

Hours spent playing video games	# Male Subjects	# Female Subjects
None	2	17
1-4 hours	8	1
5-9 hours	2	3
10+ hours	7	0

As in the previous two experiments, one-way ANOVA tests revealed that male subjects outperformed female subjects using the joystick. This performance difference was also true with multimodal control (F=8.06, P=0.007), with a mean completion time of 12.101 (std. dev.=4.345) versus female subject's 18.196 (std. dev.=8.391). As in the previous experiments, however, this is probably attributable to hours spent playing video games. Male subjects reported spending more time playing video games per week, as shown in Table 4.1 and Figure 4.9, and there exists an inverse relationship between level completion time and time spent playing video games, as shown in Figures 4.10 and 4.11. It is worth noting that the slope for multimodal completion time nearly levels off after the first drop, suggesting that while some video game experience increases performance, returns diminish fairly quickly.

4.2.3 Subject Preference

As in the previous experiment, 40 subjects were asked to rate their experience with the joystick and speech based on the following 6 bipolar semantic categories:

- Boring or Fun
- Detached or Engaging
- Difficult to Control or Easy to Control

- Frustrating or Enjoyable
- Unnatural or Natural
- Complex or Simple

Multimodal received a higher mean rating than the joystick for all 6 categories. Five of the 6 tested categories were statistically significant with an alpha value of .05. Subjects rated multimodal more fun (P=.000), more engaging (P=.000), easier(P=.000), more enjoyable (P=.000), and more natural (P=.016). The results are shown in Figure 4.12.

4.2.4 Conclusion

This section provides a brief summary of the results of this experiment.

Subjects overwhelmingly preferred NameTags over the joystick for object selection, employing it 84% of the time. Subjects preferred the joystick over the grid cursor for object movement; however, those that did use the grid cursor tended to continue to use it throughout the game.

Multimodal control outperformed joystick for level completion time by 25.7%. Further, results suggest that this performance gap would only widen as the number of target objects increases.

As in the previous two experiments, one-way ANOVA tests revealed that male subjects outperformed female subjects using the joystick. This performance difference was also true with multimodal control, though this is probably attributable to hours spent playing video games. Also, because the slope for multimodal completion time nearly levels off after the first drop, even a moderately experienced gamer may be able to use multimodal control ably.

Multimodal control received a higher mean rating than the joystick for all 6 categories. Five of the 6 tested categories were statistically significant. Subjects rated multimodal more fun, more engaging, easier, more enjoyable, and more natural.

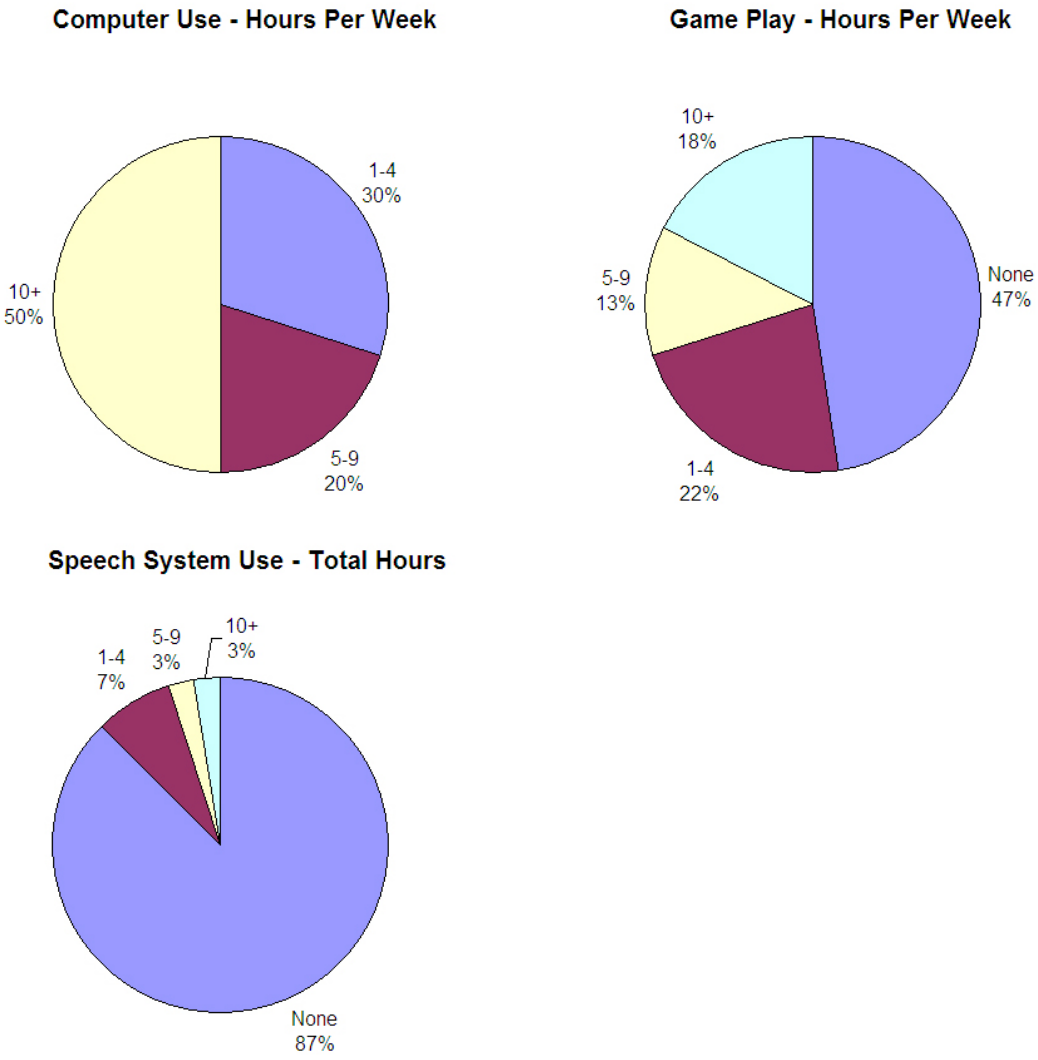

Figure 4.1: Subjects' computer use and game play (hours per week), as well as speech interface experience (total hours).

