

# ChibiPoint: Accessible Pointing for Web Applications

Alexander Birch

November 2013

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Context and Motivation . . . . .	3
1.2	Literature & Technology Review . . . . .	3
1.3	Problem statement and Hypothesis . . . . .	3
1.3.1	Hypothesis . . . . .	3
1.4	Goals and Methods . . . . .	4
1.5	Results / Contributions . . . . .	4
1.6	Thesis overview . . . . .	4
<b>2</b>	<b>Literature &amp; Technology Review</b>	<b>5</b>
<b>3</b>	<b>Requirements &amp; Design</b>	<b>6</b>
<b>4</b>	<b>Implementation</b>	<b>7</b>
<b>5</b>	<b>Evaluation</b>	<b>8</b>
5.1	Overview . . . . .	8
5.2	Study 1: Usability Study . . . . .	9
5.3	Study 2: Quantitative Comparison of Pointing Systems . . . . .	9
5.3.1	Outline . . . . .	9
5.3.2	Demographic . . . . .	9
5.3.3	Methodology . . . . .	10
5.3.4	Metrics . . . . .	10
5.3.5	Pilot Study . . . . .	11
<b>6</b>	<b>Results</b>	<b>12</b>
6.1	Overview . . . . .	12
6.1.1	Keystrokes . . . . .	12
6.1.2	Time . . . . .	12
6.2	Analysis used . . . . .	12

6.3	Keystroke Performance . . . . .	13
6.3.1	Task 1: [Within form] Immediate related traversal . . . . .	13
6.3.2	Task 2: [Within form] distant related traversal . . . . .	13
6.3.3	Task 3: Visually early element . . . . .	13
6.3.4	Task 4: Visually late element . . . . .	14
6.3.5	Task 5: [Within form] Spatially close, markup distant traversal . . . . .	14
6.3.6	Task 6: Low visual indicator of focus . . . . .	15
6.3.7	Task 7: Scrolled element . . . . .	15
6.3.8	Task 8: Link surrounded by links . . . . .	15
7	Conclusions & Future work	16
8	Figures	17
9	Appendices	19

# Chapter 1

## Introduction

### 1.1 Context and Motivation

### 1.2 Literature & Technology Review

### 1.3 Problem statement and Hypothesis

#### 1.3.1 Hypothesis

**Hypothesis 1: ChibiPoint is generally more efficient at web navigation than tabbing**

We hypothesise that our system, **ChibiPoint**, **requires significantly fewer keypresses to navigate webpages, than does ‘tabbing navigation’** (the current standards-prescribed method for keyboard navigation). ‘Navigate webpages’ is a broad term, as there are many classes of pointing that need to be analysed. The more detailed hypothesis is that, **given several distinct classes of pointing, ChibiPoint requires fewer (or equal) keypresses, for a large majority of these scenarios**. Since we analyse two versions of ChibiPoint, it should be understood that we require at least one of these to outperform tabbing in a majority of tasks.

**Hypothesis 2: ChibiPoint’s ‘flyouts’ feature, reduces further the required keypresses for web navigation**

We hypothesise that the novel ‘flyouts’ feature of ChibiPoint, which assigns hotkeys to suggested buttons on the webpage, contributes to reducing the number of keypresses required by ChibiPoint for web navigation, compared to using just its standard pointing method of hierarchically drilling through the page and pointing at elements using crosshairs. In other words, we hypothesise that ChibiPoint ‘with crosshairs and flyouts enabled’ performs a large majority of pointing scenarios in fewer keypresses than ChibiPoint ‘with just crosshairs enabled’.

