World Pointing: Improving Natural Pointing Interaction with Real-World Landmarks

Adrian Reetz
University of Saskatchewan
110 Science Place
Saskatoon SK S7N 5C9
+1 (306) 966–2327
Adrian.Reetz@USask.ca

Carl Gutwin
University of Saskatchewan
110 Science Place
Saskatoon SK S7N 5C9
+1 (306) 966–2327
Gutwin@CS.USask.ca

ABSTRACT

Selecting items from digital systems using full-body pointing gestures recently became popularized by mainstream devices such as the Wii and the Kinect. Although this kind of gesture has long been subject to research, we still do not know much about how to design pointing targets, which act as selection proxies, in order to make them easily learnable and quickly selectable for users. In this research note, we argue for the use of real-world objects as selection proxies and show that using them allows significantly better performance in terms of selection time, accuracy and learnability over virtual pointing targets such as Virtual Shelves and Air Pointing.

Keywords

Gestural interfaces, Real-world interaction, Human performance, Human memory

1. INTRODUCTION

Air-pointing selection techniques, i.e. techniques in which people use their arms to select system functionality through natural pointing, have long been researched (e.g., [4], [13], [9]) and are now used commercially as well (e.g., Nintendo Wii, Microsoft Kinect). However, little is known about how these systems should be designed, and how these designs perform in different scenarios. In particular, we do not know much about the conceptual design of selection proxies, i.e. pointing targets or areas to which system functionality is assigned. Properly designing selection proxies is important for the success of natural pointing techniques because careless design can lead to selection techniques that suffer from low learnability, low selection time, high selection error rate, user frustration, and subsequently the failure of the technique.

Most published research in this area focused on using virtual user-relative regions in space as selection proxies, for example, Virtual Shelves [9] and Air Pointing [6]. These techniques reported steep learning curves and low initial selection accuracies, which usually frustrates and discourages users. We suspect—and ultimately show—that pointing at abstract user-relative regions in space and learning their mapping to specific system functionality requires a lot of training. During this process, selection accuracy is lower than with the alternative pointing target design technique we introduce.

To address the shortcomings of abstract user-relative pointing targets we present “World Pointing”, which makes use of absolutely positioned real-world objects to provide pointing targets for system selections. With World Pointing, users map system functionality to landmarks, e.g., architectural features, real-world objects, or real-world regions. An example would be mapping the command “turn light on” to the ceiling lamp; pointing at the ceiling lamp would execute the command “turn light on”. The major advantage of landmarks over abstract user-relative targets is that learning landmarks involves more suitable systems of human memory. An analysis of the memory processes involved in remembering pointing targets lead us to the hypothesis that people should perform better with natural pointing techniques that use strong semantic associations (such as World Pointing), instead of ones that use weak associations (such as Air Pointing). Our results confirm this hypothesis by showing that people perform better with World Pointing in terms of learnability and selection accuracy, while maintaining selection time. We also show that pointing target recall after mundane tasks, such as moving around in an environment, is better with World Pointing. Using landmarks implies that people can use a particular set of mappings in one particular environment only, the one with the landmarks in; therefore, World Pointing is best used in static environments that are familiar to users, such as their living room, kitchen, or office.

GRAND 2013, May 14-16, 2013, Toronto, ON, Canada

In this article, we present three main findings:

1. Participants become proficient (accuracy and completion time) in a selection task with World Pointing substantially sooner than with Air Pointing
2. Participants are at least as fast using World Pointing as Air Pointing. With higher levels of training, people become significantly faster with World Pointing than Air Pointing
3. Participants retain a higher level of proficiency after moving around in an environment with World Pointing than with Air Pointing

2. RELATED WORK

2.1 Manipulation-based Pointing

The core idea of manipulation-based pointing techniques [3] is that interaction with digital systems happens through manipulation, mostly selection, of a proxy object, such as an on-screen icon; the WIMP desktop metaphor is a classic example. In contrast, sign-language-based pointing techniques let users issue commands with hand gestures instead of proxies, for example [7] [10]. Of interest for air-pointing selection techniques are mostly full-arm or full-body pointing gestures, which are recently popular with the Nintendo Wii and the Microsoft Kinect, although research goes back to 1980 [4]. Initially, researchers were mostly concerned with exploring what it means to use full-arm manipulation-based pointing techniques for user interfaces ([4], [8], [3]). Later, the focus shifted towards hardware and implementation details ([5], [12], [13]). Only recently, the community started to look into human-performance-related details, such as selection accuracy and selection time, e.g. [6], [9]. So far, very little research has been dedicated to more fundamental issues that influence user performance, most notably the underlying cognitive processes.