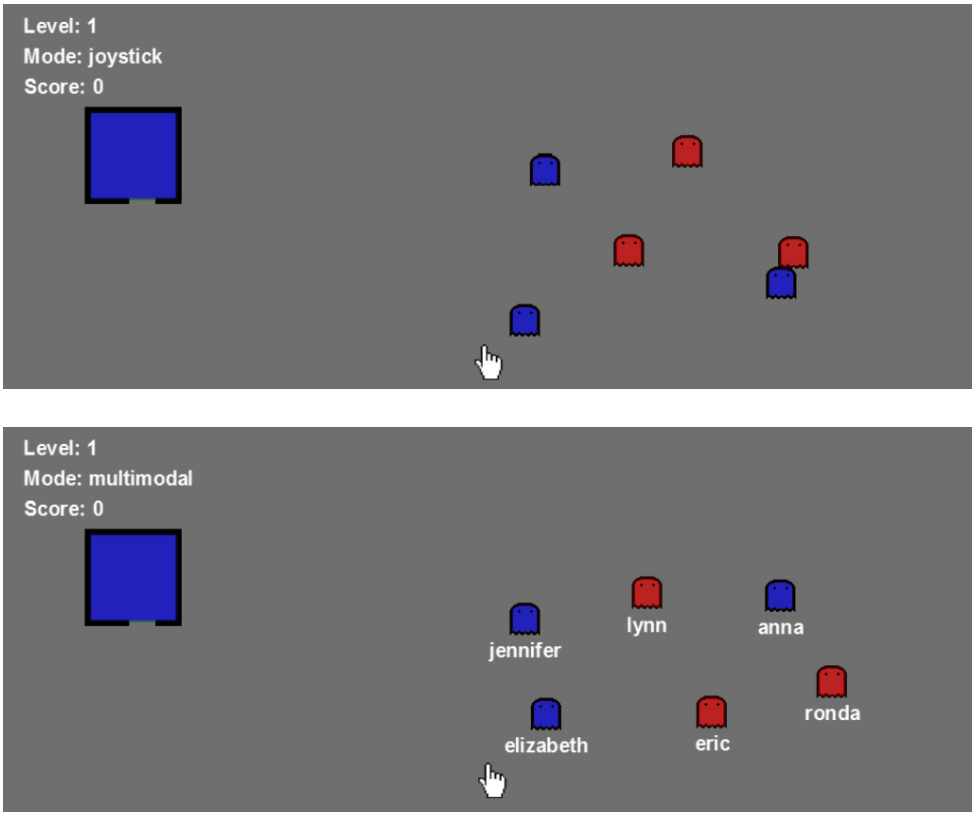

Figure 4.2: Joystick-only version of the game (top), and the multimodal version (bottom).

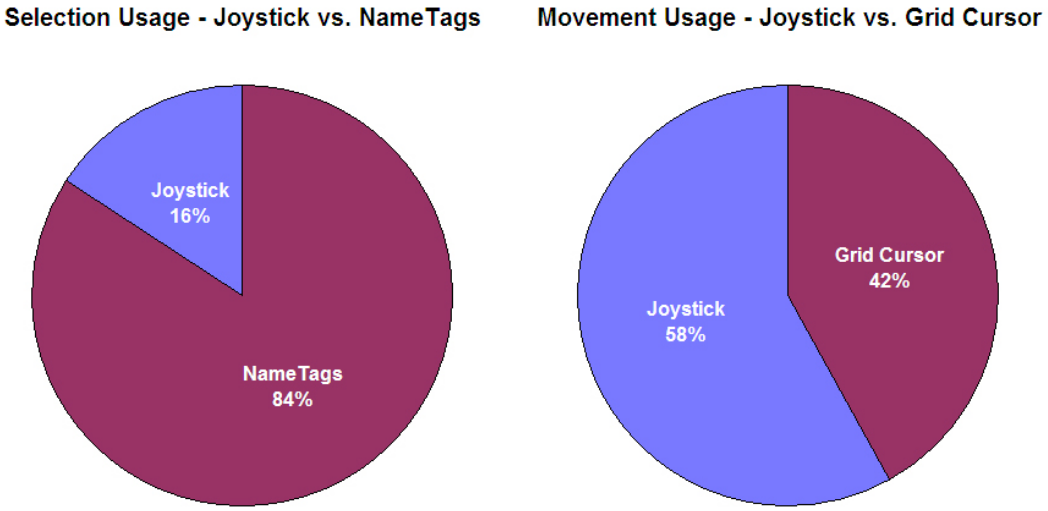

Figure 4.3: Subjects' overall usage percentages for selection and movement.

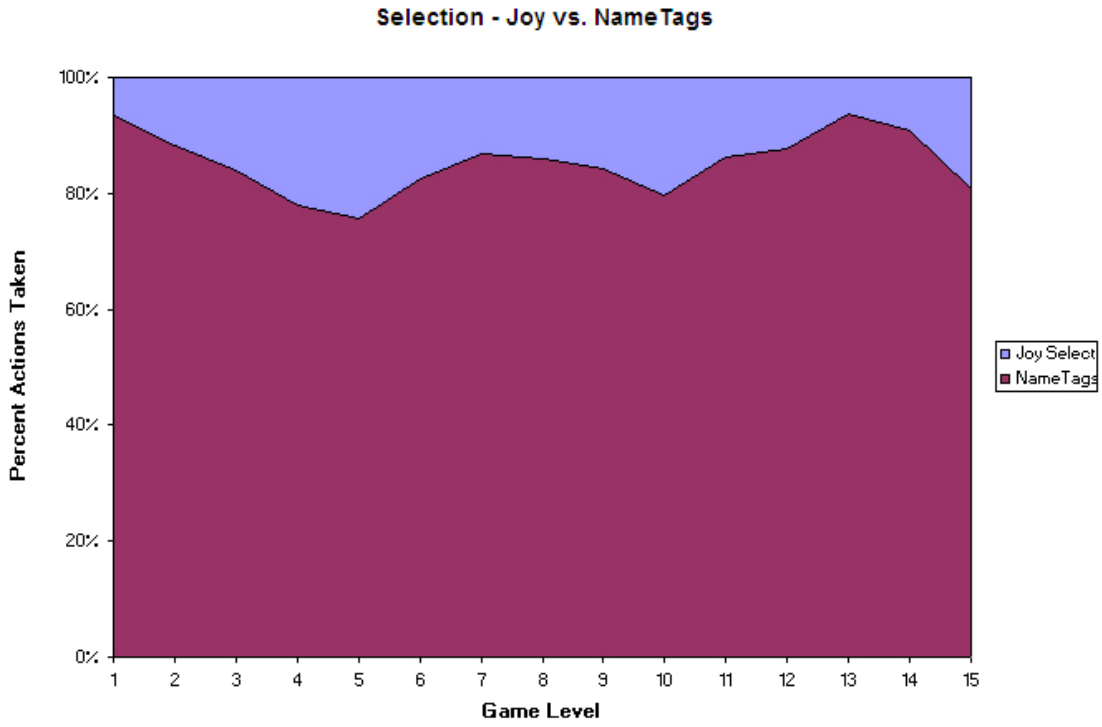

Figure 4.4: Subjects' usage percentages for selection based on game level.