## **1.4 Goals and Methods**

## **1.5 Results / Contributions**

## **1.6 Thesis overview**

## **Chapter 2**

### **Literature & Technology Review**

## **Chapter 3**

# **Requirements & Design**

# **Chapter 4**

## **Implementation**



# Chapter 5

## Evaluation

### 5.1 Overview

The primary system comparison we sought to pursue was ‘ChibiPoint’ versus ‘tabbing navigation’. We aimed also to evaluate the contribution made by the ‘flyouts’ feature of ChibiPoint: whether it helps or hinders. As such, we evaluated three systems:

1. Tabbing navigation
2. ChibiPoint (with just the crosshairs feature enabled)
3. ChibiPoint (with both the crosshairs feature AND the flyouts feature enabled)

We aimed to quantitatively evaluate the following things:

1. Which system is the most efficient for each class of navigation tested
2. Which system is the fastest for each class of navigation tested
3. Which system is preferred by the participants
4. Which system is considered by the participants to be easiest to use
5. How reasonable is the amount of keypressing demanded by each system, according to the participants

Qualitative data was useful also, as it offered feedback on specific advantages or disadvantages of each system, as well as providing insight to usability and the reasons for user preferences.

We performed first a usability study, as a precursor to the more detailed quantitative testing. The usability study was expected to reveal how a new user approaches the system — what level of instruction is necessary, and what problems are encountered with usage. Addressing issues here would reduce problems in the larger, quantitative study that followed.

Regarding the quantitative study: a pilot study was conducted first, to ensure that the study approach was effective, before widening the participation. Omissions from that pilot study were addressed. A larger study then recruited 12 users to evaluate the systems in a counterbalanced  $3 \times 8$  ANOVA (three systems, 8 navigation types).

## 5.2 Study 1: Usability Study

## 5.3 Study 2: Quantitative Comparison of Pointing Systems

### 5.3.1 Outline

A pilot study set the sequence for the full study to follow. The main output of the study was the evaluation of pointing systems; participants were instructed to complete several tasks of the following ilk:

1. Go to a specified website.
2. Using the specified pointing system, click a specified button/page element.

After completing this set of tasks with one pointing system, the participant was asked to repeat it with another of the pointing systems under test, until all three systems had been tested. We counterbalanced the order in which pointing systems were allocated to participants, to reduce biases in learning the tasks or systems.

In addition to this evaluation of system performance, participants were asked to fill out two questionnaires. The first questionnaire was posed before the pointing tasks: participants declared their proficiencies with the computing concepts involved, as well as describing their demographic. The second questionnaire was posed after the pointing tasks, and asked the participants to give subjective feedback on the systems they used.

### 5.3.2 Demographic

All participants were known personally by the researcher, with a majority (9) explicitly identifying as being current Computer Science students.

The mean average age for participants was 22.58, with a standard deviation of 2.151. The range was 21–28. These data are tabulated in Figure 8.2. A quarter of participants were female, the rest were male (Figure 8.3).

#### Proficiencies

Most participants were frequent users of computing devices; 5 using computers or smartphones for 13+ hours per day, 5 using the same for 7-9 hours, and just 2 using devices for 4-6 hours. Other categories (0-1, 2-3, 10-12) were offered, but no participants identified with these. These frequencies can be seen in Figure 8.1.

Many participants (5) preferred to navigate using a mouse, with a majority (6) recruiting hotkeys in addition to the mouse. Only one participant actually preferred keyboard controls, and no participant identified with completely requiring full mouse support or full keyboard support (Figure 8.4).

Touch-typing proficiency was rated on a five-point scale ranging from the sentiment ‘I look at every key before pressing’ to ‘I always type without even a reference glance at the keyboard’ (Figure 8.5). Participants on (mean) average considered themselves a 3.42. A standard deviation of 1.084 was observed. Two primary users of the Dvorak keyboard layout were asked to use the (familiar, but not preferred) Qwerty layout during this study, and rated their touch-typing with respect to the constraint that they would be typing in Qwerty during this time.

A majority (11) used mainly Windows as their desktop Operating System. Browser choice was more divided, with 7 participants using mainly Google Chrome, 3 Firefox and 1 ‘Other’ (at the time disclosed to be Safari, which was not an option offered on the questionnaire).

