HAL
open science

Mobile Pointing Task in the Physical World: Balancing Focus and Performance while Disambiguating

William Delamare, Céline Coutrix, Laurence Nigay

To cite this version:

William Delamare, Céline Coutrix, Laurence Nigay. Mobile Pointing Task in the Physical World: Balancing Focus and Performance while Disambiguating. Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI 2013), 2013, Munich, Germany. pp.89-98, 10.1145/2493190.2493232 . hal-00953356

HAL Id: hal-00953356

https://inria.hal.science/hal-00953356v1

Submitted on 28 Feb 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mobile Pointing Task in the Physical World: Balancing Focus and Performance while Disambiguating

William Delamare, Céline Coutrix and Laurence Nigay
CNRS, Joseph Fourier University UJF-Grenoble 1
Grenoble Informatics Laboratory (LIG), UMR 5217, Grenoble, F-38041, France
{William.Delamare, Celine.Coutrix, Laurence.Nigay}@imag.fr

ABSTRACT
We address the problem of mobile distal selection of physical objects when pointing at them in augmented environments. We focus on the disambiguation step needed when several objects are selected with a rough pointing gesture. A usual disambiguation technique forces the users to switch their focus from the physical world to a list displayed on a handheld device's screen. In this paper, we explore the balance between change of users' focus and performance. We present two novel interaction techniques allowing the users to maintain their focus in the physical world. Both use a cycling mechanism, respectively performed with a wrist rolling gesture for P2Roll or with a finger sliding gesture for P2Slide. A user experiment showed that keeping users' focus in the physical world outperforms techniques that require the users to switch their focus to a digital representation distant from the physical objects, when disambiguating up to 8 objects.

Author Keywords
Mobile interaction; Physical interaction; Pointing; Disambiguation; Focus; Performance.

ACM Classification Keywords
H.5.2. Information interfaces and presentation (e.g. HCI): User Interfaces – Interaction styles.

General Terms
Human Factors; Design; Experimentation; Performance.

INTRODUCTION
Mobility gives users the opportunity to interact with the physical world around them, anywhere and anytime. A mobile user carrying a handheld device within smart environments can initiate interaction with many objects. In order to avoid unnecessary travel for selecting one of these objects, a usual solution is to point at them from a distance [1, 29, 31].

Two metaphors exist for physical pointing: ray casting (e.g. laser-based systems) and volume selection (e.g., Infrared-based systems). While studied in Virtual Environment (VE), the physical world has specificities that make existing techniques difficult to transpose:

- The location of physical targets is not easily modifiable by the system, making difficult the spatial reconfiguration of targets as the Flower Ray does [11].
- The location of the pointer representation (e.g., the laser dot) is mapped to the pointing device location (e.g., the laser device) with a fixed scale, making difficult the use of improved Control-Display Gain [12].
- There is no visual representation of the pointing tool. For instance, when pointing with a laser, there is no representation of the ray. Thus, techniques like the Flexible Pointer are difficult to transpose [9].

Yet, ray casting and volume pointing metaphors have been transposed to mobile physical interaction. However, as expected, pointing with a bare laser has accuracy limitations due to limited precision [12, 15]. Jitters cause ±5° deviation [10], i.e., 17cm when interacting at a distance of 2m, possibly causing an error. The existing techniques that address these problems have other limitations: (1) the management by the system of a model of the physical world [14, 34], restricting the dynamicity of the physical world (e.g. moving a lamp has to be reported in the system by the user), and (2) invasive devices, such as large sensors [16] or a platform attached to the ceiling [34].

Instead of using a ray, techniques based on the selection volume metaphor are good alternatives to avoid accuracy problems. However, several objects may fall in the volume selection. Thus, a disambiguation step is needed in order to refine the selection.

Some existing disambiguation techniques use heuristics in order to automatically disambiguate which target in the volume the user intended to select: In a spatial-geometric approach [19], a metric is used to automatically select the target closest to the volume axis. While avoiding additional action from the user for the disambiguation step and clearly letting the users focus on the primary physical target, heuristics can fail in some particular situations. For instance, if the scoring mechanism is based on the size of the intersection between objects and the selection volume, users would not be able to select a small object located right in front of a large one. In addition, the automatic predictive

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
MobileHCI 2013, Aug 27–30, 2013, Munich, Germany.
Copyright 2013 ACM xxx-x-xxxx-xxxx-x/xx/xx-xx....$10.00.

mechanism can prevent the users from understanding why they perform incorrect selections, and thus adjust their manipulation. As a consequence, in this paper we address explicit disambiguation mechanisms.

The explicit disambiguation step can be performed either by keeping users' focus on the physical target, or by forcing them to switch their focus to a digital representation distant from the physical objects, such as a list of objects' names displayed on the screen of a handheld device. In this paper, we investigate techniques that allow maintaining of the focus on the physical world.

Fitts law [3] usually helps understanding of the perceptual-motor factors of pointing tasks and therefore is used to enhance pointing techniques. Such modeling would be useful for focus-keeping physical disambiguation techniques. The law has been transposed to pointing at distant on-screen targets [13]. This work models a direct, ray-casting pointing technique with no disambiguation. As a consequence, it does not always apply to a volume selection technique where a disambiguation step might be needed when several targets are preselected. The law has also been transposed to pointing at distant physical targets through a phone used as a Magic Lens [22, 23]. On the contrary, we aim for our users not to look at the screen but to maintain their focus on the physical target itself. As a consequence, we aim at a different perceptual-motor task and its Fitts modeling needs to be further studied. Prior to doing so, we would like to assess if such techniques are worth considering in the first place since alternate focus-changing techniques already exist. Towards this aim, we investigate in this paper the tradeoff between change of users' focus and performance for novel explicit disambiguation techniques.

The contributions of this work are twofold: (1) we propose two explicit disambiguation techniques that facilitate users' focus in the physical world, namely Physical Pointing Roll (P2Roll) and Physical Pointing Slide (P2Slide), and (2) we experimentally explore the balance between change of focus and performance, considering our two techniques and a baseline technique displaying a list of objects' names on a handheld screen.