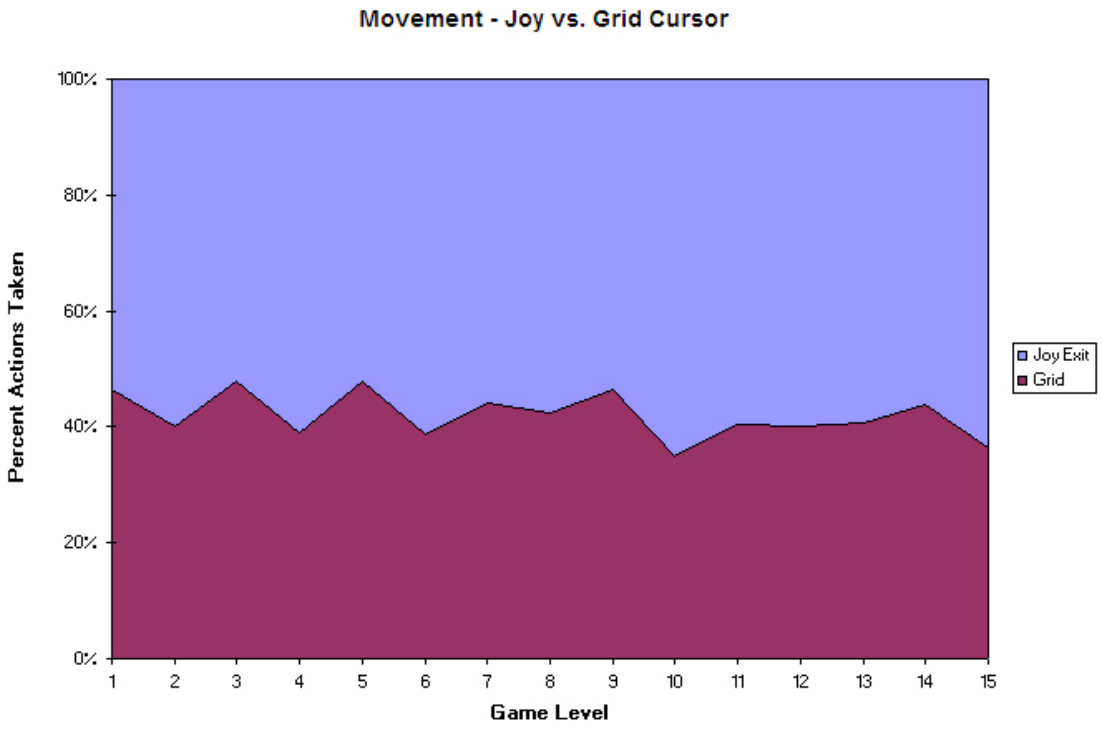

Figure 4.5: Subjects' usage percentages for movement based on game level.

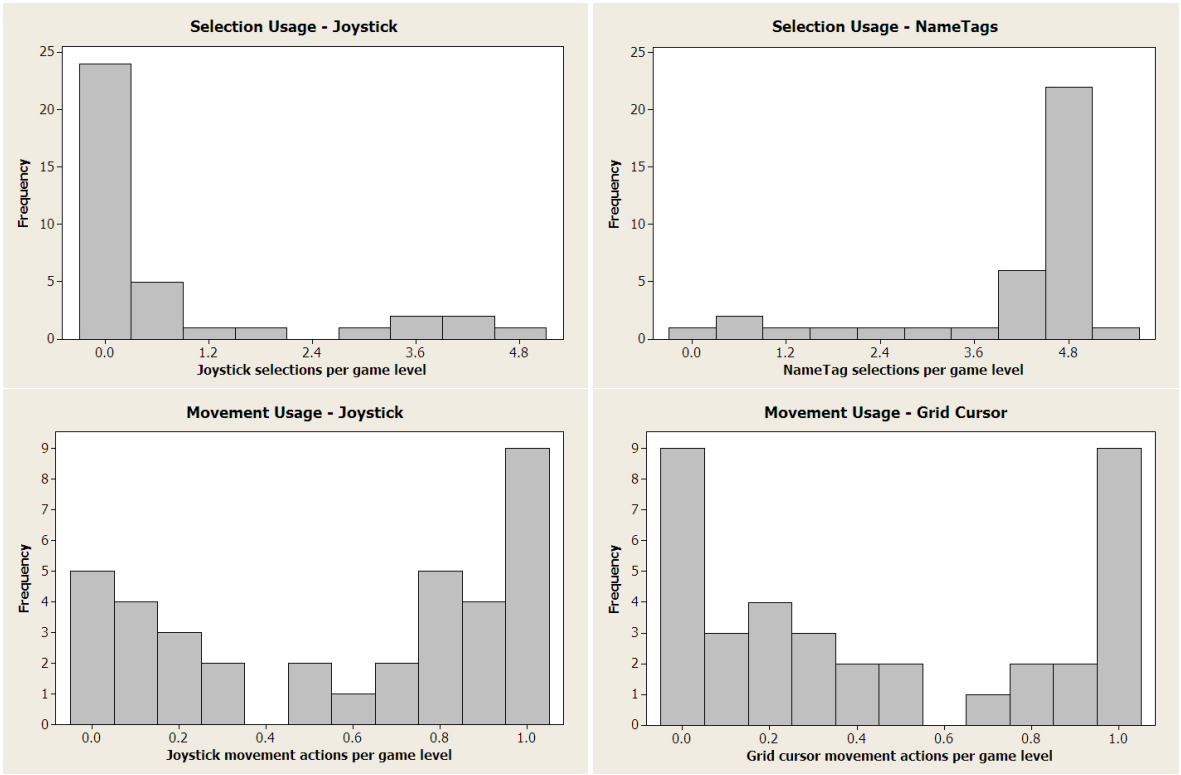

Figure 4.6: Histograms of subjects' usage. The movement usages are markedly polarized, especially the grid cursor.