## **Disability**

Regarding participant’s exposure to accessibility needs or solutions, three had used Dragon NaturallySpeaking speech-to-text. The two who elaborated on the extent of their use of speech-to-text described only limited use. No participant had — at the time, or ever — had computer-relevant vision difficulties. Two had had in the past minor experiences with RSI, with one describing “RSI from mouse use, but nothing too severe to require a large change in typing/clicking”, the other elaborating “Have changed input device due to pain in very specific circumstances.”. No participants had RSI at the time of the study.

### **5.3.3 Methodology**

The following describes the methodology for the full study. Deviations exclusive to the pilot study are disclosed in its separate section.

The researcher followed a script to ensure that all instances of the study were performed the same way.

Participants were briefed on the purpose of the dissertation and of the study, as well as what participation in the study would entail. They were assured that no identifying information would be kept about them, that it was the system performance being measured rather than their personal performance, and that they were free to withdraw participation at any point. With these explained, participants were asked to sign a permission slip, and were offered also a copy in case they wished to contact the researchers after the fact.

Participants were assigned a unique number. This enabled both the questionnaires they would fill in to be related as being from the same participant. Additionally participants were assigned a sequence in which to evaluate the three systems. Biases pertaining to the order in which systems are used, were counterbalanced by testing using all sequence permutations of the three systems equally. There were 6 permutations, so the study recruited a multiple of this: 12 participants.

The study began with the first questionnaire’s being allocated. This asked participants to disclose their demographic (occupation, gender and age).

The participants began the testing activity by reading instructions on the system they had been allocated. The researcher then guided them through a training scenario to confirm that they had understood the instructions. During this training period, the researcher would answer any questions and correct any mistakes that occurred. In the case of learning the system ‘ChibiPoint with crosshairs AND fly-outs’ before learning the system ‘ChibiPoint with crosshairs ONLY’, participants were trained in the use of both features in turn.

[INCOMPLETE]

### **5.3.4 Metrics**

The metric that was most important to study was efficiency — how many keypresses (and by extension, how much exertion) are required for a given navigation task. Secondary to this metric was time taken

by each system to perform a navigation task — it is preferred, but not essential, for the system to be fast; the main priority is reducing exertion, which is connected to efficiency.

During the navigation tasks, user keypresses were recorded, as well as the times that they occurred. We recorded only keypresses relevant to pointing with the current system. For example, attempts to use the ‘tab’ key during use of ChibiPoint were intercepted, and answered with an alert to the user that tabbing was disallowed. We disregarded keyboard input that did not invoke functionality within tabbing or the currently activated version of ChibiPoint. Page scrolling via the keyboard was considered unrelated to pointing, and so was not recorded. Attempts to invoke flyouts were only recorded when said flyout was present; if no flyouts were created (for example because no clickables are in the specified area, or if flyouts are disabled altogether), then the keypress was disregarded.

The post-task questionnaire asked participants to give subjective feedback on the systems they used. Some of the feedback was quantitative — which was the preferred pointing system, how easy was each system to use, and how reasonable was the amount of key pressing required — and some of the feedback was qualitative, asking for general thoughts on the task and systems.

### **5.3.5 Pilot Study**

# Chapter 6

## Results

### 6.1 Overview

We compared, for each of the eight navigation tasks, the number of keystrokes each system required to complete that navigation. Time was compared also, but as this was of secondary interest, less analysis was performed.

#### 6.1.1 Keystrokes

Each system was found to have different strengths. Tabbing was found to excel significantly at just one task: traversing short distances in forms (for example, from username field to password field). For longer form traversals, it was matched or beaten by the various ChibiPoint systems, and it even failed to surpass ChibiPoint at accessing visually early elements. In the remaining five cases, both versions of ChibiPoint greatly outperformed tabbing. In all but one case, the ‘flyouts’ feature of ChibiPoint reduced the number of keypresses. For the remaining case, ‘Visually Early Element’, there was no significant difference between the versions of ChibiPoint.