We first review existing explicit disambiguation techniques for mobile selection of physical objects. Next, we present design dimensions for disambiguation techniques in the physical world before describing our two mobile disambiguation techniques P2Roll and P2Slide. We then report an in-lab experiment focusing on the balance between change of the user's focus and performance. We conclude with a discussion of our results and directions for future work.

RELATED WORK

We build on previous work on volume pointing techniques for mobile interaction that include an explicit disambiguation step. We distinguish techniques that maintain the users' focus in the physical world from techniques that imply a switch of the user's focus from the physical world to the mobile device.

The GesturePen proposes a way to point at physical objects using an IR beam [29]. It combines custom tags attached to physical objects and a custom handheld stylus for pointing. Even if the prototype did not propose any disambiguation mechanism if several tags respond to the IR broadcast, authors reported that a user suggested a dial on the device as an explicit control of the beam angle and/or length. However, an IR beam is not visible and thus difficult to control without visual feedback [31]. Interestingly, the authors evaluated the GesturePen (selection in the physical world) against a selection in a graphical list displayed on a handheld device's screen, forcing the user to change her/his focus of attention. However, the users had to get close enough to the physical object in order to read its tag. On the contrary, we want to allow the user to interact from a distance. Moreover only the pointing step has been evaluated versus the list, but not the disambiguation step, as the latter was not implemented.

Other solutions use both physical and digital representations, the digital representation being co-localized with, or closely located to, physical objects.

First, PICOntrol combines light sensors attached to physical objects and an encoded projection in order to select and control smart objects [27]. Since several sensors may fall in the projected image, authors proposed introduction of a way to optically or digitally control the size of the projection, thus keeping the users' focus on physical objects. However, this solution has not been implemented or tested.

Second, the "Radio Frequency Identity and Geometry Lamps" system (RFIG Lamps) also proposes an explicit disambiguation technique with co-localized physical and digital representations [21]. The system uses two sequential pointing gestures: a rough one for volume selection, followed by a precise disambiguating one. First, to allow a coarse first pointing gesture, the pointing device embeds an RFID reader. An encoded projection allows the system to know the positions of all the objects awoken in the volume by the RFID broadcast. The selected objects are highlighted by projecting tag areas (digital part) onto the physical objects (physical part). Then, users can manipulate a laser embedded in the pointing device for selecting the desired tag by clicking in the projected tag area that is bigger than the original tag. The key aspect of the technique is that it maintains the users' focus on the physical targets. Unfortunately, the system was designed for a warehouse scenario, in which augmented objects are cardboard boxes that suit well for projection. We cannot assume that a mobile user can always find such projection-compliant surfaces in any augmented environment. Moreover, the experiment revealed that both the projection and the disambiguation step using the laser-based input, suffer from hand-jitters.

Other solutions use mixed representations through handheld Augmented Reality (AR) techniques: users perform a rough pointing gesture, positioning the handheld device until physical objects appear on the screen via a live video. In this case, the selection volume is the field-of-view of the embedded camera. A first solution for disambiguating the selection is to provide an on-screen cursor, to select the virtual representation of objects through the video [22]. A second solution is to break the spatial mapping between the digital representation on the screen and the actual physical target. Techniques such as Shift&Freeze [33] propose to freeze the video in order to ease the disambiguating touch gesture. AR techniques include a first pointing step that does not provide a set of selected objects, but rather a view of a subpart of the physical surroundings. Thus, they force users to go through the disambiguation step (e.g., selecting the target on the screen), even if there is a single target in the selected subpart of the physical surroundings (i.e. in the video).

Finally, two techniques proposed a digital representation of the selected objects for the disambiguation step [1, 31]: a list of targets' names or images displayed on a handheld device's screen when they are in the selected volume. Only one technique has been evaluated [31], but authors only reported users' satisfaction when selecting objects in the list. On the contrary, we want to explore in detail this disambiguation step and compare the technique based on a list on screen with focus-maintaining techniques.

To our knowledge, there are no experimental studies of the benefits and drawbacks of focus switching on performance while disambiguating a selection in the physical world. In addition and as highlighted in this section, the few disambiguation techniques maintaining the users' focus on the physical world (GesturePen, PICOntrol) chose to loose the benefits of volume selection when disambiguating by returning to a ray casting technique and its potential drawbacks. This is the challenge we addressed by designing our two techniques P2Roll and P2Slide: instead, our design rationale is based on maintaining the easy volume pointing gesture while exploring navigation techniques in this volume for disambiguation.

Before describing our two techniques, we present two of the design dimensions for disambiguation techniques [8] that are determinant of the users' focus of attention and performance.

DESIGN RATIONALE: DIMENSIONS FOR DISAMBIGUATION TECHNIQUES

Two aspects determine the design of explicit disambiguation techniques: (1) the representation of the selected objects determining the focus switching or not, and (2) the input method impacting on the users' performance. As the pointing step already uses a gesture, we focus on consistent gestures as input methods.

Representation of Selected Objects

As illustrated in the previous section, the representation of objects is organized along a physical-digital continuum. Between these two extremes are mixed representations, where both physical and digital elements are used to represent objects of interest. When a digital representation is involved, users may need to change their focus of attention. In this case, spatial relationships between objects' representations can be studied [32]. Five spatial relationships are defined. Among these, for instance the RFIG Lamp technique provides a digital representation of tags that *overlay* the primary physical objects. On the contrary, the list provides digital representations that are *separated* from the physical objects. In the latter case, users have to switch their focus between these representations.

We consider this design aspect of representation as important, since it has an impact on the usability of the technique and implies a change of focus of attention.

Physical Gesture Properties

The limb chosen for inputs influence the overall performance of the interaction techniques. Indeed, there is a tradeoff between speed and accuracy that a limb is able to perform, depending on:

- The limb's bandwidth [6], i.e. the rate of information that can be transmitted by the limb,
- The jitters users may have during the interaction with this limb [7].

