



**HAL**  
open science

# Ubiquitous interaction: mobile, multi-display and freeform interfaces

Marcos Serrano

► **To cite this version:**

Marcos Serrano. Ubiquitous interaction: mobile, multi-display and freeform interfaces. Human-Computer Interaction [cs.HC]. Université Paul Sabatier Toulouse III, 2018. tel-01959335v1

**HAL Id: tel-01959335**

**<https://hal.science/tel-01959335v1>**

Submitted on 18 Dec 2018 (v1), last revised 7 Feb 2019 (v2)

**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.

UNIVERSITÉ PAUL SABATIER - TOULOUSE 3

## HABILITATION À DIRIGER DES RECHERCHES

Présentée et soutenue publiquement par

**Marcos Serrano**

le 15 Novembre 2018

Spécialité: Informatique - Interaction Homme-Machine

### **Ubiquitous interaction: mobile, multi-display and freeform interfaces**

#### JURY

Caroline APPERT	CNRS - LRI	Présidente
Michel BEAUDOUIN-LAFON	Université Paris-Sud	Rapporteur
Stephen BREWSTER	University of Glasgow	Rapporteur
Andy COCKBURN	University of Canterbury	Rapporteur
Emmanuel DUBOIS	Université Toulouse 3	Examineur
Antonio KRUGER	Saarland University	Examineur

Habilitation à Diriger des Recherches préparée au sein de l'Institut de Recherche  
en Informatique de Toulouse





# Acknowledgements

First of all, I want to thank Michel Beaudouin-Lafon, Stephen Brewster and Andy Cockburn for accepting to review this manuscript. I also want to thank the members of my jury, Caroline Appert and Antonio Kruger.

Special thanks to Emmanuel Dubois for his help in preparing this HDR, and more generally for his support since I arrived to the University of Toulouse. It has been a fruitful collaboration so far.

The work presented in this manuscript results from multiple collaborations and I want to acknowledge my co-authors involvement, especially Pourang Irani, Eric Lecolinet, Yves Guiard, Christophe Jouffrais, Anne Roudaut, Barrett Ens, Xing-Dong Yang, Louis-Pierre Bergé, Andrés Lucero and Mathieu Raynal.

Some of the presented contributions have been carried by three PhD students that I co-supervised at the University of Toulouse, Sandra Bardot, Housseem Saidi and Gary Perelman. It has been a pleasure working together and I believe I have learned as much as them during this journey.

Thanks to the ELIPSE group members, with whom I have shared many discussions that surely have an impact on the work presented here, especially Bernard Oriola, Marc Macé and Antonio Serpa.

Finally, thank you Anne-Cécile, Félix and Carmen for filling my life with love and joy.



# Abstract

My research activities have focused on solving some of the main challenges of three types of ubiquitous displays: mobile, multi-display, and freeform interfaces.

By 2017, the number of smartphone users in the world was beyond 2 billions, i.e. one third of the world population. Not only the number of smartphones has dramatically increased, but also their computing capabilities and hence the number of available applications and complex data to manipulate. This raises numerous challenges on how to easily and rapidly interact with such growing quantity of data on mobile platforms. We present four major contributions in this context: we created a novel type of gestures called Bezel-Tap Gestures to solve the problem of rapidly launching commands from sleep mode; we proposed using on-body gestures on the face, i.e. hand-to-face input, to facilitate navigating spatial data on head-worn displays; we explored mobile true-3D displays to facilitate mobile interaction with volumetric data; and finally we studied using the smartwatch output capabilities to facilitate non-visual exploration of spatial data for visually-impaired people.

Multi-display environments (MDEs) are common on desktop environments and more and more common on professional and public environments. However they also bring new challenges due to the complexity of the overall system, which can be composed of multiple and heterogeneous devices, arranged in dynamic spatial topologies. We addressed two major challenges in MDEs: the need for fluid interactions in MDEs and complex data exploration using multiple monitors. To tackle the first challenge we adopted two approaches: either using an existing device, in our case a head-worn display interface that we named Gluey, or creating a novel dedicated device, namely TDome. To tackle the second challenge, i.e. facilitate the exploration of complex data on MDEs, we studied the use of around-device gestures to manipulate 3D data on public display and extended the overview+detail interface paradigm with multiple detailed views to explore multiple regions of the data simultaneously.

Finally, to fulfill the adoption of pervasive displays, we need to facilitate a seamless integration of displays into existing environments, ranging from in-vehicle and wearable displays to public displays. Traditional rectangular displays are not well-suited for these applications, as they can not be easily integrated. Emerging technologies allow for the creation of non-rectangular displays with unlimited constraints in shapes. With this eminent adoption comes the urgent challenge of rethinking the way we present content on non-rectangular displays. Our approach was threefold: first, we carried focus groups to gather both concrete usage scenarios and display shapes; then, we studied text content only, being the fundamental brick of any UI; finally, we explored more complex content layouts combining text and icons.

To address the previously presented research challenges, we had to face the methodological challenges of designing, prototyping and evaluating interaction. Traditional design and evaluation methods quite often failed at properly answering the previous research challenges. As a result, every new project came up with the additional challenge of rethinking our approach and research methods at multiple levels: design, prototyping and evaluation. For this reason, I deem of importance to include a chapter on research methods.



# Contents

<b>Acknowledgements</b>	<b>i</b>
<b>Abstract</b>	<b>iii</b>
<b>Contents</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Towards ubiquitous interaction . . . . .	1
1.1.1 The advent of mobile computing . . . . .	1
1.1.2 Multi-display environments . . . . .	2
1.1.3 Displays everywhere . . . . .	2
1.2 Our contributions . . . . .	2
1.2.1 Mobile interactions with complex data . . . . .	3
1.2.2 Multi-display tools and interfaces . . . . .	4
1.2.3 FreeForm interfaces . . . . .	5
1.3 Methodological challenges . . . . .	6
1.4 Manuscript overview . . . . .	6
<b>2 Mobile interactions with large data spaces</b>	<b>9</b>
2.1 Research challenges . . . . .	9
2.1.1 Gestural interaction as mobile input . . . . .	9
2.1.2 Offscreen and eyes-free mobile output . . . . .	10
2.1.3 Summary . . . . .	11
2.2 Command selection from sleep mode on tablets through Bezel-Tap gestures . . . . .	11
2.2.1 Bezel-Tap Gestures . . . . .	12
2.2.2 Bezel-Tap Menus . . . . .	13
2.2.3 Summary . . . . .	15
2.3 Exploring spatial data on head-worn displays using hand-to-face input . . . . .	15
2.3.1 Hand-to-face gestures . . . . .	15
2.3.2 Spatial navigation with hand-to-face input . . . . .	16
2.3.3 Social acceptability of hand-to-face input . . . . .	17
2.3.4 Summary . . . . .	18
2.4 Mobile True-3D Displays . . . . .	18
2.4.1 Display properties . . . . .	18
2.4.2 Spatial configuration . . . . .	19
2.4.3 Direct touch interaction . . . . .	19
2.4.4 Summary . . . . .	20
2.5 Mobile map exploration for visually impaired people . . . . .	20
2.5.1 Interaction principles . . . . .	20
2.5.2 Non-visual map exploration techniques . . . . .	21
2.5.3 Experimental results . . . . .	22
2.5.4 Summary . . . . .	22
2.6 Chapter conclusion . . . . .	23

<b>3</b>	<b>Multi-Display tools and interfaces</b>	<b>25</b>
3.1	Research challenges . . . . .	25
3.1.1	Fluid interaction in MDEs . . . . .	25
3.1.2	Exploring complex data on multiple displays . . . . .	26
3.1.3	Summary . . . . .	28
3.2	Gluey: fluid interaction in MDEs using a head-worn display . . . . .	28
3.2.1	Usage Scenario . . . . .	28
3.2.2	Gluey key components . . . . .	29
3.2.3	Gluey implementation . . . . .	30
3.2.4	Summary . . . . .	31
3.3	TDome: fluid interaction in MDEs using a dedicated multi-DoF device . . . . .	31
3.3.1	Usage Scenario . . . . .	31
3.3.2	TDome key components . . . . .	32
3.3.3	Evaluation . . . . .	33
3.3.4	Summary . . . . .	34
3.4	Interaction with Overview+Detail Interfaces on 3D Public Displays . . . . .	34
3.4.1	Interaction techniques for translating the detail view in 3D . . . . .	35
3.4.2	Experimental results . . . . .	35
3.4.3	Summary . . . . .	36
3.5	Splitting Detailed Views in Overview+Detail Interfaces . . . . .	36
3.5.1	Interface design . . . . .	36
3.5.2	Experimental results . . . . .	38
3.5.3	Summary . . . . .	38
3.6	Chapter conclusion . . . . .	39
<b>4</b>	<b>Towards Displays Everywhere: FreeForm Interfaces</b>	<b>41</b>
4.1	Research challenges . . . . .	41
4.2	Collecting usage scenarios . . . . .	43
4.2.1	Data collection . . . . .	43
4.2.2	Display shape properties . . . . .	43
4.3	Displaying text on freeform displays . . . . .	44
4.3.1	Mapping framework . . . . .	44
4.3.2	Quantitative studies on reading text . . . . .	45
4.4	Laying out content on freeform displays . . . . .	46
4.4.1	Elicitation study . . . . .	46
4.4.2	Composition principles . . . . .	47
4.4.3	Experimental studies . . . . .	48
4.5	Chapter conclusion . . . . .	50
<b>5</b>	<b>Research methods for ubiquitous interaction</b>	<b>53</b>
5.1	Design methods . . . . .	53
5.1.1	In-the-wild observations . . . . .	53
5.1.2	Research through design by device composition (*contribution) . . . . .	55
5.1.3	Elicitation studies: exploiting similarity over agreement (*contribution) . . . . .	55
5.1.4	Design probes . . . . .	57
5.2	Prototyping methods . . . . .	58
5.2.1	Rapid prototyping tools (*contribution) . . . . .	58
5.2.2	Proof-of-concept prototypes . . . . .	58
5.3	Evaluation methods . . . . .	62
5.3.1	Component-based Wizard-of-Oz (*contribution) . . . . .	62
5.3.2	Pairwise comparisons . . . . .	63
5.3.3	Conclusion . . . . .	64
<b>6</b>	<b>Perspectives</b>	<b>65</b>

<i>Contents</i>	vii
6.1 From digital to physical ubiquitous displays . . . . .	65
6.2 Immersive data analytics to explore spatio-temporal data . . . . .	67
6.3 Pervasive freeform interfaces . . . . .	68
<b>Bibliography</b>	<b>69</b>
<b>Selected publications</b>	<b>79</b>





# Chapter 1

## Introduction

### Chapter content

---

1.1	Towards ubiquitous interaction . . . . .	1
1.2	Our contributions . . . . .	2
1.3	Methodological challenges . . . . .	6
1.4	Manuscript overview . . . . .	6

---

### 1.1 Towards ubiquitous interaction

Ubiquitous computing was a term coined by Mark Weiser around 1988 and presented in his 1991 paper "The computer for the 21st Century" [Weiser, 1991]. In this paper, Weiser exposes his vision that personal computers should evolve to make computing "an integral, invisible part of people's lives". While we are not yet there, recent advances in computing are making this vision come true.

#### 1.1.1 The advent of mobile computing

One of the major evolutions of personal computing in the last decade has been the one leading **from desktop computers to mobile devices**. I started my PhD in 2007 with the goal to develop an approach to prototype the combination of multiple interaction modalities (such as touch, gestures, etc.), i.e. what is called Multimodal Interaction. It soon became evident that multimodal interaction was not limited to desktop environments: that same year the arrival of two major innovations, iPhone and Android, generalized the use of multitouch and gestural mobile interaction. By 2017, the number of smartphone users in the world was beyond 2 billion, i.e. one third of the world population. Not only the number of smartphones has dramatically increased, but also their computing capabilities and hence the number of available applications. A 2012 survey showed that Android users have an average of 177 applications on their phones, using 73 of them occasionally and 32 very often [Shin *et al.*, 2012]. This raises numerous challenges on how to easily and rapidly interact with such growing quantity of data.

Interestingly, Weiser stated that "ubiquitous computing [...] does not mean just computers that can be carried to the beach, jungle or airport." [Weiser, 1991]. In other words, the ubiquitous computing vision can not be limited to the use of smartphones in a mobile context. The next step in the evolution towards ubiquitous computing is the emergence of **wearable devices**, namely smartwatches, head-mounted displays or jewellery. These devices are slowly becoming available for widespread daily use through lighter form factors.

They leverage mobility and always-available interaction, which are necessary to put in effect the concept of ubiquitous interaction. However, despite their interesting form factors, wearables devices have limited interaction capabilities, which makes it difficult using them for non trivial interactions, such as panning and zooming on a map.

### 1.1.2 Multi-display environments

Multi-display environments (MDEs) are common on desktop environments, where usually two or three displays are positioned alongside to extend the overall display space. Beyond this now ordinary setup, MDEs are also common on some professional environments, such as control rooms. More recently, multi-displays environments have become available on public environments, such as train stations, airports or malls, through the use of chained displays [Ten Koppel *et al.*, 2012].

When combined, the mobile and multi-display phenomenon result in even more complex environments, which have been called Mobile Multi-Display Environments (MMDEs) [Cauchard *et al.*, 2011]. These environments provide unique opportunities, as mobile devices can be employed to interact with these larger displays. However they also bring new challenges due to the complexity of the overall system, which can be composed of multiple and heterogeneous devices, arranged in dynamic spatial topologies (i.e. some devices/displays can move while others remain static).

### 1.1.3 Displays everywhere

The pervasive need to access and make decisions based on data is ushering another evolution in ubiquitous computing: displaying interactive dynamic content where and when needed, a vision termed as displays everywhere [Pinhanez, 2001]. This vision covers a broad set of use cases, as these **pervasive displays** can be leveraged to show energy consumption data in-situ in smart environments; facilitate collaborative pedagogy in university classes; leverage interactive presentations in museums, showrooms or shop fronts; or facilitate planning and decision-making during professional meetings.



Figure 1.1: The three types of ubiquitous interfaces that we explored in our research projects.

## 1.2 Our contributions

My research activities have focused on solving some of the main challenges of the three previous types of ubiquitous interfaces (Figure 1.1): mobile, multi-display, and pervasive freeform interfaces.

### 1.2.1 Mobile interactions with complex data

Our work on mobile interaction addressed both the challenges of how to extend the limited input capabilities of mobile devices, particularly concerning gestural interaction, and how to access information beyond the display space, i.e. offscreen or eyes-free output (Figure 1.2).

#### Novel gestures for mobile input

During my 1-year post-doctoral fellowship at Telecom ParisTech (Paris, France), we addressed the issue of how to facilitate micro-interactions with a very large number of commands when the mobile device is asleep. Since mobile devices constantly switch to sleep mode to save energy, interaction is hampered by the need to reactivate them whenever they have gone to sleep, typically by pressing a physical button and sliding a widget on the screen. We addressed this problem through the development of a novel interaction technique, Bezel Tap [Serrano *et al.*, 2013], allowing micro-interactions on mobile devices. Not only can the technique serve to open an application and launch a favorite command rapidly, it can also wake the device if asleep.

During my 1-year post-doctoral fellowship at the University of Manitoba (Winnipeg, Canada), we addressed the issue of how to access commands and facilitate 2D interaction on head-worn displays. Currently, HWDs provide onboard microphones and small capacitive sensors for user input. Voice recognition is useful for command-based tasks such as for search queries but has limited use in certain settings (i.e. noisy environments). The capacitive surface on the temple of HWDs presents a viable on-device method for input, but it has limited input space. We proposed hand-to-face input as a novel, alternative method for interacting with HWDs [Serrano *et al.*, 2014a]. We define hand-to-face input as any gesture that involves touching, rubbing, scratching or caressing the face.