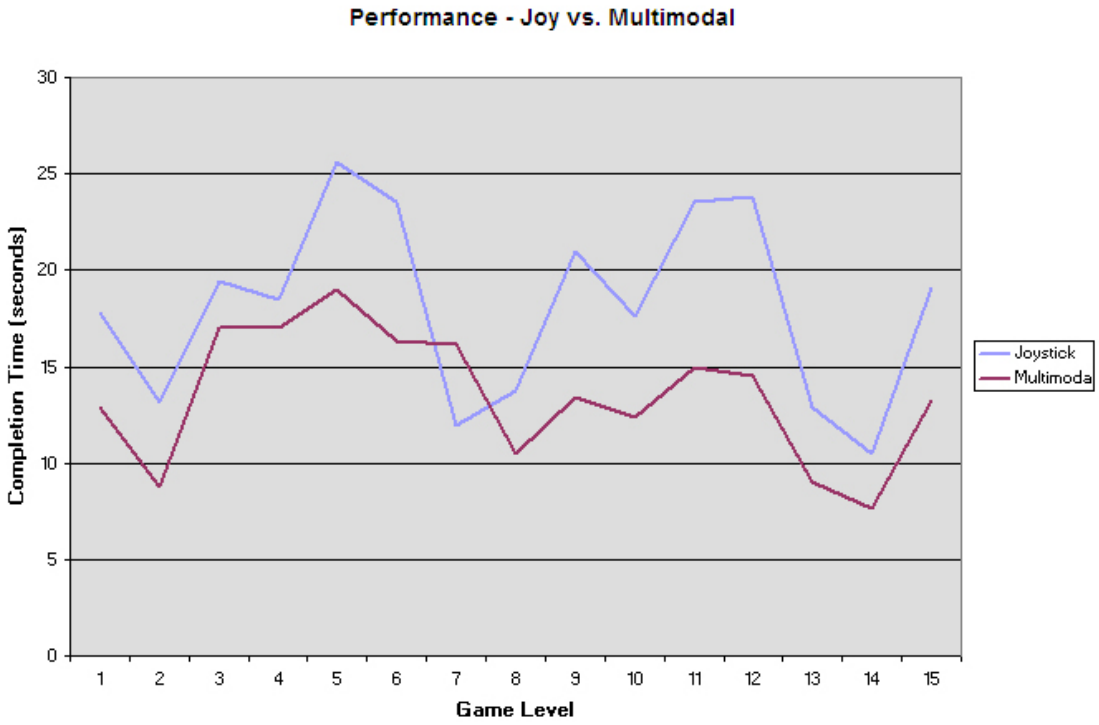

Figure 4.7: Level completion times for the joystick and multimodal control. Joystick shown in blue, multimodal in red.

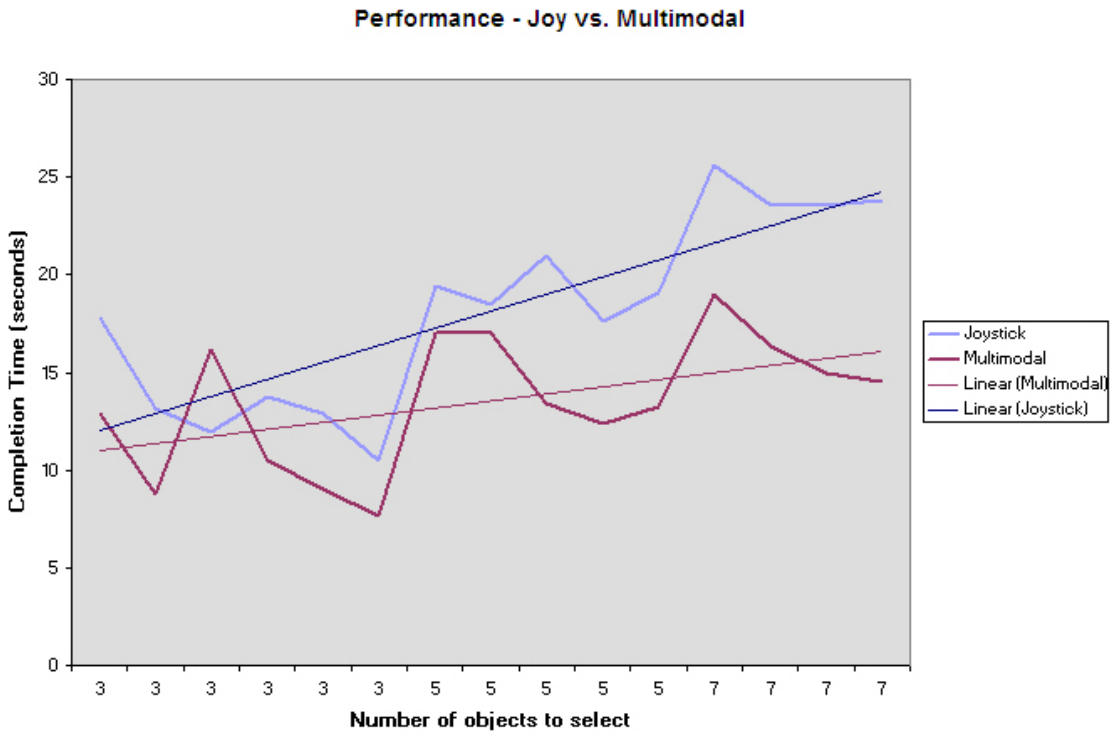

Figure 4.8: Level completion times as number of objects to select increases. Joystick shown in blue, multimodal in red.

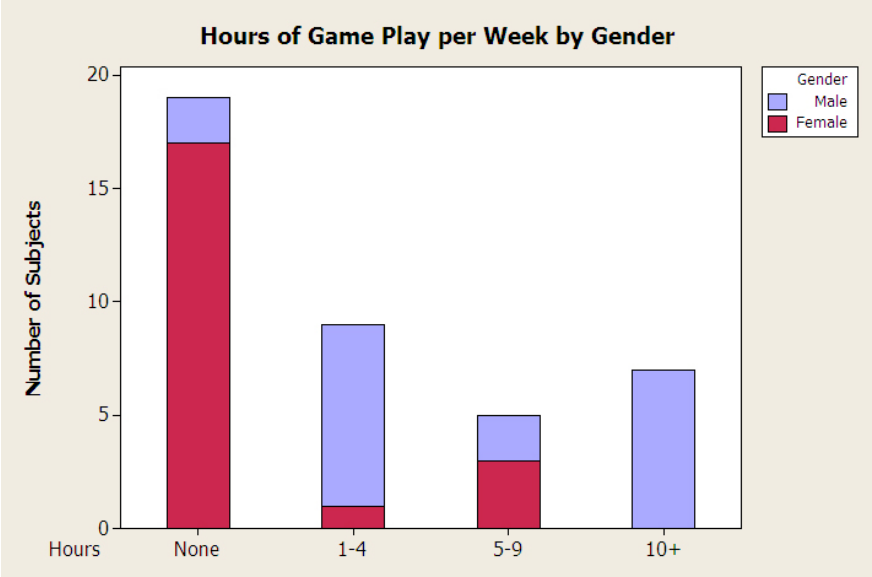

Figure 4.9: Subjects' reported hours spent playing video games per week.