But, as stated in a previous study [2], because of many differences in tasks, protocols, border effects such as learning effect between conditions, or muscles group involved in the experiment, it is difficult to clearly state which limb is best suited for a particular task. However, since the pointing step already involves an arm gesture, we focus on the next options following the kinematic chain of the upper limb: wrist and fingers. It appears that even with non-comparable results, wrist [2] and fingers [36] have been reported as good input candidates.

In the next section, we present our two disambiguation techniques designed to keep users' focus on physical targets with wrist and finger inputs, previously identified as good candidates.

P2ROLL AND P2SLIDE

We present two mobile disambiguation techniques designed for the selection of physical objects. We assume that augmented physical objects provide visual feedback to the users, for instance with Light-Emitting Diodes (LEDs) commonly attached to the physical augmented objects [29].

One technique uses a wrist rolling gesture (Physical Pointing Roll – P2Roll), and one technique uses a finger sliding gesture (Physical Pointing Slide – P2Slide). Without exiting comparative results on the two input methods, we decided to explore both options.

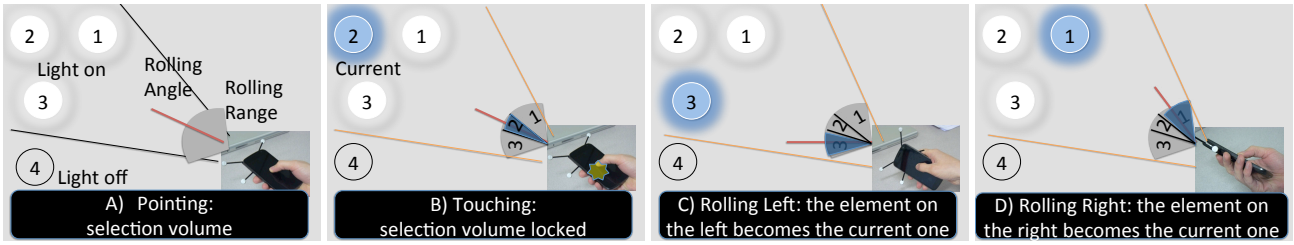

Figure 1: P2Roll walkthrough. A) The pointing step. B) The beginning of the disambiguation step with a touch gesture. The current element is '2' since the rolling angle is in its area. C) The user performs a rolling gesture to the left, so that '3' becomes the current element. D) The user performs a rolling gesture to the right extremity, so that the current element goes from '3' to '1' by traversing '2'.

Quasi-Mode

Both techniques use sequential actions performed while holding a handheld device: pointing and disambiguating. The disambiguation step can be seen as a quasi-mode: the user enters this mode by touching and holding her/his finger on the handheld device's screen, and exits the mode (i.e. validate her/his selection) by releasing her/his finger from the screen.

If the desired target is not in the pointed volume when entering the disambiguation mode (i.e. error during the pointing step), the user can exit the disambiguation mode without validating the selection by sliding her/his finger to the bottom of the touch screen. In doing so, on finger lift, the selection is not triggered and the pointing direction is unlocked.

Physical Pointing Roll: P2Roll

This disambiguation technique uses a wrist rolling gesture for browsing the set of pre-selected objects (Figure 1). After the pointing gesture (Figure 1, A), users can lock the pointing volume by touching the device's screen (Figure 1, B). At this point, objects are associated with a range of rolling angles, so that when the wrist angle is in a particular range, the corresponding current object provides feedback to the user. Nothing is displayed on the handheld device's screen, the user focuses only on the physical objects and their feedback. Once on the desired target, the user can validate the selection by removing the thumb from the touch screen.

A previous study [18] reported a comfortable rolling range of 130˚ while holding a laser, in [-70˚, 50˚]. But holding a laser can be different from holding a smartphone. So we conducted a pilot study (7 participants) showing that right-handed participants were comfortable holding the smartphone in the motion range [-80˚, 50˚]. We used mirrored values for left-handed people. The discretization of the rolling range was quadratic (Figure 1, B C D), since it has been demonstrated to allow more precision than a linear one [20]. We empirically optimized this discretization and then validated it through pilot experiments. Discretized parts were then linked with physical objects, linearly ordered on the horizontal axis (Figure 1, B C D).

We expect that expert users could anticipate the rolling angle for a given physical target, so that when locking the pointing direction, the desired target would be the current element (i.e. selected if the selection is validated). This would increase performance since only a fast 'touch and release' gesture on screen would be needed for the selection of the desired target. In a real-world scenario, we expect this expert behavior for recurrent situations e.g., selecting the TV from the couch.

Physical Pointing Slide: P2Slide

This disambiguation technique uses a horizontal sliding gesture with the thumb on the device's touch screen (Figure 2). It has been shown that horizontal thumb gestures on a one-handed device have an equivalent or better throughput

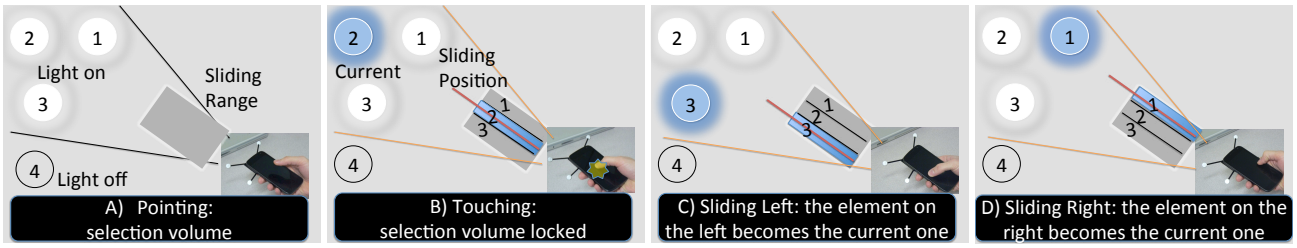

Figure 2: P2Slide walkthrough. A) The pointing step. B) The beginning of the disambiguation step with a touch gesture. The current element is '2' since the thumb is in its area. C) The user performs a sliding gesture to the left, so that '3' becomes the current element. D) The user performs a sliding gesture to the right extremity, so that the current element goes from '3' to '1' by traversing '2'.