#### Offscreen and eyes-free mobile output

During my post-doctoral year at the University of Manitoba we also explored the vision of mobile volumetric (or true-3D) displays, i.e. displaying volumetric 3D content in mid-air around a mobile phone. This can be accomplished through different means: stereoscopic displays with head-tracking, optical illusions, moving parts or augmented reality glasses. The challenge is that we possess limited knowledge on both the numerous constraints imposed on viewing and interacting with mobile true-3D interfaces and the usage scenarios suitable for such displays. To this end, we studied the various factors that can potentially influence the effective deployment of true-3D on mobile devices in an emulated environment [Serrano *et al.*, 2014b].

While mobile content is difficult to access for sighted people, they are inherently inaccessible to visually impaired (VI) users. Visually impaired users wanting to use mobile or pervasive displays have to deal with the lack of any haptic cue on the touchscreen and the lack of visual feedback. However, previous work has shown the interest for VI people to use wearables [Ye *et al.*, 2014] and particularly smartwatches: they are small, easily accessible and unobtrusive to wear, and can improve information access and social interactions. Our work in this area has the goal of overcoming the inherent limitations of digital content by providing a visually impaired user with a way to independently explore and/or reconstruct data. To this end, we explored how smartwatches can be used to access digital geospatial data [Bardot *et al.*, 2016, Bardot *et al.*, 2017].

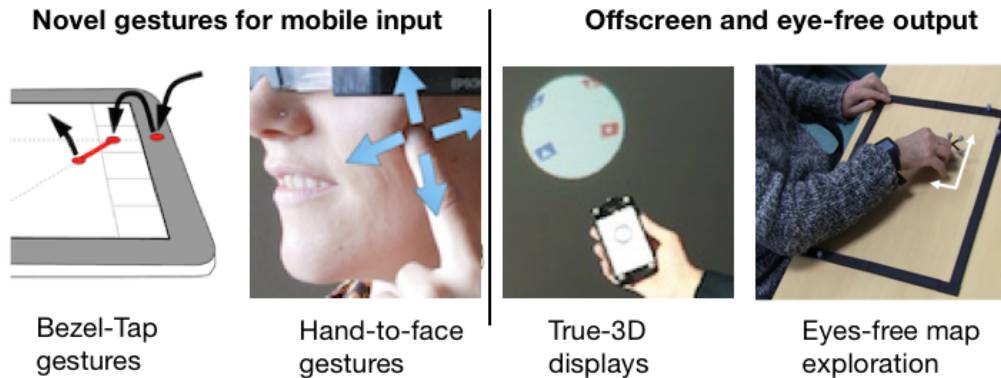


Figure 1.2: We proposed two types of contributions to mobile interaction: novel gestures for mobile input (left), and data exploration techniques beyond the display space, i.e. offscreen or eyes-free output (right).

### 1.2.2 Multi-display tools and interfaces

Multi-displays environments (MDEs) offer numerous advantages for organizing information across displays, for enhancing individual and group work, and for extending the interaction space. Our contributions to interacting with MDEs can be classified along two main axes (Figure 1.3): novel tools for fluid interaction in MDEs (1) and data exploration on MDEs using overview + detail interfaces (2).

#### Tools for fluid interaction in MDEs

The incidental emergence of MDEs has resulted in a device vacuum: no device has been specifically implemented for optimizing interactions in such spaces. To this end, we developed two systems to facilitate the overall interaction on MDEs. First, we introduced Gluey [Serrano *et al.*, 2015], a head-worn software interface that acts as a glue for bonding information across the various displays in a distributed environment. Second, we presented a novel touch-enabled device, TDome [Saidi *et al.*, 2017], designed to facilitate interactions and address a range of tasks in MDEs. TDome is the combination of a touchscreen, a dome-like Mouse [Perelman *et al.*, 2015] providing 6 DOF, and a camera that can sense the environment. TDome thus inherits properties of other existing mouse-like devices but includes many novel features to tackle the needs of common MDE tasks.

#### Data exploration on MDEs using overview + detail interfaces

We also worked on overview + detail interfaces, which are very common on multi-display environments: in such setups, the mobile display shows a detailed view of a section of the large display (the overview). While this technique is well known for 2D overviews with one detail view, its use in 3D environments or with multiple detail views had not yet been fully explored. To this end, first we focused on how to move the detail view (displayed on the smartphone) on a 3D environment [Bergé *et al.*, 2014]. In particular, we explored the use of mid-air gestures for controlling the translation of the Detail view, by moving the hand around the mobile device or by moving the phone itself. Second, we studied the use of different number of detailed views to interact with very large graphs in an overview + detail setting composed of a large screen and a mobile device [Saidi *et al.*, 2016]. Our

goal was to find the optimal number of detailed views to perform tasks involving multiple graph nodes

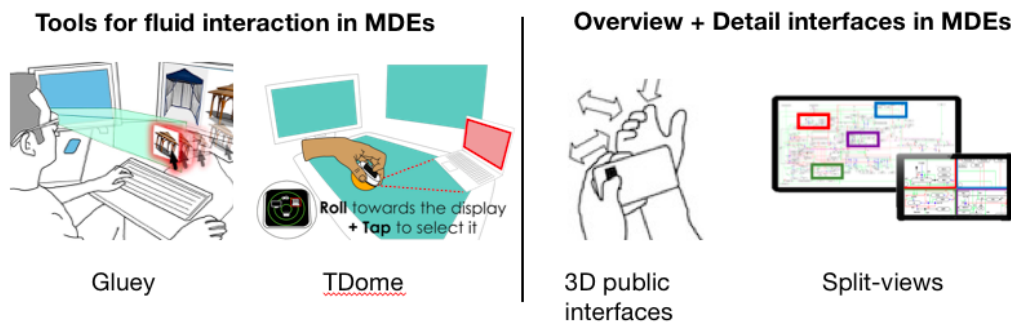


Figure 1.3: Our contributions to interacting with MDEs can be classified along two main axis: novel tools for fluid interaction in MDEs (left) and data exploration on MDEs using overview + detail interfaces (right).

### 1.2.3 FreeForm interfaces

To fulfill the adoption of pervasive displays, we need to facilitate a seamless integration of displays into existing environments, ranging from in-vehicle and wearable displays to public displays. Traditional rectangular displays are not well-suited for these applications, as they can not be easily integrated. Emerging technologies allow for the creation of nonrectangular displays with unlimited constraints in shapes. For instance, a single non-rectangular display can replace current instrument panels on car dashboards, in-between other on-board instruments; non-rectangular displays will also facilitate inserting displays on nonrectangular objects, furniture and/or urban architecture. In the context of mobile computing, non-rectangular displays can adopt shapes which will better fit wearable devices or replicate existing jewellery. With this eminent adoption comes the urgent challenge of rethinking the way we present content on non-rectangular displays (Figure 1.4).



Figure 1.4: Examples of 2D non-rectangular interfaces.

I started exploring this research challenge in 2016 in collaboration with Pourang Irani (University of Manitoba, Canada) and Anne Roudaut (University of Bristol, UK). We have published two preliminary explorations for displaying text and laying out content on non-rectangular interfaces [Serrano *et al.*, 2016, Serrano *et al.*, 2017]. This initial exploration suggests that the research question is timely, promising and interesting. The first work revealed

that text legibility is affected by shape factors such as text layout and alignment. The second work showed that content layout properties (grid type, type of symmetry, etc.) affect user perception differently according to the display shape. Overall, these studies reveal that different approaches apply to different content (such as text or icons layout).

### 1.3 Methodological challenges

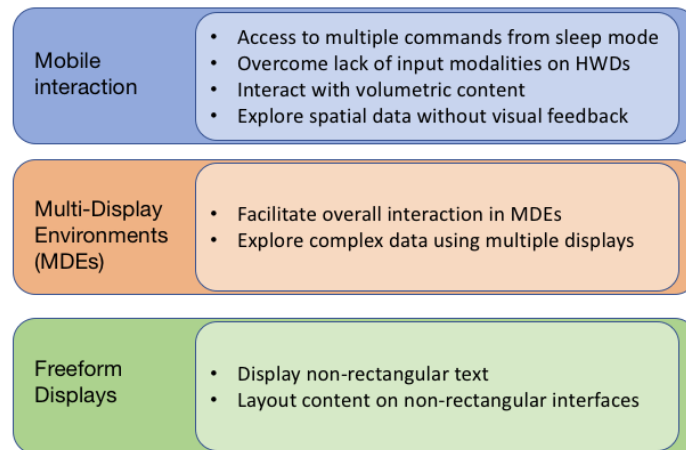


Figure 1.5: Research challenges we addressed for mobile, multi-display and freeform interfaces.

To address the previously presented research challenges, summarized in Figure 1.5, we had to face the methodological challenges of designing, prototyping and evaluating interaction. Traditional design and evaluation methods quite often failed at properly answering the previous research challenges. As a result, every new project came up with the additional challenge of rethinking our approach and research methods at multiple levels: design, prototyping and evaluation. For this reason, I deem of importance to include a chapter on research methods for ubiquitous interfaces (chapter 5). While the ultimate goal of our work was not to address these methodological challenges, we proposed novel methods or adapted existing ones.

Concerning design methods for HCI, we detail our experiences carrying in-the-wild observations [Lucero *et al.*, 2016], user-elicitation studies [Serrano *et al.*, 2014a] and design probes [Serrano *et al.*, 2017]. We also proposed a novel design process to combine existing interaction devices, DECO [Perelman *et al.*, 2016]. Prototyping advanced interactions can be difficult: we report our solutions for developing proof-of-concept prototypes for ubiquitous interaction, such as on-body input [Serrano *et al.*, 2014a] or mobile holographic (or true-3D) displays [Serrano *et al.*, 2014b]. Finally, we discuss evaluation methods with a focus on two key aspects: wizard-of-oz studies [Serrano and Nigay, 2010] and pairwise comparison experiments [Serrano *et al.*, 2017].

### 1.4 Manuscript overview

This manuscript is organized as follows. One chapter is dedicated to each one of the three types of ubiquitous interfaces explored in our research work: Mobile interfaces (Chapter 2), Multi-Display environments (Chapter 3) and Freeform interfaces (Chapter 4). These chapters start with a section detailing the research challenges addressed in our work, followed

by one section for each major contribution in the domain. Then the Chapter 5 presents our work on the research methods in three main sections: Design, Prototyping and Evaluation methods. Finally, Chapter 6 presents the conclusion and perspectives to our work. Figure 1.6 sums up the organization of this manuscript.

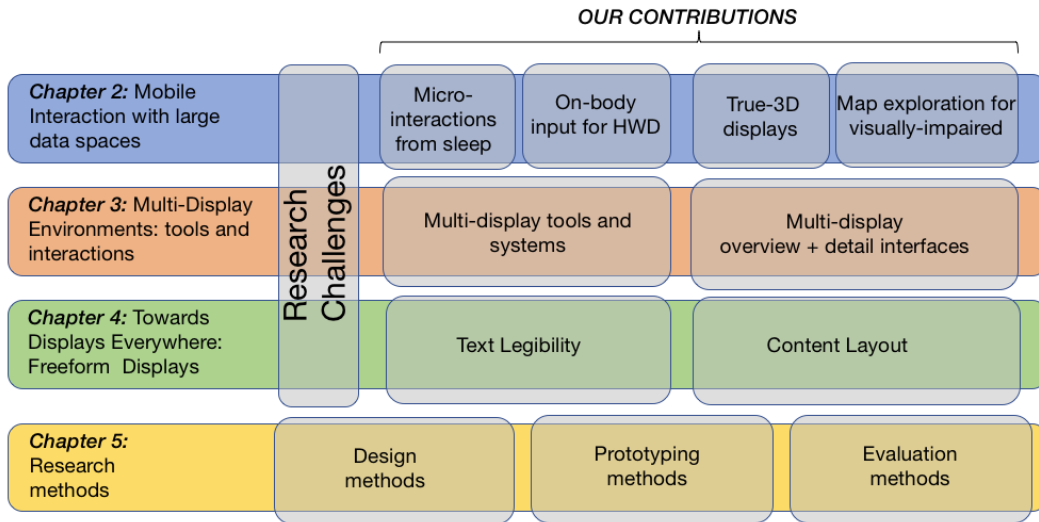


Figure 1.6: Outline of the manuscript organization.





## Chapter 2

# Mobile interactions with large data spaces

### Chapter content

---

2.1	Research challenges . . . . .	9
2.2	Command selection from sleep mode on tablets through Bezel-Tap gestures	11
2.3	Exploring spatial data on head-worn displays using hand-to-face input .	15
2.4	Mobile True-3D Displays . . . . .	18
2.5	Mobile map exploration for visually impaired people . . . . .	20
2.6	Chapter conclusion . . . . .	23

---

With the evolution in mobile computing technologies, smartphone, tablets and wearable devices are more powerful than ever. They can now easily process large volumes of data, such as multiple applications and commands, geospatial information or volumetric data. The bottleneck to fully exploit such data is the interaction capabilities of mobile devices, both in terms of input and output modalities. In this chapter I present our work on facilitating mobile interaction with large data spaces.

### 2.1 Research challenges

As said in the introduction, our work on mobile interaction addressed both the challenges of how to extend the limited input capabilities of mobile devices, particularly concerning gestural interaction, and how to access information beyond the display space, i.e. offscreen or non-visual output.

#### 2.1.1 Gestural interaction as mobile input

Our work on extending current input on mobile devices concentrated on two main research questions that we detail below: how to rapidly access multiple commands anytime, i.e. from the device sleep mode, and how to overcome the lack of input modalities to explore spatial data on head-worn displays.

##### Rapid access to numerous commands from sleep mode

A limitation of mobile devices is that they provide little support for quick commands. Micro-interactions, which rest on unconditional availability and fast access [Ashbrook, 2010], are especially desirable for frequent actions such as, for instance, controlling a media player, checking emails and SMS, calling friends and family, getting the local map or the weather forecast, and more basic actions like copying, pasting, and application switching. This is specially difficult due to the large set of mobile applications.

A 2012 survey showed that Android users have an average of 177 applications on their phones, using 73 of them occasionally and 32 very often [Shin *et al.*, 2012].

Since mobile devices constantly switch to sleep mode to save energy, interaction is hampered by the need to reactivate them whenever they have gone to sleep, typically by pressing a physical button and sliding a widget on the screen. This problem is exacerbated when mobile devices are used to control multimedia devices (TV, set-top box, etc.) and home equipment (home automation, domotics systems). In this scenario remote commands expand the large set of mobile applications. The challenge, then, is to allow always available and rapid access to a relatively large number of commands.

Gestural interaction provides an efficient means for activating commands rapidly. Marking menus [Kurtenbach and Buxton, 1991] are a well-known technique relying on 2D gestures. They have inspired many HCI studies, some of them dedicated to mobile devices [Kin *et al.*, 2011]. One major merit of these techniques is to make it easy to discover and learn the gestures: all commands are visible in novice mode and gestures are learned incidentally from repetitive use. These techniques are well suited for triggering a few dedicated commands at the application level, but can not be used as a global shortcut mechanism, i.e. without interfering with application shortcuts. Finally, gestural interaction does not work with the device in sleep mode, and unfortunately that is often the case due to the high power consumption of the touchscreen. In summary, there is a need for a novel type of gesture interaction on handheld devices that can be employed whether the device is alive or in sleep mode.