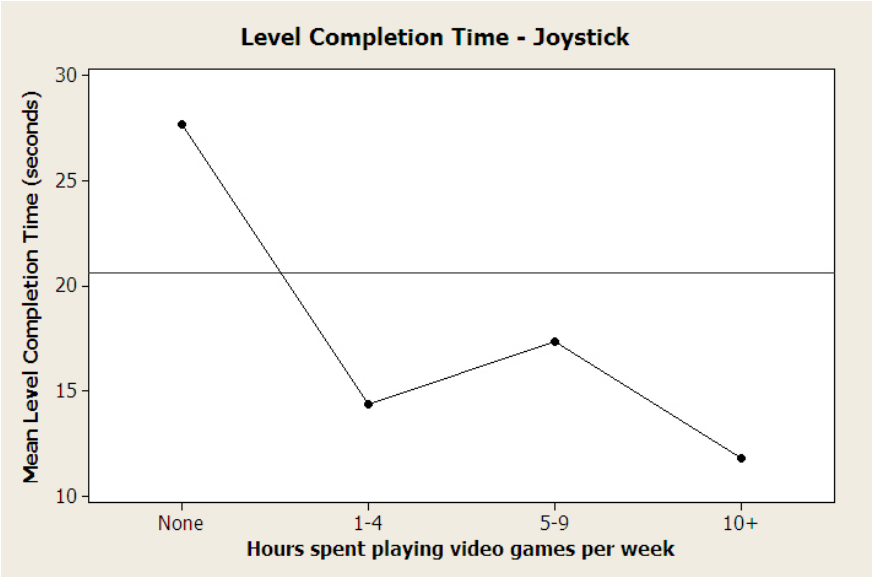

Figure 4.10: Level completion times with joystick based on time spent playing video games.

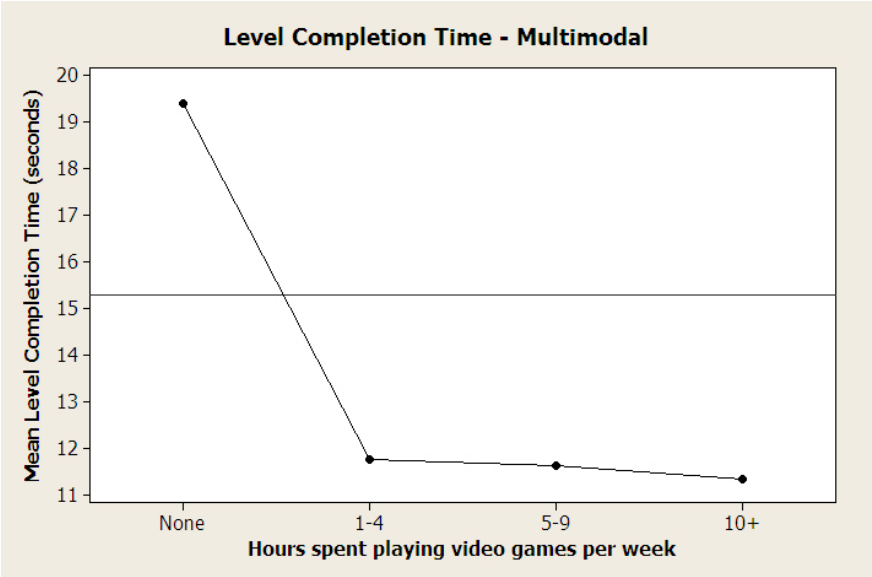

Figure 4.11: Level completion times with multimodal based on time spent playing video games.

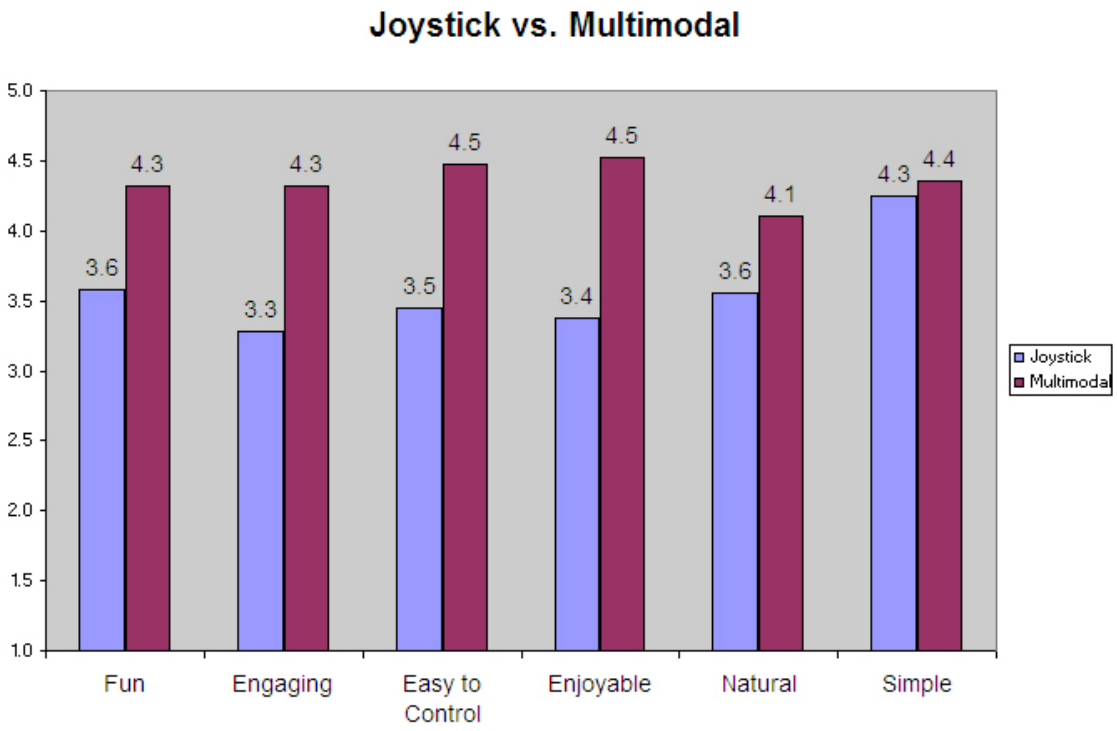

Figure 4.12: User ratings for the joystick and speech. Joystick shown in blue, speech in red.

Chapter 5

Conclusions and Future Work

This chapter covers overall study conclusions as well as ideas for future work.

5.1 Conclusions

In Experiment 1, the following hypotheses were tested:

- The joystick will outperform the grid cursor in completion time.
- The joystick will adhere tightly to Fitts' Law (greater than .90 Pearson's correlation).
- The grid cursor will adhere tightly to Fitts' Law (greater than .80 Pearson's correlation) with modifications to the formula.

The first and second were proven, but the third was not. More detail of this, as well as subject preference results, is shown below.

The joystick outperformed the grid cursor in mean selection time by a factor of 3.5, making the joystick the clear winner for completion time. Therefore, the grid cursor would be useful only when the joystick is not accessible or when time is not a factor. Joystick performance was highly dependent on the amount of time subjects played video games per week. Speech performance was highly dependent on the amount of time subjects spent using the computer. Because of these two factors, it is possible that users who are new to gaming but are familiar with computers may prefer the grid cursor from a usability standpoint. This is reinforced by the fact that the grid cursor received a higher mean rating than the joystick for 4 out of the 6 subjective preference categories, especially the "engaging" category.