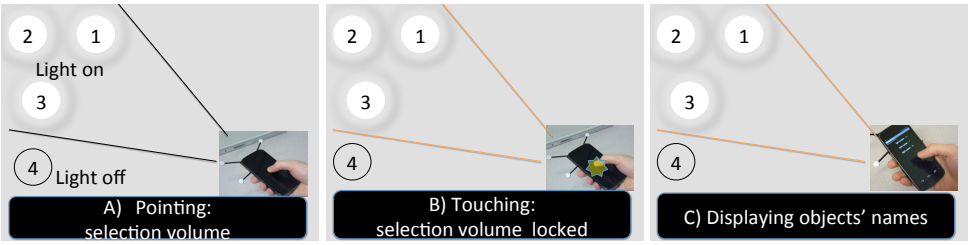

Figure 3: Illustration of List. A) The pointing step. B) The beginning of the disambiguation step with a touch gesture. C) The list of objects' names displayed on screen.

than vertical ones [35]. As P2Roll, no feedback was provided on the device: browsing was performed with an input on the smartphone and an output on physical augmented objects. After the pointing gesture (Figure 2, A), users can lock the pointing volume by touching the device's screen (Figure 2, B). At this point, objects are associated with a vertical area of the screen. All vertical areas are equally distributed along the horizontal axis of the touch screen. Each area is associated with a pre-selected physical object, linearly ordered on the horizontal axis, so that when the finger is detected in an area, the corresponding object provides feedback to the user (Figure 2, B, C and D).

As for P2Roll, we expect the same anticipation for expert users: validating the selection with a fast 'touch and release' on the remembered area of the screen.

EXPERIMENT

The aim of this experiment was to explore the relationship between change of users' focus and selection performance. We therefore studied the differences in performance of two classes of disambiguation techniques:

- Techniques that imply a switch of users' focus from the physical targets to a digital representation of targets: in the experiment we evaluated a commonly used technique: a list of objects' names displayed on a handheld screen (baseline technique).
- Techniques that keep users' focus on the physical targets: in the experiment we evaluated our two techniques P2Roll and P2Slide.

We chose to evaluate two techniques that keep users' focus on physical objects in order to control the effect of the input method. Having two focus-maintaining techniques also allowed us to explore the benefits and drawbacks of both wrist rolling and finger sliding inputs.

Performance of the selection techniques was evaluated by selection time and error rate. This evaluation took place along several difficulty levels: an increase of the density of physical objects is expected to augment the selection task's difficulty, as it was expected to augmented the search task's difficulty using AR techniques [24]. However, as the tasks are different, the two experiments cannot be compared.

Baseline: List

As reported in the related work section, a list displaying the names of pre-selected objects on a handheld device's screen is a commonly used technique for disambiguating the pointing selection. Thus, we chose this technique for the baseline of our experimental study.

First, the user points at a target (Figure 3, A). Once the selection volume is locked by tapping on screen (Figure 3, B), the user is able to look at the screen without the need to maintain the position of her/his arm. All pre-selected objects are displayed on the handheld device's screen (Figure 3, C). Then, the user can either directly select the desired target's name by tapping on it, or scroll the list until the name appears on screen and then tap on the desired target's name. If the first pointing gesture needs to be canceled (i.e. the desired target is not in the selection volume), users can touch the default 'return' button of the phone and retry the pointing selection.

We chose to represent physical objects by their names, since (1) images may be misinterpreted if several objects are similar, and (2) textual names can easily be ordered to help with the selection task. This choice therefore favors the list technique. We used the standard vertical list widget and the standard size of Android items: the user therefore needs to scroll when the list contains more than 8 items.

Hypotheses

For the experiment, we formulated the three following hypotheses. Two hypotheses concern the comparison of performance between focus-maintaining vs. focus-switching techniques and the third one concerns the comparison of performance of our two techniques:

- H1a: The disambiguation techniques that keep users' focus on the physical targets have better performance than the disambiguation techniques that switch users' focus to a digital representation of the physical targets.
- H1b: Users prefer focusing on the physical objects while disambiguating the pointing selection.
- H2: In the class of disambiguation techniques that keep users' focus on physical targets, the disambiguation technique using a smaller finger sliding gesture (P2Slide) outperforms the disambiguation technique using a larger wrist rolling gesture (P2Roll).

Apparatus

Targets were Ping-Pong balls embedding RGB Light Emitting Diodes (LEDs) controlled through Arduino[1] microcontrollers. LEDs were placed in Ping-Pong balls for light diffusion purposes, and numbers (i.e., names of the physical targets) were written on Ping-Pong balls (Figure 4, right) to be used by the List disambiguation technique. Numbers were used so that names could be read from afar and quickly found in the ordered list on screen. Targets were randomly assigned a number, as in the real world objects are not necessarily placed in the alphabetical order. There was also no pattern in the disposition of targets.

Only the density of targets influences the performance of all three disambiguation techniques, thus the size of Ping-Pong balls or the distance from Ping-Pong balls are not relevant once in the disambiguation step. In order to provide different densities, sets of 2, 4, 8 and 16 light balls were fixed on 4 different 31 cm × 22 cm frames (Figure 4, right). The same densities were used to evaluate AR search techniques [24]. We chose 16 objects as a maximum since previous work found that it is the upper bound of the wrist rolling inputs [20]. This limit also suits the P2Slide technique: accuracy with the index finger on a surface could rise up to 150dpi, i.e. targets of 0.17mm [4]. However, these results are for the index and not for the thumb. The maximum of 16 targets requires sectioning of the screen into areas of width of about 3.5mm, thus 20 times larger than the index limit. This is sufficient for the thumb. We also assessed that 16 objects in a selection volume represent a realistic scenario, e.g. when pointing in a smart home. Previous studies used 4 [25, 31] or 5 [29] objects. Interviewing 13 subjects, it appears that up to 6±2 (mean ± standard deviation) objects are likely to be disambiguated in their home.