### Exploring spatial data on head-worn displays

Head-mounted devices are becoming widely available for daily use through lighter form factors and with transparent displays. We refer to these modern accessories as head-worn displays (HWDs). As consumers get affordable access to HWDs, ways in which they interact with content on such devices is being actively investigated. These devices are particularly useful for way-finding, and therefore it is important to facilitate map-related tasks, such as pan and zoom. However, interacting with such spatial information is particularly difficult given the limited input capabilities of HWDs.

Currently, HWDs provide onboard microphones and small capacitive sensors for user input. Voice recognition is useful for command-based tasks such as for search queries but has limitations in certain settings (i.e. noisy environments). The capacitive surface on the temple of HWDs presents a viable on-device method for input, but it has limited input space. Other less self-contained options, such as a wearable device or an auxiliary smartphone, can also allow for input. However, these require carrying them and may be occluded by the HWD content.

Mid-air and on-body gestural interaction can overcome the above limitations. However, mid-air input [Bailly *et al.*, 2012] suffers from the lack of tactile feedback. On-body interaction offers an input surface, the human skin, with the advantage of leveraging human proprioception as an extra feedback mechanism. This can overcome some of the limitations with mid-air interactions. A large body of work has considered appropriating the body as an interaction surface. Much of this prior work has considered coupling on-body input with on-body projection using wearable pico-projectors [Harrison *et al.*, 2010]. An open question then is which body area could better fit the specific needs of interacting with HWDs.

### 2.1.2 Offscreen and eyes-free mobile output

The second main challenge, i.e. exploring information beyond the display space, evolved around two different questions detailed below: how to interact with volumetric content around mobile devices (offscreen space); and how to facilitate the non-visual exploration of spatial data for visually-impaired people.

### Interacting with volumetric content

Benefiting from 3D on mobile devices is pertinent, beyond video games, for mobile scenarios such as 3D interior design or 3D map exploration. Mobile devices (smartphones and gaming consoles) incorporate at present auto-stereoscopic displays as a first step toward achieving true-3D content. True-3D [Kimura *et al.*, 2011], i.e. displaying volumetric 3D content in mid-air, can be accomplished through different means: stereoscopic displays with head-tracking [Nii *et al.*, 2012], optical illusions [Hilliges *et al.*, 2012], moving parts or augmented reality glasses.

Integrating true-3D on mobile devices, apart from facing hardware challenges, presents a number of unresolved human factors questions concerning its use. We possess limited knowledge on both the numerous constraints imposed on viewing and interacting with mobile true-3D interfaces and the usage scenarios suitable for such displays. These include knowing about the ideal angular position, size and distance of the volumetric projection, relative to the mobile device, the projection limits on visual search and direct interaction and how to coordinate the mobile and true-3D views. Answers to such questions will equip manufacturers and designers with tools to begin exploring a range of technologies for true-3D input on mobile devices.

### Mobile exploration of spatial data without visual feedback for visually-impaired people

Visually impaired (VI) people need regular access to geospatial information during education but also during everyday life. Previous work has already explored how VI people can access virtual maps. Besides static solutions based on using the keyboard or haptic devices (e.g. mouse or phantom), the main approach for exploring digital maps in mobile context is to use a touchscreen (on a smartphone or a tablet). For instance, [Su *et al.*, 2010] investigated the sonification of simple navigation maps on handheld touch screens. The evaluation showed that users of the proposed prototype recognized a few shapes in a matching to sample protocol, but can also develop a rough understanding of an indoor floor plan. However, an obvious drawback of handheld devices is that they only provide a limited surface for exploration, and hence require recurrent panning and zooming operations that are very difficult to perform by VI users. The challenge then is how to provide VI people with mobile techniques to explore spatial data that do not rely on the use of a touchscreen.

#### 2.1.3 Summary

In this chapter we present our solutions to address the aforementioned research questions (see Figure 2.1): to rapidly launch commands from sleep mode, we created a novel type of gestures called Bezel-Tap Gestures (section 2.2); to facilitate navigating spatial data on head-worn displays, we proposed using on-body gestures on the face, i.e. hand-to-face input (section 2.3); to facilitate the visualization of volumetric data, we explored mobile true-3D displays (section 2.4); finally, to facilitate non-visual exploration of spatial data for visually-impaired people, we studied the smartwatch output capabilities (section 2.5).

## 2.2 Command selection from sleep mode on tablets through Bezel-Tap gestures

During my post-doctoral fellowship at Telecom Paristech, I worked with Eric Lecolinet and Yves Guiard on solving the challenge of enabling micro-interactions from sleep mode on mobile devices. We proposed Bezel-Tap Gestures, a novel interaction technique that not only serves to open an application and launch a favorite command rapidly, but can also wake the device if asleep. This work led to one publication at CHI 2013 [Serrano *et al.*, 2013].

Challenges	Gestural input		Offscreen output	
Research question	Launch commands from sleep mode	Navigate spatial data	Interact with mobile true-3D data	Perform non-visual exploration
Data	Multiple commands	Spatial data	Volumetric data	Spatial data
Devices	Smartphone/Tablets	Wearables: HMD	Smartphone/Tablets	Wearables: watch
Chapter section:	2.2	2.3	2.4	2.5
Contributions	Bezel-Tap gestures	Hand-to-face input	Evaluation of mobile true-3D	Watch-based map exploration

Figure 2.1: Research challenges and our contributions to the domain of mobile interaction with large data spaces.

### 2.2.1 Bezel-Tap Gestures

In its basic form a Bezel-Tap gesture involves two successive events recorded in different input channels, a tap on the bezel immediately followed by a tap (or a slide) on the screen (Figure 2.2). These gestures were primarily designed for tablets, which have quite large bezels, and do not interfere with common touch-based interaction techniques. Most importantly, the technique makes it possible to both wake up the device and activate a command without the risk of draining the battery. This property is useful even if the device is password protected as many commands are not security threatening, this being especially true when the device serves to control home equipment.

#### Gesture detection

A Bezel-Tap gesture involves a first tap on the bezel, detected by accelerometers, and a subsequent tap on the touchscreen that must occur within a short time interval (Figure 2.2). So we have two temporal constraints: 1) the time delay between the two events must be greater than a few milliseconds (50ms with our hardware) so that a Bezel-Tap gesture cannot start with a screen tap (this could otherwise happen in the case of a double tap on the screen if the user taps hard); and 2) the delay must not exceed 600ms to avoid taking unrelated events into account (i.e., a screen contact occurring long after a tap).

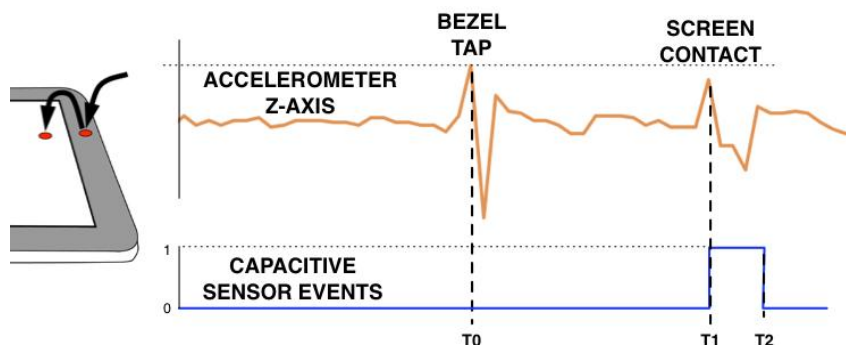


Figure 2.2: A Bezel-Tap gesture involves a first tap on the bezel, detected by accelerometers, and a subsequent tap on the touchscreen.

Incidentally, it is worth noticing that Bezel-Tap gestures allows identifying a location on the device (that serves to select a command) but only rely on the touchscreen to do so. While accelerometers could theoretically serve to detect tap locations, hardware currently available on mobile devices would make this hardly feasible. Using two input modalities (the accelerometer and the touchscreen) solves this problem: the touchscreen provides information the accelerometer is unable to provide.

### **Power considerations**

Bezel-Tap Gestures make it possible to trigger commands quickly, even with the device in sleep mode. This reactivation feature requires the accelerometer to remain permanently powered. The small power consumption of accelerometers makes this feasible without considerably reducing the battery life. Our estimations were that our prototype would only reduce the battery life of about three quarters of an hour every 45h of battery life.

We also investigated the effect of carrying the device in mobile context to see how many taps would be detected by the accelerometer just because of the motion of the device. A high number of detections would drain the battery because, using our technique, taps reactivate the capacitive sensor of the screen. We hence conducted an experiment carrying a tablet inside a backpack during subway, bus and bike journeys. We collected 8 hours of tablet's internal accelerometer log in mobile context. On average 6 taps per hour were detected: 9 taps/hour on bus, 5 taps/hour on subway and 4 taps/hour on bike. Therefore the added power consumption is negligible.

### **Inadvertent activations**

The fast succession of two events from different input channels (a tap followed within a short time interval by a screen touch) is a low probability event that can serve as a unique signature. For instance, no Bezel-Tap gesture is recognized if the user double taps the screen or double taps the bezel. In order to evaluate the probability of inadvertent activations of Bezel-Tap gestures, we completed a field study where we gave 12 participants a tablet for 24 hours, asking them to use a web browser implementing Bezel-Tap Gestures for at least one hour in total. The results show that a few taps were detected by the accelerometer (on average 7.7 taps per hour of use) but none of them were followed by a screen contact in less than 600ms, so that no false positive was detected. This field study suggests that our technique is robust to common usage.

### **Target selection performance**

Selection performance is fast and accurate up to a fairly large number of commands, even in the absence of any screen feedback. We carried an experiment to evaluate the performance of Bezel-Tap gestures in terms of speed and precision depending on the size of the command set (4, 5, 6 or 7 items per bezel region, see Figure 2.3). Additionally, we also wanted to compare performance for the different regions of the bezel (top, bottom, left, right). The results of this experiment confirmed the usability of the Bezel-Tap technique in all four regions of the device in sleep mode. Selection times, on the order of 1.5 sec (with the reaction time included), were compatible with the micro-interaction concept [Ashbrook, 2010]. Performance accuracy was as good for  $N = 5$  (96.5%) than for  $N = 4$  (96.9%), but decreased for  $N=6$ . One practical suggestion that arises from these results is that a set size of five items is optimal for the technique and a good solution to select shortcut commands.

#### **2.2.2 Bezel-Tap Menus**

Bezel-Tap Gestures can be extended to allow selecting a large number of items. We proposed two new techniques that rest on a hierarchical organization (Figure 2.4). Bezel-Tap Slide (BTSlide) involves a tap on the bezel followed by a slide on the screen. The starting



Figure 2.3: We evaluated the target selection performance depending on the size of the command set: 4, 5, 6 or 7 items.

and ending points of the slide selects a group of items and an item in this group respectively. Bezel-Tap3 (BT3) involves three taps: one on the bezel and two on the screen. The second tap selects a group of items and the third tap an item in this group.

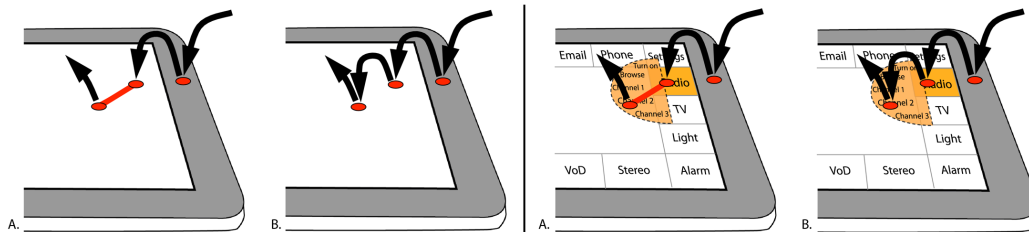


Figure 2.4: Bezel-Tap Menus in expert (left) or novice mode (right). Bezel-Tap Slide (A) involves a tap on the bezel followed by a slide on the screen, while Bezel-Tap3 (B) involves three taps.

These Bezel-Tap menus are hierarchical. Their first level consists of five rectangular items for all four bezel regions (Figure 2.5). According to our experiment results we decided to expand even items and reduce corner items by 15% in order to increase even-numbered items success rate. The second level of the menu rests on 180° radial menus. The second tap (resp. the starting point of a slide) selects one of these radial menu, for instance the "Radio" menu in Figure 2.5. The third tap (resp. the ending point of a slide) selects an item in this menu. Menus only appear if the user waits more than 600ms between the second and the third tap. A complete novice user just needs to tap on the bezel then anywhere on the screen and wait for 600ms. An expert user will bypass these stages by performing all taps (or slides) without any visual feedback. To make interaction easier, we did not assign menus to corner items (which can serve as one-step shortcuts for very frequent commands). This design allows for a total of 12 radial menus (3 per region), and each radial menu offers five items. A Bezel-Tap menu can thus comprise a total of 64 items: 4 corner shortcuts and 60 menu items.

### Bezel-Tap Menu performance

We carried a study to evaluate the usability of Bezel-Tap Menus for command selection in expert mode. Participants were asked to perform the full gestural sequence illustrated in Figure 2.4, using either BT3, BTSlide or an extension of Bezel Gestures [Bragdon *et al.*, 2011] that allows selecting more commands than the original technique. The results show that after little training all three techniques allow accurate selection in a two-level hierarchical menu. Bezel-Tap techniques (BT3 and BTSlide) were more accurate than Bezel Gestures (error rates were 5.2%, 4.5% and 8.7%, respectively), but slower. Anyway, the speed was pretty good for Bezel-Tap techniques (1.6s on average), this making them appropriate for

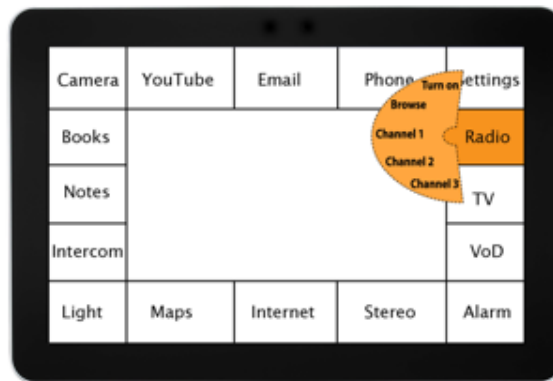


Figure 2.5: Illustration of the content of a two-level Bezel-Tap Menu.

micro-interactions. This result is even more interesting considering that Bezel-Tap gestures can not only select a command but also reactivate the device in this short amount of time.

### 2.2.3 Summary

Bezel-Tap Gestures is a technique that allows selecting commands quickly on mobile devices. Bezel-Tap Gestures do not interfere with common interaction techniques when the device is in active mode. They also work in sleep mode, both waking up the device and triggering a command. Our work explored the use of a simple but effective method to extend input interaction on mobile devices, i.e. tap gestures. Since the publication of our work on Bezel-Tap gestures in 2013, the use of tap gestures has been applied to back-of-device interaction on smartphones [Granell and Leiva, 2017] and for rapid command selection on smartwatches [Oakley *et al.*, 2015].

## 2.3 Exploring spatial data on head-worn displays using hand-to-face input

During my post-doctoral fellowship at the University of Manitoba, I worked with Pourang Irani and Barrett Ens on solving the challenge of facilitating input interaction, spatial navigation in particular, on head-worn displays (HWDs). We proposed hand-to-face input as a novel, alternative method for interacting with HWDs. This work led to one publication at CHI 2014 [Serrano *et al.*, 2014a].