#### 6.1.2 Time

### 6.2 Analysis used

Keystroke results were analysed, for each task, using a repeated measures ANOVA. Greenhouse-Geisser correction was used to determine how significantly the mean keystrokes differed between systems. This correction was used because sphericity could not be assumed of our data; there is no homogeneity of variance, as tabbing variance emerges very differently to ChibiPoint variance (tabbing is deterministic excepting mistakes, whereas ChibiPoint navigation can take multiple valid routes).

After ascertaining whether there existed within that task a significant difference in performance between the systems, we performed — using Bonferroni correction — post-hoc tests pairwise between each combination of systems used. This told us whether, between any two systems, there existed a significant difference in keystrokes.

## 6.3 Keystroke Performance

For shorthand, the ChibiPoint systems will be named: CX (ChibiPoint with just crosshairs feature enabled) and CX+F (ChibiPoint with crosshairs and flyouts enabled).

We consider statistical significance for  $p < 0.01$ .

### 6.3.1 Task 1: [Within form] Immediate related traversal

System	Keypresses Mean (s.d.)
CX	5.67 (.492)
CX+F	3.00 (.000)
Tabbing	1.00 (.000)

Mean keystrokes differed statistically significantly between systems.  
( $F(1, 11) = 814.000$ ,  $p < 0.001$ ).

CX keystrokes significantly greater than Tabbing ( $p < 0.001$ ).

Analysis failed to produce significance value for CX+F vs Tabbing.

CX+F keystrokes significantly fewer than CX ( $p < 0.001$ ).

Hypothesis 1 refuted in this case.

Hypothesis 2 supported.

### 6.3.2 Task 2: [Within form] distant related traversal

System	Keypresses Mean (s.d.)
CX	4.67 (.492)
CX+F	2.25 (.452)
Tabbing	4.33 (1.155)

Mean keystrokes differed statistically significantly between systems.  
( $F(1.574, 17.311) = 33.543$ ,  $p < 0.001$ ).

CX keystrokes non-significantly different to Tabbing ( $p = 1.000$ ).

CX+F keystrokes significantly fewer than Tabbing ( $p = .001$ ).

CX+F keystrokes significantly fewer than CX ( $p < .001$ ).

Hypothesis 1 supported (for CX+F only).

Hypothesis 2 supported.

### 6.3.3 Task 3: Visually early element

System	Keypresses Mean (s.d.)
CX	5.00 (.603)
CX+F	4.67 (.888)
Tabbing	7.00 (.000)

Mean keystrokes differed statistically significantly between systems.  
( $F(1.590, 17.489) = 54.057, p < 0.001$ ).

CX keystrokes significantly fewer than Tabbing ( $p < .001$ ).

CX+F keystrokes significantly fewer than Tabbing ( $p < .001$ ).

CX+F keystrokes non-significantly different to CX ( $p = .797$ ).

Hypothesis 1 supported.

Hypothesis 2 not supported.

#### 6.3.4 Task 4: Visually late element

System	Keypresses Mean (s.d.)
CX	5.25 (.452)
CX+F	3.17 (.577)
Tabbing	72.50 (1.732)

Mean keystrokes differed statistically significantly between systems.  
( $F(1.246, 13.701) = 15056.088, p < 0.001$ ).

CX keystrokes significantly fewer than Tabbing ( $p < .001$ ).

CX+F keystrokes significantly fewer than Tabbing ( $p < .001$ ).

CX+F keystrokes significantly fewer than CX ( $p < .001$ ).

Hypothesis 1 supported.

Hypothesis 2 supported.

#### 6.3.5 Task 5: [Within form] Spatially close, markup distant traversal

System	Keypresses Mean (s.d.)
CX	4.42 (.996)
CX+F	2.92 (.289)
Tabbing	29.75 (9.324)

Mean keystrokes differed statistically significantly between systems.  
( $F(1.017, 11.188) = 91.990, p < 0.001$ ).

CX keystrokes significantly fewer than Tabbing ( $p < .001$ ).

CX+F keystrokes significantly fewer than Tabbing ( $p < .001$ ).