For tracking the position and orientation of the smartphone in our experiment, we used the OptiTrack[2] system using 5 V100:R2 cameras and the Arena[3] software. The system provided 100Hz, 6 degrees of freedom (DoF) tracking of markers attached to the device and ran on a 3.06Ghz Core 2 duo workstation.

In order to avoid a form factor and to have consistency between techniques, all three techniques used the same smartphone as an input device. The device was a Galaxy Nexus I9250 smartphone, used to capture touch inputs on the screen and to display the list of targets' numbers.

The three techniques were implemented in python and ran on a 2.2 GHz Intel Core i7 computer.

[1] http://www.arduino.cc/
[2] http://www.naturalpoint.com/optitrack/
[3] http://www.naturalpoint.com/optitrack/products/arena/

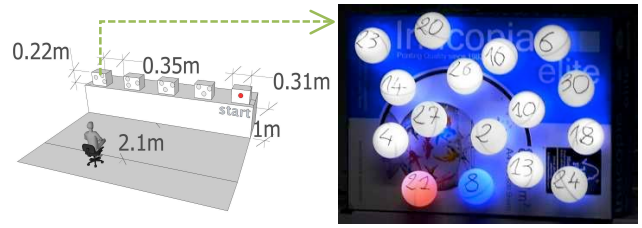

Figure 4: Left: setup of the experiment. Right: a frame with 16 Ping-Pong balls embedding RGB LEDs. All (white) balls are in the selection volume; #21 (red) is the target of the current trial; #8 (blue) is the current element.

Participants

Six male and six female unpaid volunteers, ranging in age from 22 to 33 years old (26.75±2.70), participated in the experiment. One participant was left-handed, and 8 used tactile smartphone on a daily basis. All participants controlled the input device with their dominant hand.

Procedure

The task in the experiment was to select a physical object, namely a Ping-Pong ball with an embedded LED. Users were sitting in front of the 4 frames exactly in the middle, at a distance of 2.1m (Figure 4, left). Although the techniques are mobile, the users were sitting in order to minimize the fatigue since the experiment was very long in time and to fully control the distance to the targets. A special frame, with only one object corresponding to the 'start target', was placed on the utmost right of the four other frames. Frames were positioned at 35 cm distance from each other.

The trial begins by pointing at a red 'start target'. For subjects to be as fast as possible, there was no constraint about the initial angle of the wrist or the initial finger position on the screen. Once pointed at, the 'start target' turned green, meaning that its selection could be validated. After the subject validated the 'start target', another target turned red, randomly chosen among the 30 other targets. When pointing at a frame or a subpart of the frame, all targets of the frame were switched on in white, meaning that they were in the selection volume. Our pilot study showed that users tended to only aim at a subset of the frame when the target was on the border, thus creating new uncontrolled densities and scattering data. In order to avoid this uncontrolled parameter in our experiment, frames were completely selected even if they were not completely inside the selection volume. This design choice allowed us to ignore the size of the selection volume and to control the number of selected items in the experiment.

When beginning the disambiguation step, feedback was provided:

- In the physical world for the P2Roll and P2Slide conditions, a target turned blue in the pre-selected volume of white targets in order to indicate what the

current element was. When the current element was on the desired red target, the target turned from red to green.

- On the device's screen for the List condition. A sound feedback was additionally provided by the smartphone to notify the user about the appearance of the list of targets' number on screen. This was done in order to guide visual attention to the screen of the mobile phone [30].

Once the disambiguation was validated (either right or wrong), a sound indicated that the trial was complete and the 'start target' turned red again for the subject to launch the next trial.

Design

We used a repeated measure within-participant design. The independent variables were the disambiguation technique T (List, P2Roll, P2Slide) and the density D (2, 4, 8, 16).

The experiment lasted approximately 90 minutes per participant, and was divided into 3 sections corresponding to the three techniques. The order of the disambiguation techniques was counterbalanced between subjects. A Latin square design was used for balancing the different positions of the target frames between subjects, so that all 4 frames happened to be at the 4 different distances from the 'start target'. Between each section, users were asked to fill a qualitative questionnaire about the technique they just used. Each section consisted of 6 blocks: 3 training blocks and 3 blocks for measuring performance. The 3 blocks of training aimed at (1) avoiding the "wow effect" that occurs during our pilot study when users first manipulate new interaction techniques, (2) recording data from "equally expert" users, since some participants were already used to select an item in a list, and (3) allowing users to learn the new techniques, in particular the horizontal/angular mapping between physical objects and Roll and Slide inputs as it may be difficult, e.g., for objects 10 and 13 figure 4, right. Each block consisted of 40 random selections (10 for each density condition). This design resulted in 3 techniques × 3 blocks × 40 selection tasks, i.e. 360 trials per participant and 4320 trials in total.

RESULTS

The main dependent measures for the task were trial completion time and error rate (for H1a and H2). In order to obtain a deeper insight into the differences between the techniques, we also considered their footprint. The footprint of a device is the length of the path the device went through during a trial. This is a commonly used measure for evaluating fatigue while manipulating a device [11].

A total of 1.58% of the trials were removed as outliers (data off by at least three standard deviation from the mean). There was no need to test for learning or order effect since the techniques and the target frames were counterbalanced.

We performed non-parametric tests on our data, instead of ANOVA, since (1) a Shapiro-Wilk normality test showed that there was a violation of the normality assumed for parametric tests ($p < 0.0001$), and (2) a Bartlett's test showed a violation of homogeneity of variances for all factors ($p < 0.0001$).