### 2.3.1 Hand-to-face gestures

We define hand-to-face input as any gesture that involves touching, rubbing, scratching or caressing the face (Figure 2.6). This approach is especially well-suited for interaction with HWDs for many compelling reasons: (i) the face is often touched making it a promising area for subtle interactions; (ii) it offers a relatively large surface area for interaction, but not normally clothed as are other areas; (iii) it facilitates eyesfree, single-handed input, which can be invaluable in mobile settings (e.g. riding a bike, holding on in a bus); and (iv) is in close proximity to the HWD, making it likely to accommodate device-borne sensors and creating a natural extension of the device temple.





Figure 2.6: Hand-to-face input for navigation. a) Panning, b) Pinch zooming, c) Cyclo zooming, d) Rotation zooming. Our studies show that Cyclo was not socially acceptable while Rotation was not efficient.

### Suitable areas for interaction

To explore the breadth of potential hand-to-face gestures and their mapping to interactive tasks, we elicited user input through a guessability study [Wobbrock *et al.*, 2009]. For a set of common mobile tasks [Ruiz *et al.*, 2011], divided into action or navigation tasks, we asked participants to suggest suitable gestures on the face (above the neck) and on the HWD.

On the face, participants produced gestures for a total of 11 different areas, such as the cheek, forehead, ear, chin or jaw. The results reveal a distribution (Figure 2.7) with gestures concentrated on the cheek and then on the forehead. Other areas saw an equal distribution of gestures: jaw, ear, temple, and chin. Areas such as the eyes, nose, lips, neck and hair were least used. On the HWD, participants used 5 different interaction areas: temple, hinge, frame, bridge and glass. Most of the gestures were situated on the temple.

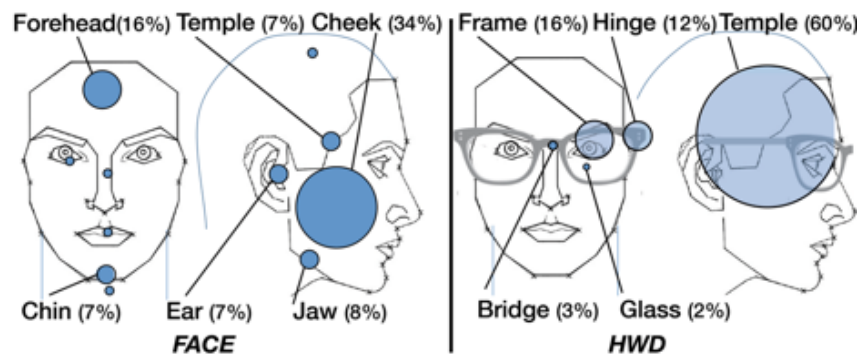


Figure 2.7: Main areas identified by participants as suitable for input on the face (left) and on the HWD (right). The circles' size is proportional to the percentage of gestures (in brackets).

Overall users preferred interaction on the face for the navigation tasks, while opinions were mixed for the actions tasks. Users particularly preferred using the face for panning and zooming. Users indicated that "the face provides a larger area", which is perceived as a benefit for panning and zooming. This is particularly true when using the cheek, since it is "the best part to interact with the face" and it is "like a touchpad".

### 2.3.2 Spatial navigation with hand-to-face input

We carried two studies to evaluate the performance, physical effort and user preference of hand-to-face gestures for panning (study 1) and zooming (study 2).

For the first study we implemented three panning interactions: displacement-based (Pan-D), flick (Pan-F) and rate-based (Pan-R). With Pan-D, finger movement is directly mapped to the movement of the map. In Pan-F, the user flicks to pan, mimicking the iOS flick behaviour. In Pan-R, the distance panned from the initial touch position is mapped to the finger velocity movement. Participants used four different interactive surfaces, two on the face (Cheek and Forehead) and two on the HWD (Oversized and Regular temple). We include two different temple sizes in order to study the impact of its size on navigation.

This first study revealed that the best facial area for input is the Cheek. The Forehead and the Regular HWD Temple not only showed worse performance, but also result in higher fatigue. Overall there was no difference between the Cheek and the Oversized temple, but both were favored over the Regular temple. The Oversized temple, however, is far larger than most HWDs, suggesting that the Cheek is a preferred interaction region.

For the second study, we selected three different zooming techniques, based on a combination of prior known methods and from the guessability study: Linear, Rotation and Cyclo. Linear zooming, by pinching with two fingers, is the classical touchscreen technique. Circular zooming with two fingers (using the angle of rotation) is based on the metaphor of adjusting an optical lens. Cyclo is a one finger zooming technique proposed as a way to avoid clutching when zooming [Malacria *et al.*, 2010]. It consists of doing a circular movement with one finger. The orientation of the rotation is mapped to the zoom direction (in or out). From the previous study we dismissed the Forehead area due to its low overall results.

The results of the second study extend further our exploration of Study 1, providing insight into hand-to-face interaction for document navigation. The main finding is that the Cheek is more efficient than both the Oversized and Regular temples for zooming. While the Oversized temple was efficient in Study 1 for one finger panning, it becomes inefficient with a two-finger gesture. Both the classical Pinch and the single-finger Cyclo are equally efficient in our study.

### 2.3.3 Social acceptability of hand-to-face input

While we demonstrated that hand-to-face interaction techniques improve navigation performance, we know little on how comfortable users would feel in different social contexts. We therefore carried a controlled exploration of the social acceptance [Rico and Brewster, 2010] of our hand-to-face gestures. Participants watched a video of an actor performing panning and zooming gestures in front of a wall and then performed themselves the same gestures 3 times. Participants were asked to rank on a 5-point Likert scale the social acceptability of each hand-to-face gesture.

We found no difference in social acceptability between Face and HWD, but with a trend showing better acceptance for interaction on the HWD. The acceptance rate for both face and HWD gestures in any social context is above 50%. Results were rather homogeneous on the HWD, with a constant 10%-12% of disagreement for all social contexts except in front of strangers, where this value is 18%. We found more differences on the Face, with no disagreement when at Home or Alone, but with 31% disagreement in Public places and 25% in front of strangers. Comments from participants also show that most of them don't mind using the face: "I don't think it would disturb me to do the gesture either on the skin or on the temple.". One female participant indicated the problem of dirty hands on the face: "the face can be affected when perspiring".

Concerning the acceptability of the different hand-to-face gestures, we found a significant difference between the panning techniques, Pan-D being better perceived than Flick. We also found a difference between the zooming techniques, Linear being better perceived than Cyclo. Participants commented that Cyclo might be perceived as insulting, as it could signal that "you are crazy" in many cultures. This gesture seemed also more visible: "I feel all the gestures are quite subtle except Cyclo which might attract attention".

### 2.3.4 Summary

We explored hand-to-face gestures for HWDs without emphasizing the technology that would ultimately support this style of input. Interestingly, in 2017 researchers at Keio University (Japan) proposed a technology that senses touch gestures on the cheek by detecting skin deformation [Yamashita *et al.*, 2017]. Our work on face input was the first to propose the use of such an area for interaction, and has been succeeded by other interesting explorations of head-based input for mobile computing involving the ear [Kikuchi *et al.*, 2017] or the nose [Lee *et al.*, 2017].

One of the remaining challengers of on-body interaction is to consider the Midas-touch problem [Hinckley, 2003], i.e. how to differentiate casual and explicit on-body gestures. Two obvious but cumbersome solutions include touching the HWD to initiate the face detection or using a voice command. Another solution is to use gestures that are very different from casual ones.

## 2.4 Mobile True-3D Displays

During my post-doctoral fellowship at the University of Manitoba, I investigated with Pourang Irani (University of Manitoba, Canada) and Sriram Subramanian (University of Bristol, UK) how to explore volumetric content on mobile true-3D displays. As said earlier, true-3D refers to any 3D digital display capable of producing mid-air, full-depth-cue (or volumetric), multi-angle and/or multi-user images without the need for user instrumentation. This work led to one publication at MobileHCI 2014 [Serrano *et al.*, 2014b].

### 2.4.1 Display properties

We first studied the various factors that can potentially influence the effective deployment of true-3D on mobile devices (Figure 2.8). We focus on mobile-mounted 3D projection, which means that the true-3D projection moves with the mobile device as if both were attached.

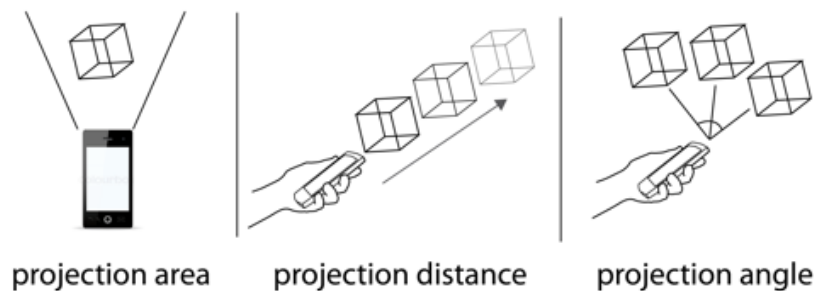


Figure 2.8: Mobile true-3D properties: projection area, distance and angle.

*Projection area:* While pico-projectors need a projection surface, true-3D projectors may display an image in mid-air around the mobile device. Prior work has generally kept the mobile projection pointing downward, straight or steerable (using a motor to direct the projector) [Cauchard *et al.*, 2012]. These solutions provide significant flexibility in finding a suitable projection surface. A true-3D mobile display needs not be constrained by the position of the projection throw. Therefore after considering the potential projection areas around the smartphone, we decided to focus on the top area of the phone (Figure 2.8-left). This area always remains visible when the user rotates the phone to inspect the 360° true-3D image.

*Projection distance to the 3D object:* The distance between the mid-air 3D projection area and the smartphone (Figure 2.8-center) may have an impact on users' visual perception and direct input. If the projection is far from the device, it may affect the continuity of the visual search but even further limit direct-touch interaction with the true-3D and require indirect forms of interaction.

*Projection angle:* We define the projection angle as the angle between the phone's y-axis and the 3D object (Figure 2.8-right). Traditional depictions of mobile true-3D envision the 3D content at a  $90^\circ$  angle relative to the phone's plane or displayed directly over the touchscreen. These depictions assume the best projection extends the mobile display into a 3D volume. However, this vision is limited as it considers the true-3D simply as an extension of the touchscreen instead of viewing it as a secondary display that can extend the mobile phone's capabilities.

Beyond these three main properties, we also considered the *volume* of the projection and the user's *point of view*. The display *volume* may affect visual search as well as direct-touch interaction. The visual exploration (hence the user's *point of view*) of a mobile true-3D display relies on wrist rotation dexterity to avoid complex interactions for rotating it. Given the restrictions in wrist rotation angles, we expect limited accessibility to occluded areas on the true-3D projection.

### 2.4.2 Spatial configuration

We carried an experiment to identify the best spatial configuration for the projection of a mobile true-3D display to ensure effective visual search. We explore the case of natural user interaction, i.e. wrist rotation for search without interface support, but do not explore any type of user input. We focus on the properties described in the previous section (Figure 2.9): projection's angle to the phone plane, distance to the phone, volume and pattern position on the true-3D (point-of-view). Participants were required to identify the location of a graphical pattern on a 3D opaque sphere on the true-3D display. The sphere was separated into eight parts, each one containing a unique pattern (Figure 2.9-right).

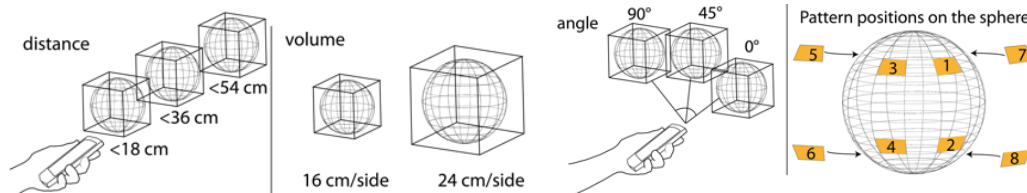


Figure 2.9: Mobile true-3D properties evaluated in our first experiment.

The results of this first study demonstrates that visual search is easier when the 3D object is positioned at an angle of  $0^\circ$  or  $45^\circ$  and at a distance of less than 36 cm. We also found that the region in the back and opposite to the hand holding the device is the weakest for object search. This is primarily due to the wrist dexterity as observed during the experiment and from participant feedback. Wrist dexterity also affects objects located further away, i.e. these become hard to inspect under all angles. Thus, our results recommend shorter distances if the device is to solely rely on wrist rotation for viewing the display.

### 2.4.3 Direct touch interaction

We carried a second experiment to investigate the effect of volume size on visual search and direct input. Participants were required to identify the location of a graphical pattern on the true-3D display and to select it with direct mid-air touch. We added one larger

volume size than in experiment 1 to further investigate this factor. The results of this second experiment informed us on the suitable values for the projection volume for direct interaction. Overall, completion time increased with the volume of the display. Projections smaller than 24cm/side improved efficiency. The slight cost in accuracy at smaller volumes suggests that target sizes need to be considered carefully for such displays.

#### 2.4.4 Summary

In HCI we find numerous examples of novel technologies whose adoption, from discovery to commercial use, take decades. Buxton refers to this process as the Long Nose of Innovation [Buxton, 2008]. Our work was motivated by our will to reduce the long nose for mobile true-3D. Extensive research is taking place to engineer mobile true-3D [Nii *et al.*, 2012] and following our work, researchers at Queen's University (Canada) presented a 3D flexible smartphone rendering "holographic" images providing motion parallax and stereoscopy to multiple users without glasses [Gotsch *et al.*, 2016]. Our contribution was to identify application guidelines, limitations and scenarios to help future adoption of this technology.

### 2.5 Mobile map exploration for visually impaired people

In this section we present the work carried by Sandra Bardot, a PhD student at the University of Toulouse that I co-supervised with Christophe Jouffrais (CNRS - IRIT Lab, Toulouse). We designed a mobile technique, based on hand tracking and a smartwatch, in order to provide pervasive access to virtual maps for visually impaired (VI) people. This work led to one publication at the conference MobileHCI 2016 [Bardot *et al.*, 2016].

#### 2.5.1 Interaction principles

Our work focused on the spatial exploration of geographic maps with associated data (e.g. demographic or weather maps). We focused on the complementary objectives of improving the spatial component of exploration, while providing the user with large surfaces and collocated feedback. To this end, we first established the principles of such an interaction in terms of which data should be explored, the tracking to use and the interaction modalities to employ.

##### Map content for VI people

We first analyzed the layout and content of regular raised-line maps, which are the most commonly used tools by VI people for exploring maps. Tactile raised-line maps have two main advantages: information is tactile, and spatial exploration is direct. They are made according to guidelines (simplification of contours, reduction of the number of elements, legends, etc.). The most important elements of these maps are the contour of areas, the points of Interest (POI) and a Braille legend describing each area and POI (Figure 2.10). Finally these maps present data associated to each area or POI, for instance the population of a region. This information is usually written outside the map with Braille.

In addition, tactile exploration of novel spaces relies on behavioural strategies such as scanning the image from left to right and top to bottom. We aimed to preserve these strategies during spatial exploration of virtual maps, therefore using hand tracking to locate the hands and render the information that is under the hand.