2.2 Environment-based Interaction & Storage

As mentioned above, selection proxies for screen-based desktop GUIs can be simple on-screen icons. When using full-arm pointing gestures in non-standard environments, such as domestic or ubiquitous computing, however, there is no reason why we cannot extend the notion of selection proxies to landmarks, i.e. anything users can point at, for example, abstract virtual regions in space, architectural features, real-world objects, or abstract real-world regions.

Although the idea of interacting with the physical world as a means to control interactive systems is not new (see Ubiquitous Computing, Tangible User Interfaces, and Augmented Reality) and although attempts have been made to assess users' physical pointing performance, e.g. [11], [13], there is very little work that examines the underlying cognitive user performance of natural pointing at real world objects to control interactive systems. One example is Virtual Shelves by Li et al. [9] who proposed an interaction technique that uses virtual locations as storage areas. Virtual Shelves are positioned relative to a user, and each shelf contains specific system functionality. An evaluation showed low selection time with Virtual Shelves, but it suffered from a high error rate. Cockburn et al. extended Li's idea with the Air Pointing Design Framework [6], a taxonomy for categorizing air-pointing selection techniques, but showed similar performance issues.

2.3 Human Memory Systems

Going into much detail about human memory systems exceeds the scope of this research note. Still, we want to mention two basic concepts because they are relevant to the models presented in the following section. First, many behavioral memory models incorporate the idea of associationism, which interprets memory as a collection of associations between stimuli and responses. Associations can be confabulated or created by experience; the nature of either stimulus or response is not restricted. People can use existing associations as device for remembering new pieces of information (mnemonics), and associations can be of different strength, i.e. some are easier to remember than others [2]. Second, people have excellent spatial memory. People can easily remember the location of hundreds of objects (though they may forget some of them occasionally). They do this by creating categorical spatial memory (i.e. knowing where objects are in relation to other objects or themselves) with very little effort through locomotion. Furthermore, they have a strong coordinate spatial memory (i.e. knowing how far objects are apart from each other) in familiar environments [1].

3. AIR POINTING VS WORLD POINTING

Air-pointing is a group of interaction techniques that uses physical full-body pointing gestures for making selections. With the Air Pointing Design Framework [6], Cockburn et al. set a useful foundation for exploring the design space of air-pointing techniques. The authors described three different techniques: ray-casting air-pointing (RCAP), 2D-plane air-pointing (2DAP), and 3D-volume air-pointing (3DAP).

With RCAP, digital items are stored into virtual pigeon holes on a two-dimensional plane. This plane can be imagined like a pattern of tiles on the wall where each tile is associated with a different digital item and pointing at a tile selects the associated digital item. With 2DAP and 3DAP, digital items are stored in in small virtual boxes, and users select an item by moving their hand into the appropriate box. These boxes form a two-dimensional array in 2DAP or a three-dimensional (meta) box in 3DAP.

The problem is, however, that users have to remember the association between a digital item (stimulus) and an invisible virtual box (response). Both Li [9] and Cockburn [6] reported that people have problems learning these associations. In our opinion, the reason for this a weak associative connection between stimulus and response: since there is little inherent meaning in "60° left, 30° down", people have difficulties to remember that this is the storage location for "TV on/off". More systematically, we believe that the choice of selection proxies makes learning the aforementioned air-pointing-techniques difficult.

For us, the solution to this problem is changing selection proxies from abstract regions in space to real-world objects. Because of real-world objects' countless properties (such as function, size, color, location, and history), they can invoke rich associations. Therefore, it will be easier for people to remember that "TV on/off" is stored at "TV screen". We call this new selection technique, in which people point at real-world objects to invoke the associated digital item, *World Pointing*. One major advantage of World Pointing is that people can use pre-existing associations in memory, which makes this technique easy to learn. If these associations are strong, people will also retain them over extended periods of time. One might argue that using real-world objects as proxies involves the danger of requiring people to visually search for the object when they want to make selections. However, in familiar environments, such as living rooms and offices, people have a detailed spatial memory model and a firm understanding of the location of these objects. Therefore, finding real-world proxy objects should not increase selection time significantly.

4. STUDY DESIGN

The goal of our study was to compare user performance with World Pointing (WP) and Ray-Casting Air-Pointing (RCAP). We were interested in confirming three hypotheses and answering one research question:

- People learn mappings faster with WP than with RCAP.
- People make less selections errors with WP than RCAP.
- People have no significantly higher selection time with WP than RCAP.
- Are users better able to maintain their level of performance with WP or RCAP after moving in the environment?