CX+F keystrokes significantly fewer than CX ( $p = .002$ ).

Hypothesis 1 supported.

Hypothesis 2 supported.

### 6.3.6 Task 6: Low visual indicator of focus

System	Keypresses Mean (s.d.)
CX	6.42 (3.204)
CX+F	2.33 (.492)
Tabbing	48.50 (30.485)

Mean keystrokes differed statistically significantly between systems.  
( $F(1.012, 11.128) = 24.176, p < 0.001$ ).

CX keystrokes significantly fewer than Tabbing ( $p = .002$ ).

CX+F keystrokes significantly fewer than Tabbing ( $p = .001$ ).

CX+F keystrokes significantly fewer than CX ( $p = .001$ ).

Hypothesis 1 supported.

Hypothesis 2 supported.

### 6.3.7 Task 7: Scrolled element

System	Keypresses Mean (s.d.)
CX	5.83 (1.115)
CX+F	3.50 (1.087)
Tabbing	23.33 (11.155)

Mean keystrokes differed statistically significantly between systems.  
( $F(1.349, 14.837) = 1304.682, p < 0.001$ ).

CX keystrokes significantly fewer than Tabbing ( $p < .001$ ).

CX+F keystrokes significantly fewer than Tabbing ( $p < .001$ ).

CX+F keystrokes significantly fewer than CX ( $p = .001$ ).

Hypothesis 1 supported.

Hypothesis 2 supported.

### 6.3.8 Task 8: Link surrounded by links

System	Keypresses Mean (s.d.)
CX	6.67 (1.371)
CX+F	3.75 (.452)
Tabbing	48.67 (2.934)

Mean keystrokes differed statistically significantly between systems.  
( $F(1.387, 15.257) = 2296.627, p < 0.001$ ).

CX keystrokes significantly fewer than Tabbing ( $p < .001$ ).

CX+F keystrokes significantly fewer than Tabbing ( $p < .001$ ).

CX+F keystrokes significantly fewer than CX ( $p < .001$ ).

Hypothesis 1 supported.

Hypothesis 2 supported.



## **Chapter 7**

### **Conclusions & Future work**

# Chapter 8

## Figures

How many hours a day do you use a computer or smartphone?

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 4 - 6	2	16.7	16.7	58.3
7 - 9	5	41.7	41.7	100.0
13 +	5	41.7	41.7	41.7
Total	12	100.0	100.0	

Figure 8.1: Device usage by participant

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Age	12	21	28	22.58	2.151
Valid N (listwise)	12				

Figure 8.2: Age of participant

Gender of Participant

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Female	3	25.0	25.0	25.0
Male	9	75.0	75.0	100.0
Total	12	100.0	100.0	

Figure 8.3: Gender of participant

**Participant Keyboard Navigation Proficiency**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	I prefer a mouse/pointing device	5	41.7	41.7	41.7
	I use hotkeys in addition to a mouse/pointing device	6	50.0	50.0	100.0
	I prefer keyboard controls where possible	1	8.3	8.3	50.0
	Total	12	100.0	100.0	

**Figure 8.4: Keyboard Navigation Proficiency of Participants**

Participants were asked to pick which of these sentiments they identified best with. The statements attempted to describe a proficiency scale for keyboard usage:

1. I need a mouse/pointing device
2. I prefer a mouse/pointing device
3. I use hotkeys in addition to a mouse/pointing device
4. I prefer keyboard controls where possible
5. I strongly prefer full keyboard support

Votes were cast only for the middle descriptors, 2–4.

**Descriptive Statistics**

	N	Minimum	Maximum	Mean	Std. Deviation
Touch-Typing Proficiency	12	1	5	3.42	1.084
Valid N (listwise)	12				

**Figure 8.5: Touch-Typing Proficiency of Participants**

Touch-typing proficiency was rated on a five-point scale ranging from the sentiment ‘I look at every key before pressing’ to ‘I always type without even a reference glance at the keyboard’.

## **Chapter 9**

## **Appendices**