##### Hand tracking for ubiquitous map exploration

Spatial exploration by using hand tracking instead of touch input offers several advantages for VI people. First, VI people tend to put multiple fingers and hand palm down on the surface, which, in absence of visual feedback, generates unexpected events. Instead, hand

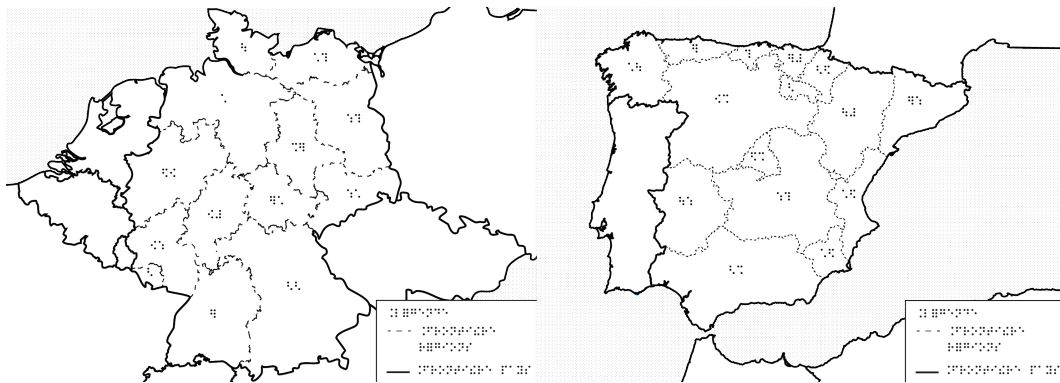


Figure 2.10: Raised-line maps of Germany (left) and Spain (right) showing regions, points of interest and legends.

tracking can simply associate one point with each hand. Second, hand tracking allows performing mid-air gestures, for instance to change the information level of the map when raising the hand. Although mid-air exploration may seem difficult for VI people, we explored its use to perform coarse-grained spatial exploration. Finally, mobile and low-cost hand tracking solutions have recently been proposed, which could leverage making map exploration possible in different contexts and on different surfaces, such as a regular desk, without the need for an underlying touch sensitive surface. Coupling hand tracking with a wearable device for input and feedback (e.g. a smartwatch) makes it possible for VI people to explore virtual maps in many places such as at school, in a mall or at home.

### Using a smartwatch for localized feedback

Previous work has shown the interest for VI people to use wearables [Ye *et al.*, 2014] and particularly smartwatches: they are small, easily accessible and unobtrusive to wear, and can improve information access and social interactions. Current smartwatches have the advantage of including speakers and vibratory feedback. We decided to use a smartwatch to provide hands-free map exploration. We used the smartwatch both as input and output. As input, the device's touchscreen is used to filter or brush data by performing simple horizontal or vertical swipe gestures. As output, the device is used to render localized Text to Speech (TTS), for instance the name of regions. The vibratory feedback is also used to render information, such as the border between regions.

### 2.5.2 Non-visual map exploration techniques

We identified different mappings between the smartwatch input/output modalities and the map exploration task, resulting in three different exploration techniques (Figure 2.11).



Figure 2.11: We investigated three non-visual map exploration techniques: plain (left), filter (center) and grid-filter explorations (right).

### Plain exploration

Plain exploration is the equivalent to the exploration performed on a raised-line map: each element on the map is rendered. The smartwatch is only used as an output for this technique. We combined auditory and vibratory feedback. TTS reads out information underneath the hand, such as the name of the region and its population. A 100 ms vibration notifies the transition from one region to another one. A continuous vibration means that the hand is outside the map.

### Filter exploration

Filtering data before exploration allows reducing the amount of information to render through TTS, and thus reduces the user's cognitive load. The filtering allows selecting a sub-range of values, for instance regions with more than a hundred thousand residents. To perform the filtering, users make swipe gestures on the smartwatch (Figure 2.11-center). After selection, only the data that corresponds to the selected filter is read out, similar to the Plain exploration technique.

### Grid-Filter exploration

With the previous techniques it can be difficult to find certain regions in a map especially if they are small. To get a full glance of a map without missing any region, one solution consists in using a 3x3 grid, i.e. reading out the information concerning all the regions contained in each cell of the grid [Zhao *et al.*, 2008].

However, when gathered within a grid, the spatial relationships between regions are masked. To overcome this limitation, we combined the Filter exploration mode with a Grid-based exploration mode. The user can use one or the other interaction level according to hand height above the map. When the hand is lying onto the table, the user explores the map in Filter mode (i.e. only the data that corresponds to the selected filter is read out). When the hand is moving over the table, the user explores the map in grid mode (i.e. reading out the information concerning all the regions contained in each cell of the grid).

## 2.5.3 Experimental results

We carried a study to compare the effectiveness of our virtual map exploration techniques against exploring a raised line printed version (Figure 2.12). The task was to explore a map, and answer a question as fast as possible. Twelve visually impaired participants volunteered for this experiment.

Overall, Grid-Filter was faster than the other techniques: on average, answering a question with the Grid-Filter technique took 40 s, with Filter 83 s, with Raised-line 127 s and with Plain 172 s. Response times were significantly longer with Raised-Line and Plain. This is due to the fact that users had to thoroughly explore the map in order to find the targeted region and the associated data to answer a question. On contrary, Filter and Grid-Filter renderings quickly provide access to the answer. Interestingly, the Grid-Filter technique was the more efficient but not the preferred one. Most participants ranked the Filter technique first on user preference criteria. These results can be explained by the use of mid-air gestures in the Grid-Filter technique: many participants reported that it is tiring, and that it is difficult to build a mental representation of the map when their hand is moving above the map.

## 2.5.4 Summary

Our work focused on the usability of virtual maps, including different exploration techniques (Plain, Filter and Grid-Filter), as opposed to the regular raised-line maps. Overall, the results suggest that VI people are able to explore geo-spatial data in virtual maps.



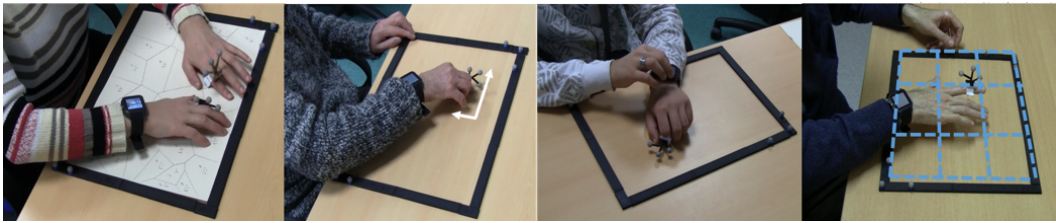


Figure 2.12: From left to right: exploration of a raised-line map with an infrared tracker on the finger; plain exploration of a virtual map; left/right swipes on the smartwatch to filter the data (filter exploration); and raising hands to gather the data within a virtual grid layout (grid-filter exploration).

More precisely, they show that when filtering functions are added to virtual maps, which is impossible to achieve with tactile maps, they provide the user with an efficient mean to retrieve specific answers about the elements of the map. These results could be of interest to design eyes-free map exploration for sighted users, as a way to overcome the limited display space of smartwatches. To this end, the performance of our non-visual techniques would need to be evaluated against current visual techniques for map exploration on watches.

## 2.6 Chapter conclusion

In this chapter we presented our contributions to mobile interaction with large data spaces, which evolved around gestural input and offscreen data exploration. We proposed two novel types of input gestures: Bezel-Tap gestures, for selecting discrete commands from sleep mode, and hand-to-face input, a type of on-body gestures which proved to be valuable for continuous exploration of spatial data on head-worn displays. Concerning offscreen data exploration, we investigated how to access volumetric content around the smartphone's displays using true-3D, and how to explore spatial data without visual feedback using the smartwatch output capabilities.





## Chapter 3

# Multi-Display tools and interfaces

### Chapter content

---

3.1	Research challenges . . . . .	25
3.2	Gluey: fluid interaction in MDEs using a head-worn display . . . . .	28
3.3	TDome: fluid interaction in MDEs using a dedicated multi-DoF device . . . . .	31
3.4	Interaction with Overview+Detail Interfaces on 3D Public Displays . . . . .	34
3.5	Splitting Detailed Views in Overview+Detail Interfaces . . . . .	36
3.6	Chapter conclusion . . . . .	39

---

Multi-displays environments (MDEs) have shown significant value for interacting with heterogeneous data sources and in multiple contexts such as collaborative analysis [Bortolaso *et al.*, 2013], crisis management [Chokshi *et al.*, 2014] and scientific data visualisation [Wang Baldonado *et al.*, 2000]. MDEs allow organizing information across displays, enhancing individual and group work, providing support to peripheral information and extending the interaction space. However, interaction in such environments is challenging due to the heterogeneity of existing solutions, and the benefits of extending the display space have been insufficiently studied for complex data (such as 3D content or very large graphs).

### 3.1 Research challenges

Our contributions to interacting with MDEs address two main research challenges: novel tools for fluid interaction in MDEs (1) and complex data exploration on MDEs (2).

#### 3.1.1 Fluid interaction in MDEs

The emergence of MDEs has resulted in a device vacuum: to our knowledge no device has been specifically implemented for optimizing interactions in such spaces. Researchers have mainly proposed adapting existing devices such as the mouse for multi-monitor pointing [Benko and Feiner, 2005], or smartphones for cross-display data-transfer [Boring *et al.*, 2010] or distant pointing [Nancel *et al.*, 2013].

However such adaptations can result in undesirable side effects: mice do not support well input on large displays and smartphones held in mid-air can be tiring and cumbersome for long interactions. Recent research has demonstrated the use of wearable devices to perform cross-device interactions [Houben and Marquardt, 2015]. However, current wearables lack proper input mechanisms and mainly serve private purposes. If MDEs are to become the office of the future (see Figure 3.1), as envisioned by [Raskar *et al.*, 1998],

can we design a device specifically tuned for such an environment? Adopting a unique device would avoid the homing effect when switching from one device to another, enhance privacy in such environments through personal data control and visualization, lead to a coherent set of interactions with the varied MDE applications, and ultimately contribute to a more fluid interaction [Bederson, 2004]. The design of such an unifying device could be based on two complementary approaches: using an already existing device, to facilitate the spread of the solution, or creating a new tailored device to replace existing ones.



Figure 3.1: Office of the future as envisioned by [Raskar *et al.*, 1998].

### 3.1.2 Exploring complex data on multiple displays

To explore large information spaces such as web pages or maps, research on visualization proposes three solutions to separate focused and contextual views: Overview+Detail (spatial separation), Zooming (temporal separation) and Focus+Context (seamless focus in context) [Cockburn *et al.*, 2009]. Overview+detail configurations fit particularly well MDEs, since overview and detail (O+D) views can be displayed on different screens. In our work we addressed two under-explored research questions around the use of multi-display O+D to explore complex data (3D content and very large graphs): interacting with 3D data on public displays, and exploring multiple regions of very large graphs simultaneously. We present and motivate the importance of these two research questions in the following paragraphs.

#### Interaction with 3D data on public displays

Public displays allow pedestrians to interact with 2D content such as city maps or tourism information. In this context, existing systems have used smartphones as personal views (Detail) of the public display (Overview), leveraging multi-user access to one display [Cockburn *et al.*, 2009]. Visualisation of 3D content on public displays is emerging to visualize scientific 3D data [Song *et al.*, 2011]; to explore culture heritage 3D scanned objects; to play public 3D games or to navigate a city 3D map.

Most of these examples already include the use of a personal device to interact with the 3D content (to orient and position a slice plane or to navigate in the 3D environment) but few consider how to apply the Overview and Detail (O+D) paradigm using the smartphone. Using the O+D on mobile devices will provide the user with the ability to pri-

vately visualize details of the 3D environment while still taking advantage of the mobile device input to interact with the 3D content. However, the public context imposes certain constraints in terms of user's profiles (mainly beginners) and appropriate interaction techniques (which need to be easy to understand and perform). One of the main challenges then is to provide an easy to perform technique to control the position of the Detail view in a 3D Overview.

### Exploring multiple regions of very large data spaces simultaneously

As said earlier, MDEs are well suited for scientific data visualization. We identified the following challenge during a collaboration with biologists carrying research on cancer. These biologists archive knowledge in graphs called molecular interaction maps (MIM [Kohn, 1999], see Figure 3.2), which contain several types of nodes (molecules, protein, etc.) and connections. As research on cancer progresses, results are added to existing MIM maps, which grow extremely large (the Alzheimer MIM map contains 1347 nodes [Kohn, 1999]) making them difficult to read and edit using the traditional panning+zooming interactions. Moreover, another consequence of this growth is that connected nodes can be located far apart from each other, thus, requiring even larger surfaces to visualize the data.



Figure 3.2: Illustration of a MIM graph containing a total of 880 nodes and 732 connections [Kaizu *et al.*, 2010].

Despite the advantages offered by O+D interfaces when working on large datasets (like graphs), these interfaces reach their limits when it comes to work on multiple regions of the overview simultaneously. An example from the previous context would be connecting distant nodes of very large graphs for example or to create a link between 2, 3 or 4 nodes. These types of multi-node links are usual in large graphs such as MIMs. Moving the detailed view repeatedly from one region to another is tedious and interaction complexity increases with the number of regions to work on.

To address this situation, several techniques have been designed in single or multi-display configurations to support the use of more than one detailed view simultaneously [Elmqvist *et al.*, 2008] [Javed *et al.*, 2012]. Earlier work had also established a set of rules for working with multiple views [Wang Baldonado *et al.*, 2000]: the "rule of diversity" recommends the use of one view per information type and the "rule of parsimony" suggests using multiple views minimally. However none of these works has investigated the optimal number of detailed views to use, most existing techniques using 2 or 4 views. Finding such optimal number is thus still an open question.

### 3.1.3 Summary

In this chapter we present our contributions to the aforementioned challenges and related research questions (Figure 3.3). To address the need for fluid interactions in MDEs, we adopted two approaches: using an existing device, in our case a head-worn display interface that we named Gluey (section 3.2), or creating a novel dedicated device, namely TDome (section 3.3). Concerning the exploration of 3D data on public displays, we studied the use of around-device gestures to manipulate the detail view on a 3D public display (section 3.4). To explore multiple regions of graphs simultaneously, we studied the optimal number of views on an overview+detail interface (section 3.5).

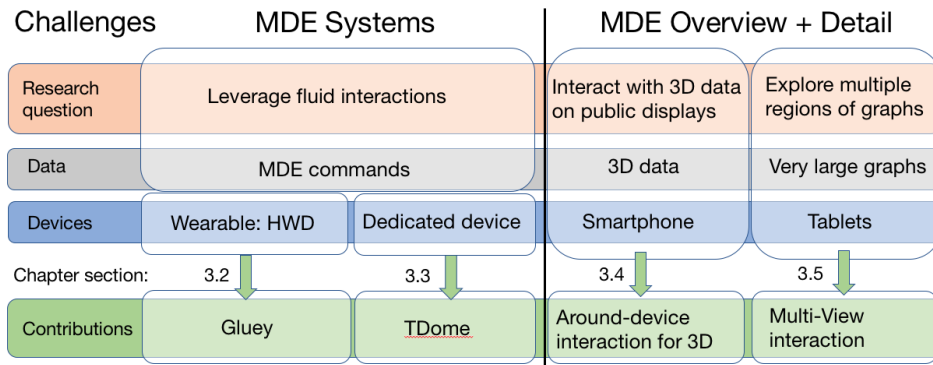


Figure 3.3: Research questions and our corresponding contributions to the domain of Multi-Display Environments.

## 3.2 Gluey: fluid interaction in MDEs using a head-worn display

To address the challenge of creating fluid interaction in MDEs, we consider what unique role can Head-Worn Displays (HWDs) play beyond simply being an additional display medium. Our contribution, named Gluey, is a HWD user interface that acts as a "glue" to facilitate seamless information flow and input redirection across multiple devices. This work carried in collaboration with Pourang Irani and Barrett Ens (University of Manitoba, Canada), and Xing-Dong Yang (Darmouth College, USA), was presented at MobileHCI 2015 [Serrano *et al.*, 2015].