The joystick adhered tightly to Fitts' Law, with a Pearson's correlation of 0.912. This implies that the joystick's performance time can be predicted with a high degreee of accuracy, much the same as other pointing devices like the mouse and touchpad.

The grid cursor did not adhere as tightly as the joystick, though it provided the tightest fit when using size as the only input parameter (correlation of 0.696). This fell short of the 0.80 goal fit.

Female subjects rated the joystick more engaging than male subjects, while male subjects rated the grid cursor more engaging than the female subjects did. Given the fact that male subjects' completion time was significantly shorter than that of their female counterparts, this may imply that many users are most engaged when they are being challenged, in accordance with psychologist Mihly Cskszentmihlyi's work on "Flow" [5].

In Experiment 2, the following hypotheses were tested:

- NameTags will outperform the joystick in selection time.
- NameTags performance will increase as object size decreases.
- The joystick will outperform the grid cursor for movement time.
- The speech mechanism (NameTags + grid cursor) will be preferred over the joystick.
- NameTags will be preferred over the grid cursor.

The first hypothesis was disproven, while the other four were proven. More detail of this, as well as subject preference results, is shown below.

Overall, the joystick narrowly outperformed NameTags for selection of stationary objects. However, NameTags outperformed the joystick for objects smaller than 44 pixels square. NameTags selection time also had less variability since it is unaffected by distance

from the target object. This indicates that NameTags is very effective in systems where the target objects are small.

The joystick again proved more effective than the grid cursor for object movement.

Subjects gave speech a higher rating than the joystick for 4 out of 6 categories, 3 of which were statistically significant. Subjects rated speech more fun, more engaging, and more enjoyable than the joystick. This implies that players may be open to speech as a primary modality if it is designed correctly.

Female subjects rated speech higher as easier to control and more enjoyable than male subjects did. Perhaps this is an indication that female gamers would prefer more speech control than is currently being offered, though this deserves further study. The more time subjects spent using a computer or playing games per week, the lower they rated speech as "simple". The more time subjects spent playing games, the less natural they considered speech as well. Both of these may indicate that users who are already familiar with a particular control modality (eg., mouse, keyboard, joystick) may be hesitant to adopt a new, unfamiliar one.

Subjects narrowly preferred NameTags to the grid cursor. Female subjects rated NameTags as more fun, enjoyable, and easier to control than male subjects did. As with speech in general, time spent playing games and NameTags' "ease of control" rating were in an inverse relationship.

In Experiment 3, the following hypotheses were tested:

- Multimodal control will outperform the joystick overall in an environment with multiple, moving target objects.

- Subjects will prefer NameTags over the joystick for object selection in an environment with multiple, moving target objects.
- Subjects will prefer multimodal control to the joystick alone in an environment with multiple, moving target objects.

All three of these hypotheses were proven. More detail of this, as well as subject preference results, is shown below.

Multimodal control outperformed joystick for level completion time by 25.7%. Further, results suggest that this performance gap would only widen as the number of target objects increases. This indicates that multimodal control is preferable when high performance is required.

Subjects overwhelmingly preferred NameTags over the joystick for object selection, employing it 84% of the time. Subjects preferred the joystick over the grid cursor for object movement; however, it is interesting to note that those subjects that did use the grid cursor tended to continue to use it throughout the game.

As in the previous two experiments, one-way ANOVA tests revealed that male subjects outperformed female subjects using the joystick. This performance difference was also true with multimodal control, though this is probably attributable to hours spent playing video games. Also, because the slope for multimodal completion time nearly levels off after the first drop, even a moderately experienced gamer may be able to use multimodal control ably. This is a promising result for video game developers who are seeking to incorporate speech control.

Multimodal control received a higher mean rating than the joystick for all 6 categories. Five of the 6 tested categories were statistically significant. Subjects rated multimodal more fun, more engaging, easier, more enjoyable, and more natural.

Simply put, multimodal control (speech + joystick) yielded both higher performance and higher subjective satisfaction than the joystick alone. This speaks strongly in favor of multimodal control in environments similar to the one tested.

5.2 Future Work

Eighty-five percent of subjects polled had never used speech control before. Because of this, it was not possible to discover how experienced speech users differed in performance and preference from inexperienced ones. Future studies may include users who have a well-distributed variety of experience using speech control. Also, time constraints limited the number of game levels. Multiple, extended sessions may uncover interesting trends not found in this study. For instance, the joystick is a familiar control mechanism for most users, while speech control is generally novel. As users have more experience with speech, their opinions about its ease of use and its engaging quality may change.

Because the bulk of subjects in this study were traditional college students, there is little information about how age may affect user performance and preference. Future studies may include well-distributed age groups to discover new information. For instance, non-traditional gamers might prefer speech since they have little experience with the joystick. Also, while speech control has been proven useful in the past for users with motor skill disabilities, perhaps elderly users would also prefer speech instead of wrestling with the eye-hand coordination required for the joystick.

It would also be interesting to test these speech mechanisms in conjunction with other software tools for disabled users in order to discover new, unexpected synergy.

Chapter 6

Further Consideration of the Grid Cursor

6.1 Model Comparison

While experiment 1 found that a subject's movement time with the joystick could be accurately predicted using Fitts' Law, movement time with the Grid Cursor did not map tightly to the Fitts' Law prediction (with a Pearson's correlation of .696). Dai theorized that selection time with the grid cursor should adhere closely to the following formula:

$$T = log_3(\frac{S}{W})$$

In this formula, S represents the size of the screen and W is the size of the target object.

The research committee also suggested a similar model, shown below:

$$T = log_9(\frac{S}{W})$$

Subjects' performance data from experiment 1 was applied to both formulations, and the results are shown in Figures 6.1 and 6.2.

6.2 Conclusions

Both Dai's model and the committee's suggested model produce the same quality of results. While they provide a marginally tighter fit than the modified formulation of Fitts' Law, neither solution approaches a desirable level of accuracy. Part of this discrepancy may be blamed on the subjects' learning curve. One can visually compare the first and second