Keeping users' focus on the physical target is efficient (H1a)

The experiment confirmed H1a: using Friedman's non-parametric tests, we found significant main global effects for T [$\chi^2(2) = 129.9$, $p < 0.0001$] and D [$\chi^2(3) = 1166.1$, $p < 0.0001$] on completion time (Table 1). A Post-Hoc test using Mann-Whitney tests with Bonferroni correction revealed that, as expected, the density D has a significant effect on completion time for each technique ($p < 0.001$). Moreover, the performance time of disambiguation techniques maintaining users' focus on the physical target is best for low densities, i.e. 2, 4 and 8 ($p < 0.0001$). There was no statistical difference between techniques in the density 16 ($p > 0.05$).

		P2Roll	P2Slide	List
D = 2	*MT*	1.56 ± .62	1.43 ± .62	1.86 ± .60
	ER	3.35	2.78	3.64
D = 4	*MT*	1.79 ± .68	1.83 ± .78	1.98 ± .47
	ER	6.15	5.01	3.39
D = 8	*MT*	2.11 ± .71	2.12 ± .81	2.23 ± .54
	ER	8.81	12.50	2.29
D = 16	*MT*	2.71 ± .83	2.71 ± .91	2.80 ± .82
	ER	17.20	19.60	4.84

Table 1: Movement Time (*MT* in seconds, *m* ± *sd* gives the mean *m* and the standard deviation *sd*) and Error Rate (*ER* in %) by technique for each density.

However, when taking into account error rates, Chi-square tests revealed that the density D had significant effects on P2Roll [$\chi^2(3) = 39.1$, $p < 0.0001$] and P2Slide [$\chi^2(3) = 55.7$, $p < 0.0001$] only (Table 1). P2Roll and P2Slide show an important increase of incorrect selections (from ~3% for density 2 to ~20% for density 16). On the contrary, the density does not impact the error rate for List (between 2% and 5%).

We further refine the error analysis by considering the error types. First, errors can happen when the desired target is not in the pointing selection (see Table 2, "Out" values). 16% of P2Roll incorrect selections were such errors and 22% for P2Slide (against 14% for List). Two observations during the experiment indicate a possible explanation:

- Some users tended to lock the selection volume faster when keeping their focus in the physical world, thus they were selecting the wrong frame. This is confirmed with

the rate of cancel action per technique: 6.3% for P2Roll, 2.7% for P2Slide and 0.5% for List.

- The cancel action was more difficult for P2Roll and P2Slide (sliding the thumb to the bottom of the screen) than for List (touching the default cancel button).

Thus, both the rush for locking the volume selection and the cancel difficulty led to more incorrect selections for P2Roll and P2Slide than for List.

Second, and more importantly, when errors happen at the disambiguation step (i.e., within the right selection volume), two kinds of errors happen: before or after reaching the desired target (see Table 2, "Before" and "After" values). Both P2Roll and P2Slide have ~60% of their errors happening *right after* reaching the target. The validating release of the finger happened on average 0.51s after the loss of the desired target for P2Roll, and 0.75s for P2Slide. One possible explanation is that the validating mechanism (i.e., release the finger from screen) is tightly coupled with the disambiguation selection (i.e., rolling or sliding) for P2Roll and P2Slide.

	Out	Before	After
P2Roll	16%	18.4%	**65.6**%
P2Slide	22.86%	18.57%	**58.57**%

Table 2: Percentage of type of error for P2Roll and P2Slide. Most errors happen after that the desired target became the current element.

Finally another explanation for the error performance is the arm fatigue over extended usage. List has a smaller footprint (18.58±9.44 cm) than P2Roll (26.73±13.24 cm) and P2Slide (24.82±13.39 cm). Friedman's tests revealed significant global effect of T [$\chi^2(2) = 217.9$, $p < 0.0001$]: users did smaller arm movements while selecting in the list than while gesturing with the wrist or even with the finger. This could explain that subjects became more tired and made more errors with P2Roll and P2Slide than with List.

Avenues for improving the P2Roll and P2Slide techniques are threefold. First, different cancel actions of the pointing step must be studied. Second, accuracy in high densities using P2Roll could be improved adapting RapMenu [26]: a rolling range would be then linked to more than one physical object, and pinch gestures would be mapped to the selected objects of a given rolling range. This solution could address the wrist accuracy limit factor. Third, decoupling the trigger from the disambiguation selection might be proposed for high densities. For instance, the trigger of P2Slide could be done using a 'tilt-forward' as for TouchOver [28]: indeed, the decoupling of the trigger action from the selection action facilitates avoidance of last minute unwanted movements.

To conclude on time and error performance related to H1a, keeping users' focus in the physical world while disambiguating a pointed volume is more time efficient. Yet, in order to decrease the error rate in high densities of the focus-maintaining techniques, P2Roll and P2Slide, we plan to explore different cancel actions of the pointing step as well as to study a validating mechanism decoupled from the selection mechanism as described above. Such a validating mechanism as well as a correct cancel action of the pointing step could prevent ~70% of errors of P2Roll and P2Slide.

Users prefer keeping their focus on the physical target (H1b)

The second hypothesis was that users would prefer to maintain focus on the physical targets when disambiguating a pointed volume. This is validated since 10 out of 12 participants reported a preference for either P2Roll (5 participants) or P2Slide (5 participants) and ranked List last (only 2 preferred List). A more detailed analysis shows that even if users found many benefits to List (e.g., they felt confident, they found it easy to learn and use), they also found it the least enjoyable technique to use.

It also appears from the analysis of the interviews that 7 participants mentioned negatively the switching of focus, saying, e.g., "*It is not cool to look at the screen and loose eye-contact with the balls*" and 4 participants mentioned negatively the feeling of having extra-actions to perform (such as reading or scrolling), saying, e.g. "*It is tiresome to have to* read". Yet, some participants were aware of the accuracy benefit of having the list representation (5 participants), saying for instance "*I do not have to do precise movements*".

To conclude on H1b, users prefer techniques that keep their focus of attention on physical objects, even though they are aware of drawbacks such as the lake of precision for the highest density. This confirms that it is worth studying focus-maintaining techniques such as P2Roll and P2Slide.