### 3.2.1 Usage Scenario

We present a scenario depicting our vision of Gluey (Figure 3.4). John, an architect, relies on numerous digital devices while juggling between ad-hoc tasks in his daily work (Fig. a - b). Most of his drawing takes place on a desktop computer attached to two monitors and a printer. John uses a tablet for sketching and for showing his drawings to clients. He uses a smartphone for his communication needs.

John completes a plan for a gazebo on his desktop computer and prepares to use his tablet for presenting it to a client. After saving his presentation, John moves it with his mouse to a clipboard on his HWD (Fig. c), so that he can later copy it onto any other device. He then glances at his tablet and uses his mouse to grab the presentation from the HWD clipboard and transfers it to the device that is in the HWD's direct view, i.e. the tablet screen (Fig. d). Meanwhile, his business partner James sends a text message to his smartphone, noting the urgency of their soon to begin meeting. He is able to quickly reply to James, without opening his phone's soft keyboard. He instead uses the desktop

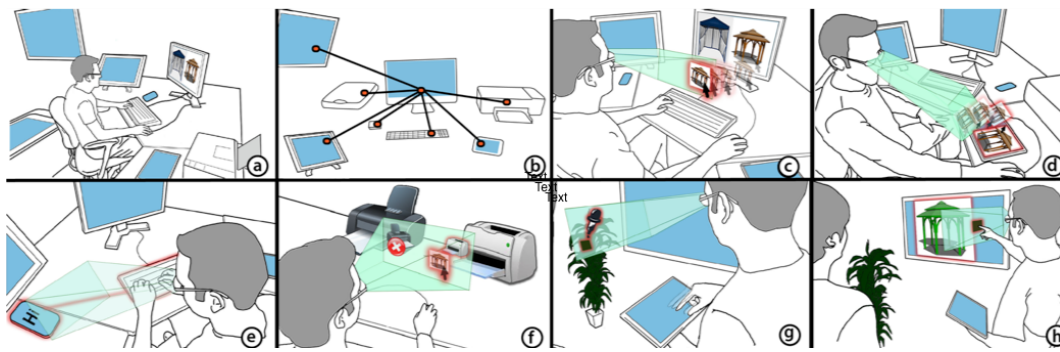


Figure 3.4: Usage scenario of Gluey.

keyboard, already near his hands, to type his reply. He does this by turning to look at his phone, which is detected by Gluey's head-mounted camera. As he types, the text gets redirected to the smartphone (Fig. e). Before leaving to meet his client, John prints a copy of the same presentation he had on the HWD clipboard by dragging its document icon to a printer icon, which comes in view when he looks in the direction of the printer (Fig. f).

At his client's office, he is led to a boardroom equipped with a large wall display for presentations. Since John has previously registered this display within Gluey's spatial configuration, he can immediately use his tablet as a trackpad to drag the presentation of his gazebo drawing from his HWD onto the large display. The client discusses different options, and suggests a different color for the gazebo, which matches the leaves of a nearby plant. John glances over to the plant, and selects its deep green color with Gluey (Fig. g), then fills in the gazebo using a finger gesture (Fig. h).

This scenario captures how John seamlessly interleaves many tasks in an ad hoc manner. He minimizes his need to switch input devices and can seamlessly migrate content to where it is needed by glancing at devices. While the actual data transfer is happening in the "cloud", John need only concern himself with simple, intuitive gestures for manipulating abstract representations of objects that suit his immediate needs. Furthermore, he can extract content from his physical surroundings (such as color or captured images) which he copies onto his digital display.

### 3.2.2 Gluey key components

To implement the previous scenario, we exploit the unique features of HWDs, such as view-fixed displays, cameras and inclusion of spatial sensors. We describe the three key components of our system using the terms Glueboard, Glueview and Gluon (Figure 3.5) as these define concepts in our implemented system but have slightly different meanings than what might traditionally be considered a "clipboard", "head-gaze input" and "input redirection".

#### Glueview

Many HWDs now include built-in sensors from which we can determine a user's head position in relation to the environment and parse details about what they are viewing. Because head motion dominates over eye movement when switching targets [Han Kim *et al.*, 2010], we can gain information based on a user's general head orientation, for instance which device they are looking at. We use the term Glueview to denote the use of Field of view (FoV) tracking (i.e. head orientation as a proxy for gaze tracking) for enabling implicit device registration: for example by simply having the smartphone in the FoV, the user can link his desktop keyboard and the smartphone (Figure 3.4 - e).



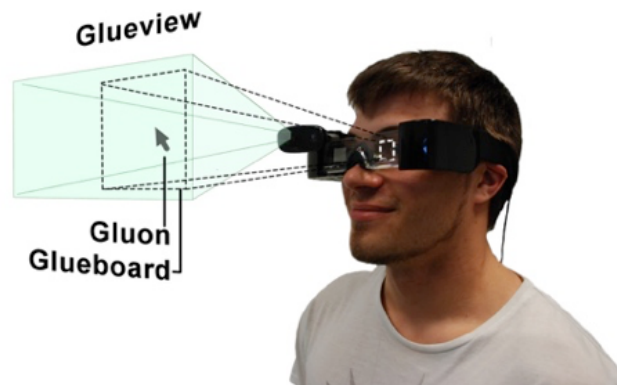


Figure 3.5: Three key components of Gluey: Glueview (field of view tracking), Glueboard (always-available feedback) and Gluon (input redirection). Together, these components enable several novel interaction techniques.

### Glueboard

Rather than treat the HWD as just another information display, we can use the device's view-fixed display as an always-available "canvas" for displaying visual feedback about interactions in the MDE. For example, this display can show text next to a keyboard, augment devices with useful status information or provide a visible clipboard space to store multiple data objects in transit between copy/paste operations to multiple destinations. We call this combined, always-available feedback and multi-object-clipboard space the Glueboard.

### Gluon

A main limitation of multi-device use is the need to change input modes when switching between devices. Since HWDs do not need to be held, the Gluey user can use any input device at hand, whether it be a keyboard, a mouse, a mobile touchscreen or mid-air finger gestures to control multiple displays. The Gluon represents the concept of pairing all available input devices with displays in the MDE to provide a unified interaction experience, independent of input mode or display type. For instance a user can use a desktop keyboard to enter text on a smartphone SMS (Figure 3.4 - e).

### 3.2.3 Gluey implementation

We created a proof-of-concept prototype using a HWD equipped with a webcam. We implemented all the features presented in our scenario: migration techniques across multiple devices; input redirection across multiple contexts; always-available visual feedback and environment's spatial model. All the implementation details can be found in the paper. We gathered preliminary user feedback by having 12 participants use Gluey for roughly 5 minutes for each technique. Overall, participants were positive at the concept of Gluey. They gave mostly favorable comments on content redirection features (copy and paste on the Glueboard) and input redirection. Seamlessly taking a snapshot and pasting it later on a device was another highly appreciated interaction. These results were encouraging taking into account the limitations of the prototype, such as its weight or the limited display FoV.

### 3.2.4 Summary

We introduced Gluey, a head-worn software interface that acts as a "glue" for bonding information across the various displays in a MDEs. Gluey exploits three key components: Glueview (i.e. using users' head-direction as a proxy to detect which device they are looking at), GlueBoard (i.e. using the device's view-fixed display as an always-available "canvas" for displaying visual feedback about interactions) and Gluon (i.e. using the input device at hand seamlessly between displays). Some challenges need still to be addressed to improve the concept of Gluey, such as helping users transition from their prior multi-display use (a mix of USB memory sticks, email and cloud services) to Gluey; avoiding unwanted device activations due to head orientation; and reduce the visual overload due to the always-available visual feedback.

## 3.3 TDome: fluid interaction in MDEs using a dedicated multi-DoF device

In this section, we present TDome, a novel touch-enabled 6DOF input and output device that facilitates interactions in MDEs. This project was carried by Houssein Saidi, a PhD student that I co-supervised with Emmanuel Dubois (IRIT - University of Toulouse), in collaboration with Pourang Irani (University of Manitoba, Canada). This work led to one publication at CHI 2017 [Saidi *et al.*, 2017].

### 3.3.1 Usage Scenario

We present an illustrative usage scenario with TDome prior to presenting its features. This scenario is inspired by our implication in neOCampus, a smart campus project at the University of Toulouse. Harry is an engineer working on a smart campus project that monitors data collected by multiple sensors on the university. To visualize and interact with the large datasets of energy consumption, the university has set up a multi-display environment composed of several displays, a large projection wall and two TDome devices resting on a tabletop.

As Harry enters the room to start his daily supervision of the energy data, he grabs one TDome and uses it to initialize the multi-display environment by simply pointing at each active display (Figure 3.6-A). He then selects the wall projection by rolling the device toward the wall (Figure 3.6-B). Harry decides to spread the data visualisation across two displays: he selects both displays with TDome and transfers the visualizations from one to the other with a TDome gesture (Figure 3.6-C). As he wants to look closer at information on the second display, he grabs TDome and walks towards the display, using the device in mid-air to perform a zoom on the data for a closer look.

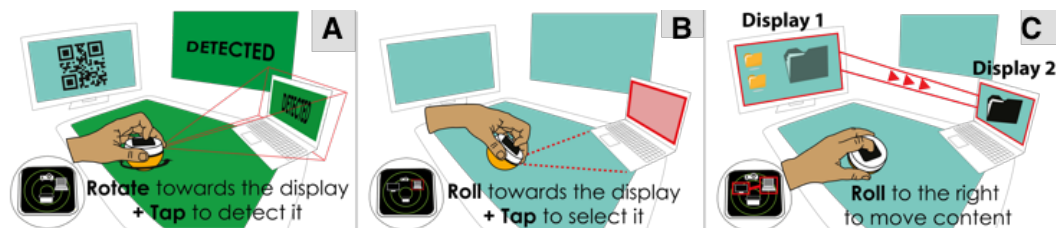


Figure 3.6: TDome facilitates several common tasks in MDEs, such as d) display registration using its embedded camera, e) device selection and f) cross-display data transfer.

Later that day, Mary enters the room and grabs the second TDome. They have a meeting to explore the university map to mark points of interest. Harry and Mary take their



personal smartphones and couple them with each TDome to benefit from personal interactions. Each smartphone shows a personal view of the projected map, which allows them to add and access personal annotations. Before ending, Harry wants to log onto the campus website and upload his annotations: he rolls TDome towards himself to display a virtual keyboard on the device's touchscreen and enter his personal password discreetly on the login page, displayed on the tabletop.

This scenario illustrates how TDome allows users to detect surroundings displays arrangement, select one display, move content between displays, reach to content at distant displays and perform personal interactions on TDome.

### 3.3.2 TDome key components

TDome results from the composition of a touchscreen with the Roly-Poly Mouse [Perelman *et al.*, 2015]. The Roly-Poly Mouse is a dome-like mouse providing rock, roll, translation and lift-up motions, initially developed at the University of Toulouse by Gary Perelman, a PhD student that I co-supervised with Emmanuel Dubois. TDome also includes a camera that can sense the environment. Regarding the touchscreens, we implemented both a Small and Large versions (Figure 3.7): to create the Small version, we enclosed a smartwatch touchscreen into TDome; for the Large version, the user can employ his personal smartphone. To support device modularity, the interchange of both touchscreens is easy and quick (Figure 3.7-right).

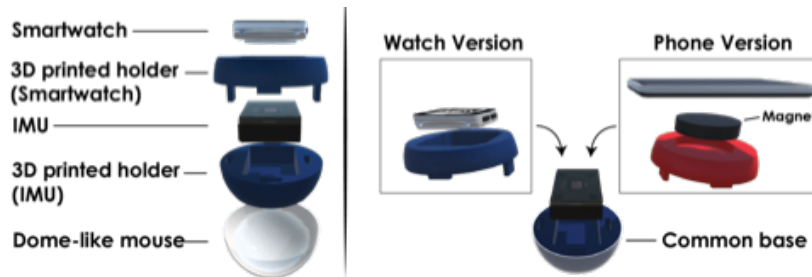


Figure 3.7: Arrangement of TDome elements.

TDome key components (in terms of spatial sensing, input and output interaction, mid-air capabilities and form factor) suit the major MDE interaction requirements [Boring, 2011]: input redirection (i.e. redirect input channels to different displays), output redirection (i.e. move content between displays), physical relationship (i.e. possess high-level information on the spatial layout of the displays), reachability (i.e. interact with a distant display) and personal data management (i.e. personal input and output interaction). We describe hereafter how each TDome component suits these requirements.

#### Spatial sensing

TDome physical manipulations allow performing 3D pointing in the surrounding space. Combined with the on-board camera, it allows sensing the environment. This can be used to detect and locate nearby displays, creating a spatial layout of the MDE displays represented through a radar-view, as illustrated in Figure 3.7 (physical relationship).

#### Input interaction

TDome allows up to 3 types of 2D pointing: by moving the device, by rolling it or by interacting with the touchscreen. These ranges of positioning facilitate input redirection.

This also offers input that best suits a given display, such as a cursor for precise tasks, or touch input for coarser input.

### Output interaction

The touchscreen display can be used as a visual buffer to move data among displays in MDEs (output redirection). It may also be useful to display a zoomed-in version of a selected area on a distant display (reachability). The built-in vibratory capabilities are an alternative to discretely provide the user with private information (personal data management). Through the easy interchange of the Small and Large TDome versions, the user can adopt the most appropriate display for each task; e.g., to visualize large graphs, the user can choose the Large version, but to display the compact radar-view (i.e. a view of the MDE spatial layout), a smaller display is more appropriate (output redirection).

### Mid-air interaction

Two of TDome's physical manipulations (roll and rotate) can be used in mid-air, thus facilitating physical displacements to interact with distant displays (reachability). It also offers more flexibility to the user to ensure the privacy for some of its tasks (personal data management).

### Form factor

TDome's tilting capabilities facilitate orienting the device towards oneself for private input and output interaction (personal data management); and attaching their personal smartphone to TDome's base allows users to access their personal applications and data (personal data management).

### 3.3.3 Evaluation

In our work we addressed two major challenges for applying TDome in MDEs: first, the device's usability, which demands the user to coordinate a physical manipulation with a touch gesture (we refer to as combined gestures - see Figure 3.8); second, the mapping between TDome gestures and MDE tasks.

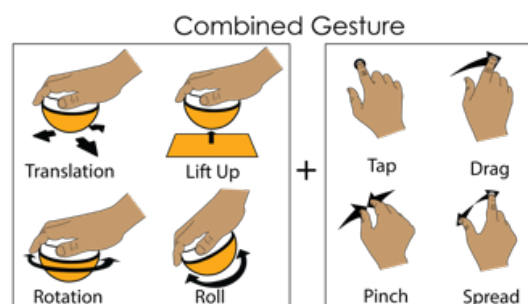


Figure 3.8: TDome combined gestures.

We explored combined gestures involving a physical manipulation (Translation, Roll, Rotation or Lift-Up) followed by a touch gesture (Tap, Drag, Pinch or Spread) through a 2-steps process. First, an exploratory study focusing on comfort established that 60 combined gestures could be comfortably performed. Second, a controlled experiment evaluated the user's performance as well as the subjective perceived difficulty. Results revealed that the

number of gestures that can be precisely and easily performed is 17 with the Small version, and 54 with the Large version.

Finally, a user survey explored the mappings between these gestures and MDE tasks. Results show that some combined gestures are more prone to be used in specific tasks than other. For instance, translation for panning or for moving a focus; roll for private pincode input; pinch and spread were preferred for zooming, drag for sending content from the tabletop to other displays and tap for display selection. In general, we found participants are able to match TDome features to MDE tasks.

### 3.3.4 Summary

We presented TDome, a device designed for interactions in MDEs. We designed two TDome prototypes: a Small version with an integrated touchscreen and a Large version based on attaching a smartphone. TDome allows versatile interactions that address major MDE tasks, which we illustrated through various proof-of-concept implementations: detect surrounding displays, select one display, transfer data across displays, reach distant displays and perform private interactions.