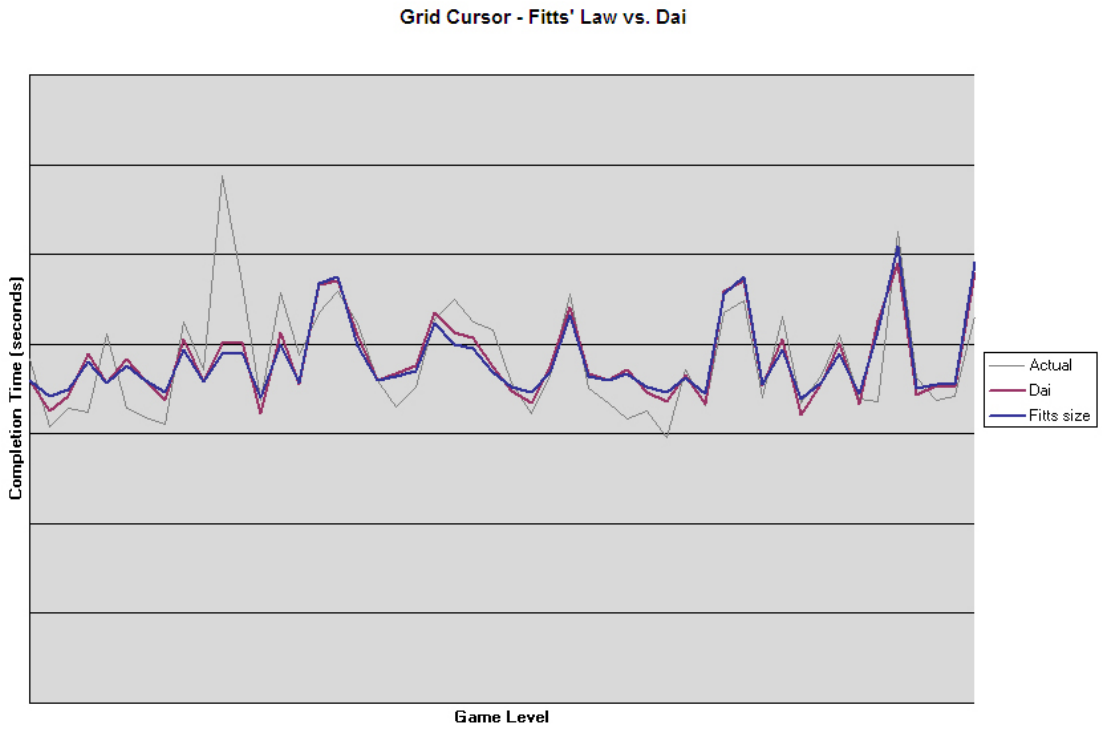

Figure 6.1: Grid cursor Fitts' prediction time (blue), Dai model (red), and the actual movement time (gray). Pearson's correlation of .732.

half of the graph to note clear correlation improvement as the game goes on. Detail of this disparity is shown for the committee's suggested model in Table 6.1.

This underscores the aforementioned need to perform similar studies that provide multiple, extended game sessions that may mitigate this issue.

Even with this factor considered, it is clear that more variables need to be considered in order to define a complete and satisfactory predictive model. These variables include, but are not limited to, the following:

- The time required for a user to determine which cell contains the desired target object
- The time required for a user to recognize the label assigned to that cell

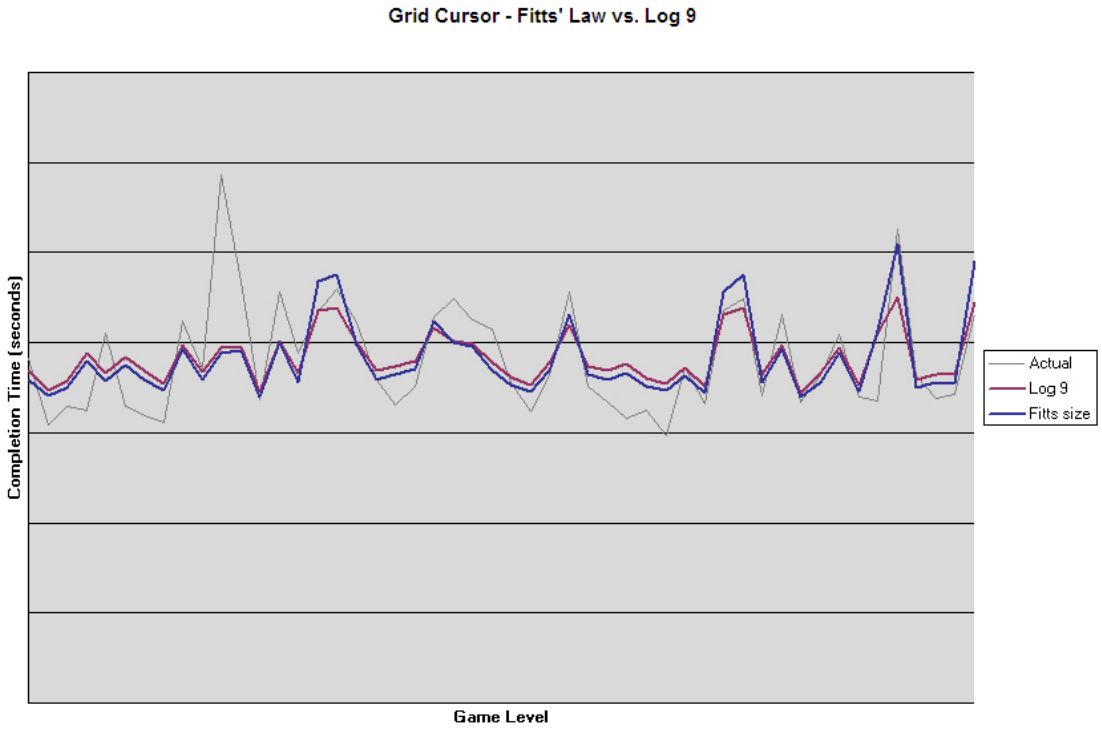

Figure 6.2: Grid cursor Fitts' prediction time (blue), committee's suggested model (red), and the actual movement time (gray). Pearson's correlation of .732.

- The time required for a user to say the cell's label
- The time required for the speech engine to recognize the user's command

While some of these variables are shared by the direction-based speech cursor control mechanisms discussed in Chapter 1, they do not cause the undesired "over-correction" problems associated with them. Instead, these factors only cause a delay between a user's formulation and realization of their goal.

Also, it bears reiterating that the version of the grid cursor employed in these experiments was bound to a worst-case constraint that required the selection cell to be entirely

Table 6.1: Comparison of first 25 levels and second 25 levels for Pearson's correlation. Subjects' performance more closely matched predictions as the game progressed.

Game Levels	**Pearson's Correlation**
1-25	.646
26-50	.864

inside the target object. A relaxation of this constraint would produce better average selection time. A highly accurate model should take this into account. When selecting an object with this constraint relaxed, a user is effectively drawing a box around the target object to isolate it from any other objects on-screen. When specifying an arbitrary point or selecting a target object with the worst-case constraint, on the other hand, the user's goal is to place the center of some cell at the desired location. If the desired point falls just outside the center of a cell, users may find themselves required to shrink the grid several times in order to specify that precise location. As a result, a highly accurate model would provide two separate formulations for these different tasks.

Bibliography

[1] Blizzard Entertainment, Inc. Warcraft III: Reign of Chaos. www.blizzard.com, 2003.

[2] Buisine, S. and Martin, J. 2005. Children's and Adults' Multimodal Interaction with 2D Conversational Agents. CHI 2005, 1240-1243.