For the last question, we rotated participants by 90° at the end of the experiment. We will refer to this as "Trial-Rotated".

4.1 Apparatus

To perform the comparative study, we implemented a testing system that allowed users to use both techniques using unrestricted full-arm pointing gestures. The system used eight NaturalPoint OptiTrack S250e IR-tracking cameras to capture participants' location and orientation. To track participants' pointing gesture, we taped a small IR-reflector to their index finger.

We set up both WP and RCAP in one section of our research lab. Participants faced a 42" (105 cm) TV screen that displayed the user interface. During all phases except Trial-Rotated (see below), participants were standing approximately 8.5' (2.6 m) away from the screen. For Trial-Rotated, we displayed the interface on a 20" (50 cm) computer screen that was approximately 4.5' (1.4 m) away. Thus, participants had a similar viewpoint in both phases.

4.1.1 *World Pointing (WP)*

World Pointing uses landmarks to store digital items, and users perform ray-casting-style pointing without system feedback to select items.

The system must be aware of the users' location, the direction of their pointing gesture, and the location of landmarks in the environment. In our implementation, we modeled landmarks through a simple four-tuple

$l \coloneqq \{X, Y, Z, \alpha\}$ (location and angular size of the landmark).

Given users' location and pointing direction $in = \{X, Y, Z, \Psi, \Theta\}$ the system calculates the pointing-ray $\overrightarrow{\mathrm{V_p}}$ and the vector between the user and the landmark $\overrightarrow{\mathrm{V_{ul}}}$.

If $\measuredangle\left(\overrightarrow{\mathrm{V_p}}, \overrightarrow{\mathrm{V_{ul}}}\right) < \alpha$ then l is selected.

4.1.2 *Ray-Casting Air-Pointing (RCAP)*

Our implementation uses a semi-circular 2D arrangement of virtual shelves.

The algorithm we used for RCAP is rather simple. Shelves are defined by the four-tuple $\mathrm{s} \coloneqq \{\Psi_{min}, \Psi_{max}, \Theta_{min}, \Theta_{max}\}$

If the input angles $in = \{\Psi_{in}, \Theta_{in}\}$ satisfies $\Psi_{min} < \Psi_{in} < \Psi_{max} \wedge \Theta_{min} < \Theta_{in} < \Theta_{max}$ then shelf s is selected.

4.2 Stimuli & Targets

We created two sets of 14 digital items in order to avoid learning effects between selection techniques. Although no item appeared in both sets, we tried to keep the lists comparable. The use of the item sets was counterbalanced between selection techniques.

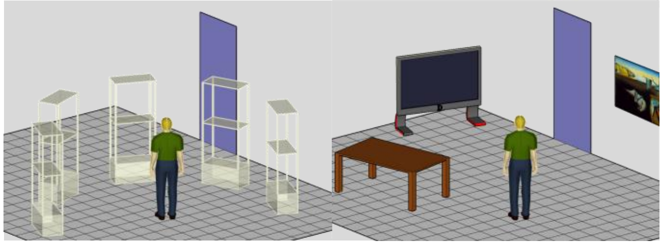

Figure 1: RCAP (left) and WP (right)

We arranged the virtual shelves for RCAP in a 7 × 2 semi-circular pattern. Subsequently, the shelves had $\Delta\Psi = 30°$ width; we decided to limit the height of each shelf to $\Delta\Theta = 60°$, so that the total size of each of the seven racks was $\Delta\Theta = 120°$. (See Figure 2 for the final shelf setup.) Since there is no obvious mapping between digital items and virtual shelves, we simply tried grouping similar items together and stored them in the same rack, for example, "Simpsons" and "Game of Thrones" (both TV shows) and "volume up" and "volume down" (both device commands). After this, we picked 14 landmarks for WP to which we mapped the two sets of digital items. We tried to pick these objects so that they would also form a 7 × 2 pattern similar to the virtual shelves because we wanted participants to perform very similar pointing gestures for both selection techniques; we tried to minimize effects caused by fatigue and pointing ability as much as possible. We then mapped the digital objects from both lists to the 14 landmarks. (See Figure 2 for the final landmark setup.) The differences between left and right are caused by the rectangular shape of the shelves in RCAP and the circular shape of landmarks in WP. We set the radius of landmarks to 34° to achieve the same overall shelf- and landmark-size for RCAP and WP.