TDome interaction techniques still need to be fine tuned and future work should compare their performance with a baseline for each MDE task. Theoretically, since TDome integrates the same capabilities as existing MDE devices, we hypothesise that it can perform similarly for each individual MDE task. TDome should however improve the overall performance by reducing homing transition times and promoting the interaction flow. Therefore, beyond individual controlled comparisons, it would be interesting to carry a longitudinal study.

## 3.4 Interaction with Overview+Detail Interfaces on 3D Public Displays

In this section we present our work on mobile-based interaction with Overview+Detail (O+D) interfaces on 3D public displays. This work results from a collaboration with Louis-Pierre Bergé, a PhD student at the University of Toulouse supervised by Emmanuel Dubois, leading to one publication at MobileHCI 2014 [Bergé *et al.*, 2014].

In this work we focused on the translation task, i.e. how to move the Detail view (displayed on the smartphone) on a 3D environment, the Overview (displayed on the public display), as illustrated in Figure 3.9-left. Generally the user controls 3 degrees of freedom (DOF) to translate the point of view and 3-DOF to rotate the point of view. In our work, we limit user control of the Detail view to a 3-DOF translation. This task is sufficient to explore public 3D content such as museum objects, and it simplifies the task in a public setting, where interaction needs to be intuitive and straightforward.

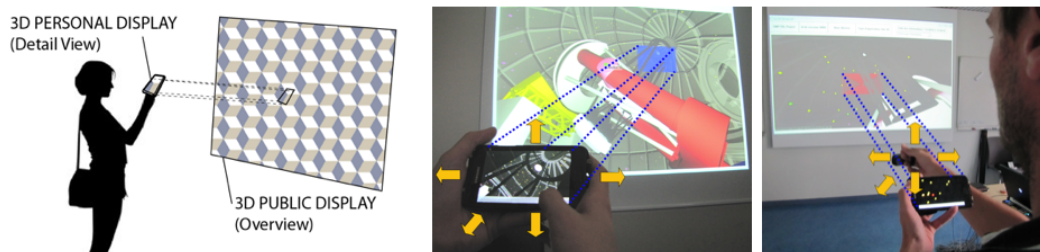


Figure 3.9: a) General setting of smartphone-based Overview+Detail interface on a 3D Public Display. We used two mid-air navigation techniques in a public installation to explore a 3D telescope visualization: b) Mid-Air Phone and c) Mid-Air Hand.

### 3.4.1 Interaction techniques for translating the detail view in 3D

We explored three different approaches to control the position of the Detail view based on previous work on mobile-based interaction with 3D content: moving the hand around the device [Kratz *et al.*, 2012], moving the device [Boring *et al.*, 2009] or using a touchscreen [Hachet *et al.*, 2009]:

- Mid-Air Hand (Figure 3.10 - a): the position of smartphone serves as a spatial reference. The position of the hand in this referential is mapped to the virtual position of the Detail view. We constrain the movement of the hand to the area behind the mobile phone. A virtual button on the mobile screen (de)activates this navigation mode.
- Mid-Air Phone (Figure 3.10 -b): translations applied to the mobile phone translate the virtual position of the Detail view. As for Mid-Air Hand, a virtual button on the mobile display (de)activates this navigation mode.
- Touchscreen (Figure 3.10 -c): inspired by commercial mobile 3D games, we use two rate-based joysticks to control the virtual position of the Detail view. The left circular joystick controls the 2D translation along the X and Y axis. The right cylinder joystick controls the 1D translation along the Z axis. Both pads can be used at the same time to control the 3-DOF navigation.

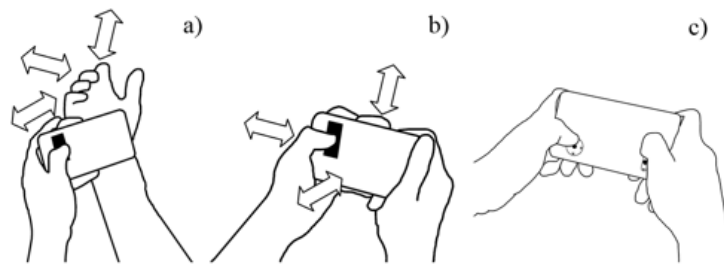


Figure 3.10: Three types of techniques we studied to translate a Detail view on a 3D Overview: a) Mid-Air Hand, b) Mid-Air Phone and c) Touchscreen..

### 3.4.2 Experimental results

We conducted a first experiment to evaluate the comparative performance of the three techniques presented in the previous section for the 3D translation of the Detail view. We asked 12 users to reach a target on a 3D scene using the Detail view. Our study revealed that techniques based on direct mapping (Mid-Air Hand and Phone) are better than those based on indirect mapping (Touchscreen) for controlling the 3D translation of a Detail view. The study also reveals Mid-Air Hand scores better in terms of attractiveness and user preference, although there is no significant difference concerning SUS score. An interesting result of our study is that Touchscreen, i.e. the most common technique, is the worst in terms of performance, of perceived attractiveness and of user preference.

These results were very encouraging and led us to further explore the two mid-air techniques in a second experiment: the goal was to evaluate the difficulty of performing the two mid-air techniques in usual public context, i.e. without training and without human explanation. Overall, results confirm that mid-air gestures can effectively be used to interact with O+D interface on 3D public display: 91.7% of participants have successfully used the Mid-Air Hand technique and 95.8% the Mid-Air Phone without any training or human explanation. These percentages would probably rise in a public context since users would

be able to imitate other participants as observed in [Walter *et al.*, 2013]. A surprising outcome of our study is that the Mid-Air Hand gesture is more difficult to understand and to perform at first than the Mid-Air Phone. Not only the success rate is higher for the Mid-Air Phone, but it also allows a faster interaction during the first trial. However, our results also reveal that after the first trial, both techniques are comparable in terms of task completion time. Interestingly our study shows that despite the initial difficulty, participants preferred the Mid-Air Hand technique.

### 3.4.3 Summary

We explored the design space of mobile-based interaction with Overview+Detail (O+D) interfaces on 3D public displays. We evaluated three mobile-based techniques from the literature to navigate in a 3D scene. One of the main findings of our experiments is that mid-air gestures are more efficient and preferred than touchscreen input for 3D interaction. Previous works on mid-air interaction with 2D content showed mixed results: some found mid-air interaction to perform as well as touchscreen [Jones *et al.*, 2012], while others found touchscreen input to perform better [Nancel *et al.*, 2011].

We employed the two mid-air techniques (Mid-Air Hand and Mid-Air Phone) in a public deployment at the University of Toulouse (Figure 3.9). The 3D scene, projected on a public display in the local university hall, represented a large telescope. The goal was to explore the different parts of the telescope and understand how it works. During two days, a large and varied audience (approx. 100 visitors composed of students, teachers and external public) explored the virtual dome of the telescope using these techniques. This in-situ installation revealed that selecting an object in the 3D scene with the hand handling the mobile phone can sometimes be difficult (due to the limited reach of the thumb). Designing alternative selection procedures should be considered in the future.

## 3.5 Splitting Detailed Views in Overview+Detail Interfaces

In this work we compared the use of different number of detailed views to interact with very large graphs in an overview + detail (O+D) setting composed of a large screen and a mobile device (tablet). This project was carried by Housseem Saidi, a PhD student that I co-supervised with Emmanuel Dubois at the University of Toulouse. This work led to one publication at MobileHCI 2016 [Saidi *et al.*, 2016].

### 3.5.1 Interface design

We designed and implemented an O+D visualization interface that consists of a large screen to display the contextual information and a tablet to show a magnified version of selected region(s) of the large space. We describe the three main views of our interface (overview, split views and translation view) and we analyze our design with the 8 rules for multiple views defined by Baldonado [Wang Baldonado *et al.*, 2000]: diversity (1), complementarity (2), decomposition (3), parsimony (4), space/time resource optimization (5), self-evidence (6), consistency (7) and attention management (8). All the details concerning each rule can be found in the original paper, we only describe here how our interface design suits these rules.

#### Split views

Our technique allows the user to have up to four independent split views at the same time, offering a detailed view on a graph region to support tasks requiring focusing on different places of the overview. We implemented three configurations for the multiple views on the tablet: 1-view, 2-views and 4-views (Figure 3.11). Using split views allows to decompose (rule 3) the complex graph rendering.

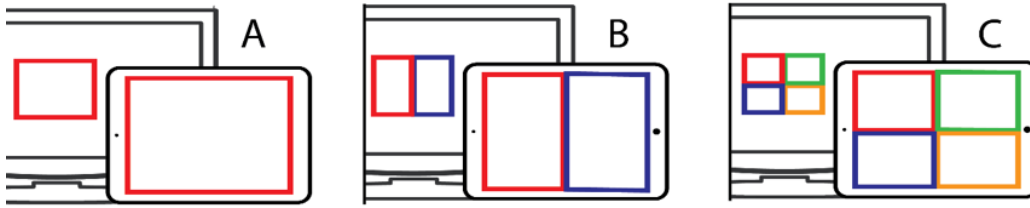


Figure 3.11: We studied splitting the detail view in 1, 2, and 4 split-views.

With the 1-view technique, the split view occupies the entire tablet display; with 2-views, each view occupies half; and with 4-views a quarter. For all of them, the zoom level is always the same, which means that as the number of views augment, the information displayed by each view decreases. This design conforms to the rule of consistency (rule 7) as the overall detailed area size is consistent over the 3 versions of our technique and when several focus are displayed their relative size is consistent as well. It also presents different conditions of space/time resource allocation (rule 5): sequential for 1-view, and side-by-side for 2-views and 4-views.

Swipe gesture inside one of the split views moves the underlying graph in the same direction: this behavior is consistent (rule 7) with regular map interactions on mobile devices. Finally, when the user selects a node in one of the split views, appropriate feedback is provided so that user's attention (rule 8) is focused on the appropriate view.

### Overview

The overview displays the entire graph on a large display (Figure 3.12). A contour color is applied to the split views on the tablet and to its representation on the overview to help the user establish the relationship between the points of view (rule 6).

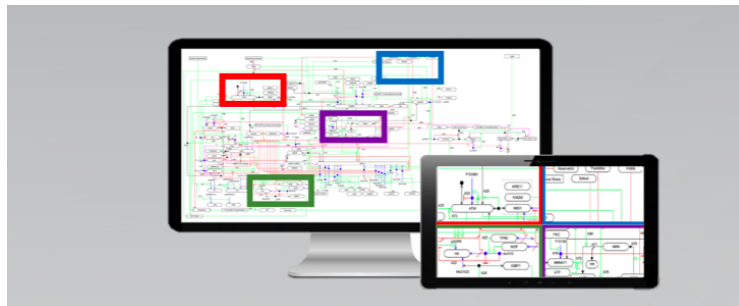


Figure 3.12: Illustration of the overview on a large display and the 4-views configuration on the tablet.

### Translation view

Positioning the split views relies on the use of the translation view on the tablet, which is activated when the user presses the "switch" button displayed on the tablet (Figure 3.13). The translation view provides a representation of the position of the 1, 2 or 4 split views on the overview. In the translation view, each split view position is represented using a view icon. Given the density of the graphs, displaying a miniature of it on the tablet would be useless. Therefore, the view icons are displayed on an empty background. By looking at

the overview, the user can use multiple (rule 1) view icons in complementarity (rule 2) for selecting multiple nodes.

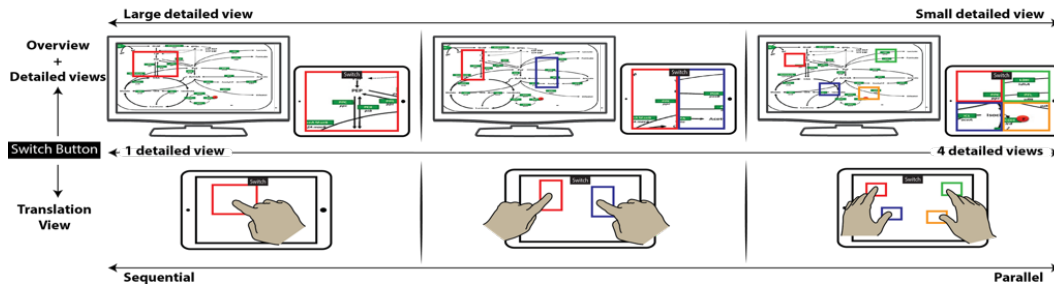


Figure 3.13: Arrangement of the 1, 2, 4 split-views configurations (top) and expected input control of the position of the views (bottom).

The user can adjust the position of one or several view icons simultaneously by direct touch manipulation. Using two hands and the multi-touch screen, the user can theoretically translate 4 view icons at the same time. Closing the translation view restores the split views. In our configuration, no zoom is allowed: this ensures a higher consistency over the split views (rule 7).

### 3.5.2 Experimental results

We conducted a controlled experiment to evaluate the effect of using multiple detailed views (1, 2 or 4) when connecting various number of nodes (2, 3 or 4) situated on different areas of large graphs. Participants were asked to create a connection between 2, 3 or 4 nodes. Overall, results show that using two or more split views is significantly faster than using only one detailed view. Results reveal that using 4 split views is only better than 2 split views for working on more than 2 regions of the graph.

An interesting finding of our experiment is that, when using 4 split views, users did not take full benefit of bimanual multitouch interaction to translate several view icons at the same time. Most of them (77%) used a sequential approach, first using one finger of each hand to move two icons, and then moving the two remaining view icons. While previous work on symmetric bimanual interaction (where each hand is assigned an identical role) has already highlighted its benefit in some settings [Balakrishnan and Hinckley, 2000] [Moscovich and Hughes, 2008], we are only aware of one work [Geyer *et al.*, 2012] exploring symmetric bimanual multitouch interaction (each finger performs a pointing gesture on a different target). In this previous work, up to 47% of the trials for some tasks were performed using multiple fingers in a bimanual setting. In contrast, our results indicate that symmetric bimanual multi-touch input is hard to perform. We believe these results are dependent on the task and we need to further explore the factors influencing symmetric bimanual multi-touch interaction.

### 3.5.3 Summary

Our work demonstrated that using two or more split views is significantly faster than using only one detailed view. Some challenges need still to be addressed to improve multi-view interaction. First, how to improve bimanual multitouch interaction to facilitate the translation of several split views at the same time. One idea could be to study combinations of fingers that can be moved synchronously and to help the user in employing these fingers. Second, as most participants used only one finger of each hand, we could consider other

potential uses of the remaining fingers: for example additional fingers might act as modifiers to bring split views together, or to move views to specific positions such as corners.

### **3.6 Chapter conclusion**

In this chapter we presented our contributions to the field of Multi-Display Environments through two main axes. First, proposing novel tools to facilitate the overall interaction on MDEs. We adopted two complementary approaches: using a head-worn display, or creating a dedicated multi-DoF device. Second, we explored Overview + Detail interfaces in MDEs with a focus on two specific contexts: 3D environments on public displays and very large graphs.





## Chapter 4

# Towards Displays Everywhere: FreeForm Interfaces

### Chapter content

---

4.1	Research challenges . . . . .	41
4.2	Collecting usage scenarios . . . . .	43
4.3	Displaying text on freeform displays . . . . .	44
4.4	Laying out content on freeform displays . . . . .	46
4.5	Chapter conclusion . . . . .	50

---

It is commonly accepted that interactive devices should have rectangular screens and, by proxy, rectilinear interfaces. However, recent breakthroughs in display technologies are enabling the design of non-rectangular displays and interfaces. Such displays are particularly adapted to fulfill the vision of pervasively displaying information anywhere. To this end, our goal is to facilitate the adoption of freeform interfaces that challenge many of the fundamental HCI principles. I started exploring this research challenge in 2016 in collaboration with Pourang Irani (University of Manitoba, Canada) and Anne Roudaut (University of Bristol, UK). The following work led to two major publications at CHI 2016 [Serrano *et al.*, 2016] and CHI 2017 [Serrano *et al.*, 2017].