[3] Christian, K., Kules, B., Schneiderman, B., and Youssef, A. 2000. A comparison of voice controlled and mouse controlled web browsing. Proc. of ASSETS'00, 72-79.

[4] CloudGarden, Inc. www.cloudgarden.com, 2008.

[5] Cskszentmihlyi, M. 1990. Flow: The Psychology of Optimal Experience'. New York: Harper and Row. ISBN 0-06-092043-2.

[6] Dai, L., Goldman, R., and Sears, A. 2004. Speech-based Cursor Control: A Study of Grid-based Solutions. ACM Assets ' 04, October 18-20, 2004.

[7] Feng, J. 2002. Improving Speech-based Navigation during Dictation. CHI 2002, 844-845.

[8] Feng, J. and Sears, A. 2004. Using Confidence Scores to Improve Hands-free Speech Based Navigation in Continuous Dictation Systems. CHI 2004, 329-356.

[9] Fitts, P.M. 1964. Information Capacity of Discrete Moto Responses. Journal of Experimental Psychology, 67, 103-112.

[10] Stategy First, Inc. Galactic Civilizations. www.galciv.com, 2003.

[11] Yo Yo Games, Inc. Game Maker. www.yoyogames.com, 2008.

[12] Gray, J., Shaik, S. et al. SpeechClipse: an Eclipse Speech Plug-in. OOPSLA Workshop on Eclipse Technology Exchange, 2003.

[13] Halverson C., Horn D., Karat C. and Karat J.(1999). The Beauty of Errors: Patterns of Error Correction in Desktop Speech systems. Proceedings of INTERACT'99, 1-9.

[14] Harada, S., Landay, J., Malkin, J., Li, X., and Bilmes, J. 2006. The Vocal Joystick: Evaluation of Voice-based Cursor Control Techniques. Proc. of ASSETS'06, 197-204.

[15] Hauptmann, A. Speech and Gestures for Graphic Image Manipulation. ACM 0-89791-301-9, 1989.

[16] Igarashi, T., Hughes, J. 2001. Voice as Sound: Using Non-verbal Voice Input for Interactive Control. ACM UIST'01, 155-156.

[17] Kamel, H. and Landay, J. 1999. The Integrated Communication 2 Draw (IC2D): A Drawing Program for the Visually Impaired. CHI '99 extended abstracts, 222-223.

[18] Kamel, H. and Landay, J. 2000. A Study of Blind Drawing Practice: Creating Graphical Information without the Visual Channel. Proc. of ASSETS'00, 34-41.

[19] Kamel, H. and Landay, J. 2002. Sketching Images Eyes-free: A Grid-based Dynamic Drawing Tool for the Blind. Proc. of ASSETS'02, 33-40.

[20] Karimullah, A., S., and Sears, A. 2002. Speech-based Cursor Control. Proc. of ASSETS'02, 178-185.

[21] Manaris, B. and Harkreader, A. 1998. SUITEKeys: A Speech Understanding Interface for the Motor-control Challenged. Proc. of ASSETS'98, 108-115.

[22] Manaris, B., McCauley, R., and MacGybers, V. 2001. An Intelligent Interface for Keyboard and Mouse Control - Providing Full Access to PC Functionality via Speech. Proc. of International Florida AI Research Symposium (FLAIRS'01), 182-188.

[23] MacKenzie, I and Buxton, W. Extending Fitts' Law to Two-dimensional Tasks. Proc. of CHI'92, 219-226.

[24] Mihara, Y., Shibayama, E, and Takahashi, S. 2005. The Migratory Cursor: Accurate Speech-based Cursor Movement by Moving Multiple Ghost Cursors using Non-Verbal Vocalizations. Proc. of ASSETS'05, 76-83.

[25] Oviatt, S. L. 1997. Multimodal Interactive Maps: Designing for Human Performance. Human-Computer Interaction, 12, 93-129.

[26] Oviatt, S. L. 2000. Taming speech recognition errors within a multimodal interface. Comm. ACM 43,9, 45-51.

[27] Palazzolo, M., Humphrey, C., Nagel, J., and Stolz, A. Adaptation of a Video Game Controller for Use by a Quadriplegic Incorporating Real-time Speech Processing, IEEE 0-7803-7740-0, 2003.

[28] Perakakis, M. and Potamianos, A. 2007. The Effect of Input Mode on Inactivity and Interaction Times of Multimodal Systems. ICMI'07, 102-109.

[29] Sanchanta, M. (2007-09-12). Nintendo's Wii takes console lead. Financial Times. http://www.ft.com/cms/s/0/51df0c84-6154-11dc-bf25-0000779fd2ac.html. Retrieved on 2008-01-23.

[30] Sargin, M, et al. Combined Gesture-speech Analysis and Speech Driven Gesture Synthesis. Enterface05 Workshop, Belgium, 2005.

[31] Sears A., Karat C-M., Oseitutu K., Karimullah A., Feng J. (2001). Productivity, Satisfaction, and Interaction Strategies of Individuals with Spinal Cord Injuries and Traditional Users Interacting with Speech Recognition Software. International Journal of Universal Access in the Information Society, 1(1), 4-15.

[32] Smith, J., and Graham, N. Use of Eye Movements for Video Game Control. ACM ACE 06, June 14-16, 2006.

[33] Sporka, A., Kurniawan, H., Mahmud, M., and Slavik, P. Non-speech Input and Speech Recognition for Real-time Control of Computer Games. Proc. of ASSETS'06, 213-220.

[34] Wang, S et al. Face Tracking as an Augmented Input in Video Games: Enhancing Presence, Role-playing and Control. CHI 2006, April 22-27, 2006.

[35] Zhang, J., Zhao, J., Bai, S., and Huang, Z. Applying Speech Interface to Mahjong Game. Proc. of 10th International Multimedia Modeling Conference (MMM'04), 2004.

Appendix

This section displays the grammars used in the speech control component of the game.

```
#JSGF V1.0;
grammar nametags;

public <namecommand> = select {SELECT} ((<moniker> [and])*);

<moniker> =
 dexter | kevin | david | susan | amanda | james | samantha | robert |
 nicholas | lynn | christopher | jennifer | kelly | laura | ashley | cindy |
 adam | gary | brandon | kimberly | vincent | lisa | pablo | eric | scott |
 patty | anna | snoopy | george | napoleon | elizabeth | alexander |
 michelle | lucy | peter | maria | charlotte | tiffany | ronda | janice |
 olivia | lesley | steve | michael;


#JSGF V1.0;
grammar grid;

public <gridcommand> =
grid (on {GRID_ON} | off {GRID_OFF})   |
(go | move) to {GOTO} <cellnumber> |

<cellnumber> {CELL_NUM} |
[go | grid] back {GRID_BACK};

<cellnumber> = 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9;
```