The order in which we presented the two selection techniques to our participants was counterbalanced using Latin square.

4.3 Procedure & Data Analyses

For our experiment, we asked participants to perform multiple selections of 14 different digital items. A set of 14 selections was called a block; within a block, all 14 digital items were selected exactly once; the order within a block was separately randomized for each block. Overall, participants had to complete 15 blocks for a total of 210 selections. Each block belonged to a certain phase and served a specific purpose: demonstration, training, or trial.

Table 1: Length of Phases

	Demo	Training #1	Trial #1	Training #2	Trial #2	Training #3	Trial #3	Training #4	Trial #4	Trial-rotated	
Length (blocks)	**1**	**1**	**1**	**2**	**1**	**2**	**1**	**2**	**2**	**2**	

For our evaluation, we only considered data collected during trial-phases. To determine the effect of the selection technique on participant's performance we analyzed the main trial phase and the rotated phase separately. For the main trial phase the analysis consisted of a 5 × 2 (Block by Technique) RM-ANOVA, for the rotated phase 2 × 2 (Block by Technique). Post-hoc tests used Bonferroni-correction for all between-block and between technique analyses.

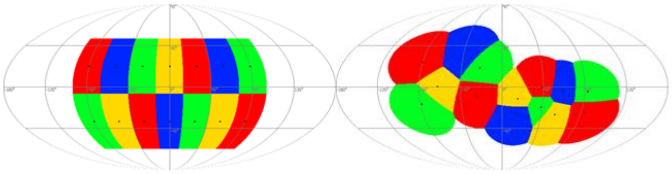

Figure 2: RCAP shelves (left), WP landmarks (right), both in Mollweide projection

We recruited 12 participants (8 male, 4 female; M=25.5 years, all right-handed) from a local university. We paid participants a $10 honorarium for participating in our one-hour-long study.

5. RESULTS – ACCURACY & LEARNING

5.1 Main Testing Phase (Trial #1 – Trial #4-2)

For the main testing phase, there was a main effect of technique on accuracy ($F(1{,}11) = 19.6, p < .001$), with participants having significantly higher accuracy with WP ($M = 92.3\%, SD = 2.0\%$) than with RCAP ($M = 72.1\%, SD = 5.0\%$). Pair-wise block-analysis revealed that participants were significantly more accurate with WP from Trial #1 through Trial #4-1 ($p < .001$, $p < .01$, $p < .05$, $p < .05$). There was no significant difference for Trial #4-2. (See Figure 3)

There was also a main effect of block on accuracy with participants ($F(1{,}11) = 49.1, p < .001$). When comparing the first ("Trial #1") and the last ("Trial #4-2") block for each selection technique separately, we found that participants were significantly more accurate when using RCAP ($p < .001$) but not for WP ($p > .05$).

There was a significant interaction between selection technique and block number ($F(1{,}11) = 11.6, p < .01$).

5.2 Effect of Rotation (Trial #4-2 – Trial rotated-1)

There was a main effect of technique on completion time ($F(1{,}11) = 10.8, p < .01$), with WP performing significantly faster than RCAP over both trials.

There was also a main effect of block ($F(1{,}11) = 16.0, p < .01$). For both techniques, average selection times increased after rotating participants ("Trial rotated-1") (RCAP: $+.39s$; WP: $+.31s$). Figure 3 illustrates the slight increase for selection time during the rotated trial phase.

There was no observed interaction effect between technique and block ($F(1{,}11) = 0.4, p > .05$).

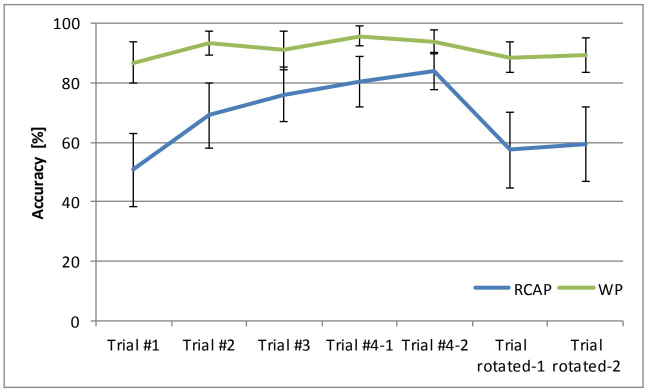

Figure 3: Overall accuracy

5.3 Recovery from Rotation (Trial rotated-1 – 2)

When examining the trials after rotation there was a main effect of technique on completion time ($F(1{,}11) = 8.2, p < .05$), with completion times being lower for WP than for RCAP.