Smaller gestures are not always more efficient (H2)

Our hypothesis H2 was that the smaller gestures of P2Slide would outperform larger gestures of P2Roll. Indeed, a Post-Hoc test using Mann-Whitney tests with Bonferroni correction revealed that completion time of P2Slide was better only for density 2 ($p < 0.0001$, $r = .048$), and equivalent to P2Roll for all the other densities. We explain this result by the fact that smaller movements are more efficient only if the task is easy. But when the difficulty increases, the smaller movements need to be performed with care and the time difference between small and large movements disappears.

Regarding error rates, Chi-square tests revealed no significant differences between P2Roll and P2Slide, whatever the density is [$\chi^2(1) = 0.8491$, $p > 0.1$]. Yet, the wrist movements of P2Roll seem to be more tiring: when analyzing the interviews, we found that physical fatigue has been reported mostly for P2Roll (7 times for P2Roll, 3 times for P2Slide and 4 times for List). However, as

previously reported, P2Roll and P2Slide have been ranked as the first preferred technique by 5 participants each.

To conclude on H2, P2Slide has better performance only for density 2. Based on both quantitative and qualitative data analysis, P2Roll and P2Slide can be considered as equivalent for the higher densities (4, 8 and 16 targets).

TOWARD DISAMBIGUATING IN THE REAL WORLD

In contrast to our experiment, objects in the real world can occlude each other in two ways. First, we voluntarily omit total occlusion since the resulting lack of feedback for the occluded object is out of the scope of this study involving a motor-*perceptual* task. Second, partial occlusion still enables left/right relationships to be defined. E.g. if a large object is behind a smaller one and the former is visible all around the latter, the former can be arbitrarily defined "on the left". This arbitrary mapping has to be learned during training, as users do in real settings.

In contrast to our experiment, the implementation of the techniques would have to be adapted for a real-world setting, e.g., with an RFID reader embedded in the device and hidden sensors attached to physical objects.

CONCLUSION AND FUTURE WORK

Mobility gives users the opportunity to select and interact with physical objects around them, anywhere and anytime, thanks to handheld devices. In this context, our contributions are two-fold. First we have presented two new selection techniques – P2Roll and P2Slide – that keep users' focus directly on physical objects, unlike GUI or AR techniques. Second we conducted an in-lab experiment mainly showing that keeping users' focus on the physical target is efficient and preferred by users when disambiguating a selection. We showed that the gain in performance is valid when up to 8 objects are included in the pointed selection volume. For the higher density of objects (16 objects in the selection volume), the focus-maintaining and focus-switching techniques are equivalent. We were able to identify a convergence at the density of 16 objects as our experiment considers a higher maximum density than previous studies (4 physical objects were used in [25, 31] and 5 objects in [29]).

In future work, in order to go beyond this experiment validating two new focus-maintaining techniques, three paths are now to be explored.

First, we plan to improve the performance of our techniques. To do so, we will explore discrete gestures instead of continuous ones (e.g. swipe instead of slide for P2Slide) and 2D gestures instead of 1D ones. In a pilot experiment we conducted, users gave positive feedback on an early prototype of the extended P2Slide technique with 2D (horizontal and vertical) finger movements.

Second, we plan to provide a better understanding of movement time by extending Fitts law to focus-maintaining techniques. This will allow our knowledge of these techniques to catch up with current knowledge of GUI or AR techniques for pointing in the physical world. Doing so will require a subsequent study, controlling the use of the techniques in order to be able to model them, e.g. the initial rolling angle for P2Roll. Contrastingly, the study presented here to validate focus-maintaining techniques allowed users to be as efficient as possible, e.g. by rotating the wrist in advance as users can do in real settings. Indeed, such 2-step interaction techniques can benefit from parallelization as for the GUI Rake-Cursor technique [5]. Fitts modeling of our techniques through a dedicated experiment will allow a better understanding of their benefits.

Third, we plan to extend our work by getting closer to real-life scenarios.

- We will investigate how users adapt themselves to high densities in real life. For instance, from which density users would rather walk closer to their target to decrease the density in the selected volume?
- We will investigate targets without visual feedback, for instance providing an audio list of pointed objects.
- We will investigate techniques for control, which usually follows the selection of an object. An interesting direction is to adapt Control Menus [17] to mid-air gesture. Doing so, users will be able to perform both the selection and the command (e.g. dimming a light) in a single, quick gesture while maintaining their focus on the physical target.
- After conducting in-lab experiments, we will conduct an in-vivo experiment of the P2Roll and P2Slide techniques. Indeed, it is common to first perform in-lab evaluation [23], and then validate results in the real world [22]. The concrete task will be a lighting source selection in the context of a restaurant and of a hotel suit.

ACKNOWLEDGMENTS

This work has been supported by the DELight project (French government's FUI -Single Inter-Ministry Fund-program) led by Schneider Electric.

REFERENCES

1. Ailisto, H., Pohjanheimo, L., Välkkynen, P., Strömmer, E., Tuomisto, T., and Korhonen, I. Bridging the physical and virtual worlds by local connectivity-based physical selection. *Personal and Ubiquitous Computing 10*, 6 (2006), 333–344.
2. Balakrishnan, R. and MacKenzie, I.S. Performance differences in the fingers, wrist, and forearm in computer input control. *Proc. CHI '97*, ACM (1997), 303–310.
3. Balakrishnan, R. "Beating" Fitts' law: virtual enhancements for pointing facilitation. *International Journal of Human-Computer Studies 61*, 6 (2004), 857–874.
4. Bérard, F. and Rochet-Capellan, A. Measuring the linear and rotational user precision in touch pointing. *Proc. ITS '12*, ACM (2012), 183–192.

5. Blanch, R. and Ortega, M. Rake Cursor : Improving Pointing Performance with Concurrent Input Channels. *Proc. CHI '09*, ACM (2009), 1415–1418.

6. Card, S.K., Mackinlay, J.D., and Robertson, G.G. A morphological analysis of the design space of input devices. *ACM TOIS 9*, 2 (1991), 99–122.