There was also a main effect of block on completion time ($F(1{,}11) = 7.1, p < .05$), with participants' completion times slightly lower in Trial rotated-2 than in Trial rotated-1.

There was no interaction between selection technique and block number ($F(1{,}11) = 0.7, p > .05$).

6. RESULTS – SELECTION TIME

Due to the limited space of this research note, we cannot go into any detail here. During no phase was WP slower than RCAP.

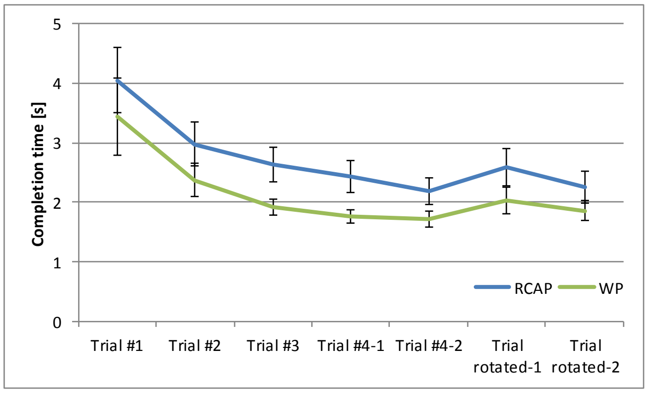

Figure 4: Overall completion time

7. DISCUSSION & CONCLUSION

In our study, we confirmed all three of our hypotheses: People can learn World Pointing faster than ray-casting air-pointing, people can initially make selections more accurately with WP than with RCAP, and people are not slower using WP than RCAP.

The answer to our research question (Are users better able to maintain their level of performance with WP or RCAP…?) surprised us. Since participants had to perform the exact same task in RCAP after being rotated (the virtual shelves are body-centric, thus rotate with the user), we did not expect accuracy for RCAP to drop that dramatically. Although we cannot explain this finding conclusively, we can try giving a possible explanation. We assume that participants do not conceptualize RCAP as virtual shelves. Some participants seemed to associate the virtual shelf with the real-world object that is located behind it ("rather than remember shelves, I mentally assigned positions to the names, e.g. 'Grand Theft Auto' is played on the TV"). This behavior would give a good explanation on the drop in performance after we rotated our participants because the environment objects obviously did not rotate and all existing associations become invalid.

Our study showed that the proper design of selection proxies, when creating selection techniques based on full-arm pointing gestures, is necessary for learnable and frustration-free interaction techniques. Based on our study and due to the associative richness of real-world objects, we recommend them for use as selection proxies.

8. FUTURE WORK

For future work, we want to gain a better understanding about the memory processes involved in air-pointing techniques work. We are particularly interested in the transition between semantic memory and procedural memory for RCAP. Furthermore, we want to explore the design space of WP to some greater detail and investigate aspects such as the effects of number of items on selection accuracy and recall rate after extended periods of time.

9. REFERENCES

1. Allen, G.L. *Human Spatial Memory: Remembering Where*. Psychology Press, 2003.
2. Anderson, J.R. and Bower, G.H. *Human Associative Memory*. Psychology Press, 1980.
3. Baudel, T. and Beaudouin-Lafon, M. Charade: Remote control of objects using free-hand gestures. *CACM 36*, 7 (1993).
4. Bolt, R.A. Put-that-there: Voice and gesture at the graphics interface. *GRAPH '80*.
5. Cao, X. and Balakrishnan, R. VisionWand: Interaction techniques for large displays using a passive wand tracked in 3D. *UIST '03*.
6. Cockburn, A., et al. Air pointing: Design and evaluation of spatial target acquisition with and without visual feedback. *IJHCS 69*, 6 (2011).
7. Gustafson, S., et al. Imaginary interfaces: Spatial interaction with empty hands and without visual feedback. *UIST '10*.
8. Krueger, M.W., et al. VIDEOPLACE: An artificial reality. *CHI '85*.
9. Li, F.C.Y., et al. Virtual shelves: Interactions with orientation aware devices. *UIST '09*.
10. Mistry, P., et al. WUW - wear Ur world: A wearable gestural interface. *CHI EA '09*.
11. Myers, B.A., et al. Interacting at a distance : measuring the performance of laser pointers and other devices. *CHI '02*.
12. Wilson, A. and Pham, H. Pointing in intelligent environments with the WorldCursor. *INTERACT '03*.
13. Wilson, A. and Shafer, S. XWand: UI for intelligent spaces. *CHI '03*.