7. Davis, B.J. and O'Connell, J. Shoulder, Elbow and Wrist Components of Physiologic Tremor Amplitudeas Measured Using a Laser Penlight. *European Neurology 43*, 3 (2000), 152–154.

8. Delamare, W., Coutrix, C., and Nigay, L. Designing Disambiguation Techniques for Pointing in the Physical World. Proc. *EICS '13*, (2013). To appear.

9. Feiner, A. The flexible pointer: An interaction technique for selection in augmented and virtual reality. *Proc. UIST '03*, ACM (2003), 81–82.

10. Forsberg, A., Herndon, K., and Zeleznik, R. Aperture based selection for immersive virtual environments. *Proc. UIST '96*, ACM (1996), 95–96.

11. Grossman, T. and Balakrishnan, R. The design and evaluation of selection techniques for 3D volumetric displays. *Proc. UIST '06*, ACM (2006), 3–12.

12. König, W., Gerken, J., Dierdorf, S., et al. Adaptive Pointing–Design and Evaluation of a Precision Enhancing Technique for Absolute Pointing Devices. *Proc. INTERACT '09*, Springer (2009), 658–671.

13. Kopper, R., Bowman, D. a., Silva, M.G., et al. A human motor behavior model for distal pointing tasks. *International Journal of Human-Computer Studies 68*, 10 (2010), 603–615.

14. De la O Chávez, F., Fernández de Vega, F., Olague, G., et al. An independent and non-intrusive laser pointer environment control device system. *Proc. ICPS '08*, ACM (2008), 37–46.

15. Myers, B.A., Bhatnagar, R., Nichols, J., et al. Interacting at a distance: measuring the performance of laser pointers and other devices. *Proc. CHI '02*, ACM (2002), 33–40.

16. Patel, S. and Abowd, G. A 2-way laser-assisted selection scheme for handhelds in a physical environment. Proc. *UbiComp '03*, Springer (2003), 200–207.

17. Pook, S., Lecolinet, E., Vaysseix, G., and Barillot, E. Control menus: excecution and control in a single interactor. *Extended Abstracts CHI '00*, ACM (2000), 263–264.

18. Qin, Y., Shi, Y., Jiang, H., et al. Structured laser pointer: enabling wrist-rolling movements as a new interactive dimension. *Proc. AVI '10*, ACM (2010), 163–166.

19. Rahman, A.S.M.M., Hossain, M.A., and Saddik, A. El. Spatial-geometric approach to physical mobile interaction based on accelerometer and IR sensory data fusion. *ACM TOMCCAP 6*, 4 (2010), 1–23.

20. Rahman, M., Gustafson, S., Irani, P., et al. Tilt techniques: Investigating the Dexterity of Wrist-based Input. *Proc. CHI '09*, ACM (2009), 1943–1952.

21. Raskar, R., Beardsley, P., Van Baar, J., et al. RFIG Lamps: Interacting with a Self-Describing World via Photosensing Wireless Tags and Projectors. *ACM TOG 23*, 3 (2004), 406–415.

22. Rohs, M., Oulasvirta, A., and Suomalainen, T. Interaction with magic lenses: real-world validation of a fitts' law model. *Proc. CHI '11*, ACM (2011), 2725–2728.

23. Rohs, M. and Oulasvirta, A. Target acquisition with camera phones when used as magic lenses. *Proc. CHI '08*, ACM (2008), 1409–1418.

24. Rohs, M., Schleicher, R., Schöning, J., et al. Impact of item density on the utility of visual context in magic lens interactions. *Personal and Ubiquitous Computing 13*, 8 (2009), 633–646.

25. Rukzio, E., Broll, G., Leichtenstern, K., et al. Mobile interaction with the real world: An evaluation and comparison of physical mobile interaction techniques. *Proc. AmI '07*, Springer (2007), 1–18.

26. Ryan, T.N., McMahan, P., and Bowman, D.A. Tech-note: rapMenu: Remote Menu Selection Using Freehand Gestural Input. *Proc. 3DUI '08*, IEEE (2008), 55–58.

27. Schmidt, D., Molyneaux, D., and Cao, X. PICOntrol: using a handheld projector for direct control of physical devices through visible light. *Proc. UIST '12 2*, ACM (2012), 379–388.

28. Scoditti, A., Vincent, T., Coutaz, J., et al. TouchOver: decoupling positioning from selection on touch-based handheld devices. *Proc. IHM '11*, ACM (2011), 1-4.

29. Swindells, C., Inkpen, K.M., Dill, J.C., et al. That one there! Pointing to establish device identity. *Proc. UIST '02*, ACM (2002), 151–160.

30. Talsma, D., Senkowski, D., Soto-Faraco, S., et al. The multifaceted interplay between attention and multisensory integration. *Trends in cognitive sciences 14*, 9 (2010), 400–410.

31. Välkkynen, P., Niemelä, M., and Tuomisto, T. Evaluating touching and pointing with a mobile terminal for physical browsing. *Proc. NordiCHI '06*, ACM (2006), 28–37.

32. Vernier, F. and Nigay, L. A framework for the combination and characterization of output modalities. *Proc. of DSV-IS '00*, Springer (2000), 32-48.

33. Vincent, T., Nigay, L., and Kurata, T. Precise pointing techniques for handheld Augmented Reality. *Proc. INTERACT'13*, (2013). To appear.

34. Wilson, A. and Pham, H. Pointing in intelligent environments with the worldcursor. Proc. *INTERACT '03*, Springer (2003), 495–502.

35. Wobbrock, J.O., Myers, B. a., and Aung, H.H. The performance of hand postures in front- and back-of-device interaction for mobile computing. *International Journal of Human-Computer Studies 66*, 12 (2008), 857–875.

36. Zhai, S., Milgram, P., and Buxton, W. The influence of muscle groups on performance of multiple degree-of-freedom input. *Proc. CHI '96*, ACM Press (1996), 308–315.