### 4.1 Research challenges

The pervasive need to access and make decisions based on data is ushering the next evolution in ubiquitous computing: displaying interactive dynamic content where and when needed, a vision termed as displays everywhere [Pinhanez, 2001]. This vision covers a broad set of application areas, such as smart cities, in-vehicle displays or wearable devices. Displays everywhere can be leveraged to show energy consumption data in-situ in smart environments; facilitate collaborative pedagogy in university classes; leverage interactive presentations in museums, showrooms or shop fronts; or facilitate planning and decision-making during professional meetings.

Displays everywhere can benefit from recent advances in display technologies which allow creating non-rectangular displays. We can divide these technologies into three large groups: electronic systems, multifaceted systems and projection systems. In the field of electronic systems, Sharp recently introduced technologies to design arbitrary 2D display shapes (Figure 4.1-left). Multifaceted systems [Poupyrev *et al.*, 2006] use display primitives to compose larger displays (Figure 4.1-center). Finally, projections can be used to create

nonrectangular displays and can take place on arbitrary surfaces or surfaces with pre-computed geometries (Figure 4.1-right). All these technologies share the common property of allowing the creation of non-rectangular or freeform displays.

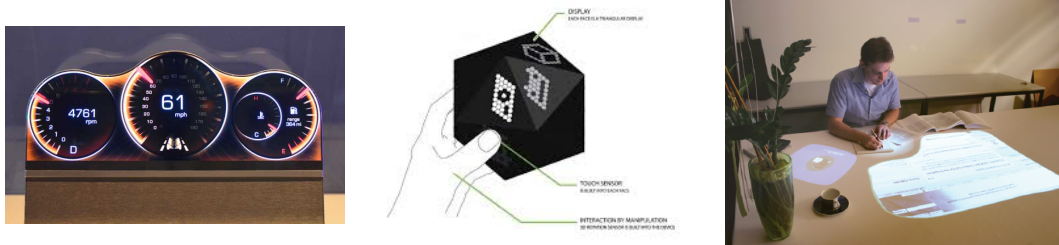


Figure 4.1: Different examples of non-rectangular technologies: Sharp electronic display (left), multifaceted system [Poupyrev *et al.*, 2006] (center) and projection-based display [Cotting and Gross, 2006] (right).

These non-rectangular displays can meet the needs of pervasive displays for which traditional displays are not well suited. For instance, a single non-rectangular display can replace current instrument panels on car dashboards. Non-rectangular displays will also facilitate inserting displays on non-rectangular objects, furniture and/or urban architecture, such as public road signs. In the context of mobile computing, non-rectangular displays can adopt shapes which could better fit wearable devices, allow more ergonomic hand grips or replicate existing jewellery, such as bracelets, pocket mirrors or round smart-watches, which are already commercially available.

However, such novel form factors challenge many of the fundamental Human Computer Interaction (HCI) principles and guidelines that have been accumulated over the past decades for presenting and interacting with content [Dix *et al.*, 2003]. Traditional WIMP interfaces are based on presentation principles such as the use of rectangular text areas and windows; the linear organization of menu bars; or the diagonal UI balance (title at the top-left, and buttons at the bottom-right). Similar principles apply to classical input interaction techniques such as rectangular selection by click and drag, or vertical text scrolling. Many of these principles will not still be valid on novel non-rectangular displays.

Previous work did only partially address the problem of interacting with non-rectangular interfaces. Specific graphical widgets have been developed to solve issues concerning occlusion on display surfaces (e.g. content occluded by the arm, or a physical object), such as nonrectilinear menus or circular layouts around physical objects. However, all these works considered very simple grid layouts (mostly circular) and used rectangular windows, which is space inefficient in a non-rectilinear surface. Cotting *et al* [Cotting and Gross, 2006] proposed mapping content that was originally designed for conventional rectangular displays into freeform bubbles (Figure 4.1-right). This technique is space-efficient but has not been evaluated to confirm its usability. In summary, previous work has not explored how presenting content on non-rectangular interfaces affects reading and viewing usability.

I started exploring the question of how to present content on non-rectangular displays in 2016 in collaboration with Anne Roudaut (University of Bristol, UK) and Pourang Irani (University of Manitoba, Canada). Our first challenge was to convene on the fundamental questions to investigate, since we were facing a vast research space resulting from the combination of multiple types of content and display shapes. Our approach was threefold: first, we carried focus groups to gather both concrete usage scenarios and display shapes; then, we studied text content only, being the fundamental brick of any UI; finally; we explored more complex content layouts combining text and icons.

## 4.2 Collecting usage scenarios

We decided to start our exploration by collecting usage scenarios of free-form displays in order to generate display shape properties that would inform our choices of shape categories in further quantitative studies [Serrano *et al.*, 2016].

### 4.2.1 Data collection

We first brainstormed with hci students to capture a subset of compelling shapes in terms of displaying and interacting with content. We ran two focus groups with 20 participants in 2 countries (France and the UK) to maximize the diversity of scenarios we could collect, and to avoid cultural biases (albeit, both countries are dominated by Western culture). We collected 62 ideas depicting 41 shapes once redundancies were eliminated. Most were 2D, and 3D ones were represented using 2D likenesses that corresponded to the user's point of view (e.g. a circle for a sphere).

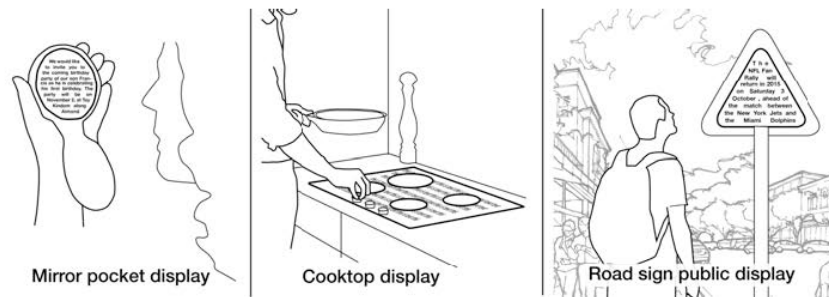


Figure 4.2: Freeform display usage scenarios collected during our focus groups.

The usage scenarios collected during the two focus groups illustrate the diversity of shapes that can hold content, such as circular mirrors for private notifications, shapes with holes such as a cooktop displays for recipes or the back of triangular road signs as public displays, as illustrated in Figure 4.2. Interestingly, in most cases, existing artifacts having nonrectangular features were suggested for content augmentation. Some examples included placing text on road signs, kitchen cooktops, pocket mirrors, puzzle pieces, bike handles, shoes, drink cans, and electric plugs, among others.

### 4.2.2 Display shape properties

To analyze the collected shapes, we used a clustering algorithm similar to the one proposed in [Roudaut *et al.*, 2014] to create groups of shapes and extract shape properties. From this analysis we observed a set of display shape properties:

- Symmetry: overall there are slightly more symmetrical shapes than non-symmetrical.
- Curvature: most of the shapes are ovoid in nature, for example car side mirrors, purses, sinks, or oval tables. Some shapes have sharp boundaries such as a triangle (road signs), miniature house shape, a tee shirt, or a cooktop.
- Porosity: we found several shapes with holes, such as bathroom elements, electrical plugs, glasses, and cooktops.
- Length-Width ratio: a good number of samples included long and thin shapes such as pencils, faucets, chair arms, or belts.

- Orientation: we observed that some displays had particular orientations, for instance the handle of a frying pan.

In addition to these display properties, we also noted additional interaction observations. For instance, many scenarios involve a display content that needs to be scrolled to access more information, such as the case of the cooking jar or the umbrella.

### 4.3 Displaying text on freeform displays

After this initial exploration of the usage scenarios and display shapes, we tackled the core concern of how to display text on a screen that is non-rectangular. Reading text is fundamental to many tasks including visually scanning, flicking through a document for specific content or displaying icons. However, running a large study comparing text legibility on multiple shapes is difficult because of the high dimensionality of possible topologies. To address this issue, we propose a text mapping framework.

#### 4.3.1 Mapping framework

We proposed a framework that aims at presenting different mappings of text content onto arbitrary shapes. The framework describes three axes with increasing levels of abstraction. This list is non exhaustive as we only considered text mappings that relate to readability, e.g. we dropped cases with upside down text.

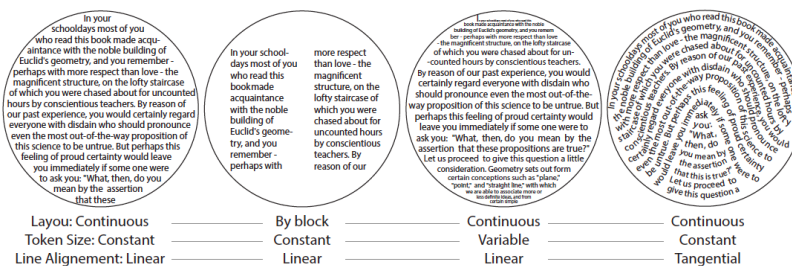


Figure 4.3: Illustrative examples on a circle of different types of text mappings according to the three axes of our mapping framework.

We proposed the three following axes, illustrated in Figure 4.3 :

- Layout: this axis describes the general text layout, which can be continuous or by block. For example, the CHI Proceedings layout is in blocks (formatted on two columns). We could have also considered the case where the layout is not continuous (e.g. random), but this would clearly disturb text readability.
- Token size: this axis describes the size of the tokens, which can be constant or variable. E.g. the fisheye menu illustrates the case variable.
- Line alignment: this axis describes the line alignments in which the text fits. It could be linear, i.e. horizontal, or oriented parallel lines, or what we call tangential, i.e. following the shape. More precisely, text could follow a vector field around the shape boundary. This is typically the case in calligrams.

We then drew on relevant text legibility work and formulated 10 hypotheses to predict how the mappings affect legibility when displayed with different display shape properties. These hypotheses relied on existing knowledge on text legibility, but also extended it

as we were unaware of any study investigating text legibility on non-rectangular shapes. For instance, because we are familiar with reading text that is aligned to the left, we can assume that return sweeps will be more difficult when the text is not left aligned. We then carried a set of quantitative studies on different display shapes to validate or invalidate our predictions. We provide an overview of the studies and the major findings in the following section.

### 4.3.2 Quantitative studies on reading text

#### Overview

To examine our 10 hypothesis on how text mapping affects legibility on non-rectangular shapes, we carried four experiments in total (Figure 4.4):

- Experiment 1: We compared shapes with different left or right text alignments.
- Experiment 2: We compared different text layout on various shapes with or without a hole.
- Experiment 3: We compared different token sizes on different shapes. We also wanted to compare the impact of continuous scrolling vs. page scrolling on text legibility.
- Experiment 4: We compared different line alignments on different shapes.



Figure 4.4: Illustrations of some of our experiment conditions. Left, different left alignments (study 1). Center, text layouts on shapes with or without a hole (study 2). Right, tangential line alignments (study 4).

#### Studies design

These four experiments were based on the same task and procedure. Reading tasks need to be carefully designed so they bear resemblance on how we commonly read. Two primary task options exist. In one case, the post-reading comprehension of users is evaluated using procedures such as the Nelson-Denning reading test. However, this test is designed primarily for gauging reading deficiencies. A second approach consists of seeking spelling mistakes or finding specific words. Such tasks promote skimming. We adopted a task similar to that of Jankowski et al. [Jankowski *et al.*, 2010], which introduces word substitution errors, forcing participants to read and comprehend sentences. Incomprehensible sentences need to be flagged for errors and subjects must read the entire passage to recognize all substituted words. The new words are common words that are placed grossly out of context.

As in prior work [Jankowski *et al.*, 2010], we measured text legibility by both examining reading time and reading accuracy. We focused on short text (150 to 170 words) as a result of our brainstorming sessions. Using longer texts may have shown more differences in results, but small passages are ecologically valid and in line with the scenarios we gathered. A total of 37 people (8 female) with normal or corrected to normal vision took part in our experiments.

### Resulting design guidelines

From the results of our studies we can provide a set of design guidelines for optimizing text legibility on non-rectangular displays:

- Both left and right irregular alignments should be avoided, as text in these are perceived to be difficult to read and overall not aesthetic. Instead, symmetric shapes are preferred.
- Shapes with circular or sharp alignments are acceptable for presenting text: they are perceived to be easy to read, and overly clean, beautiful and interesting.
- If the shape contains a hole, text should be displayed using a broken layout with two columns around the hole to prevent any impact on reading performance.
- Shapes without holes are perceived to be less interesting than with holes. Thus, using holes in freeform shapes is not only a solution to context requirements (such as the cockpit), but also an aesthetical feature to explore.
- To use dynamic scrolling on non-rectangular shapes, text should be resized so that each line contains the same amount of text. Otherwise, use page scroll with constant text size.
- While resizing text for dynamically scrolling is perceived as beautiful and clean, resizing text with page scrolling raises mixed results. Some users disliked it because of display space loss and of varying interline spacing. Thus, resizing text should be limited to dynamic scrolling.
- Shapes with continuous line alignment where lines are cut by the shape curvature should be avoided as they are perceived to be difficult to read and non aesthetic. This is similar to the effect of holes on continuous text. Even though tangential alignment does not affect reading performance on linear shapes, continuous text should be preferred as it reduces the perceived difficulty.
- Text on very sharp shapes should be avoided, as text on these is harder to read than on linear shapes. If used, such shapes should be filled with continuous text rather than tangential that impacts reading performance.

## 4.4 Laying out content on freeform displays

After studying how to map text onto non-rectangular displays, we extended our investigations to more ecological content, i.e. combining text and images, which lead us to address the fundamental question of how to layout content. Our goal was thus to investigate if the established composition principles for traditional displays generalize to non-rectangular displays, and if not how they can be adapted for designers to create the layout of freeform display content.

### 4.4.1 Elicitation study

We first performed a qualitative study that consisted in asking graphic designers to map traditional web content onto non-rectangular shapes. We recruited 5 professional graphic designers with expertise in print or web design. We gave them a webpage (the home page of The Guardian) with all associated content. We asked them to fill this content into four shapes: a circle and a triangle, with or without hole (Figure 4.5).

All designers found that shapes with hole were more difficult to fill and were particularly not satisfied with designs on the triangle with hole. Designers agreed that the easiest shape was the circle without hole. Besides collecting the designers subjective feedback on





Figure 4.5: Examples of graphical designs collected during our elicitation study.

satisfaction and difficulty, we analyzed the resulting productions using previously existing design composition principles.

#### 4.4.2 Composition principles

Presenting content on rectangular displays is reasonably well understood and there exists numerous composition guidelines. In our work, we analyzed the composition principles proposed by Galitz [Galitz, 2007]: these principles consist in aesthetic composition guidelines extracted from tacit knowledge that visual designers have accumulated over years of experience (e.g. balance, proportion or unity).

Using the probes collected during our elicitation study with graphic designers, we examined how the initial definition of the composition principles could be adapted to non-rectangular displays. As in our previous work on text mapping, we suggested hypotheses on how these existing composition principles would generalize to freeform interfaces, and we evaluated them in a set of studies. The analysis of the results showed that graphic designers' inner sense for composing layouts matches existing composition principles (simplicity, sequentiality, economy, proportion and unity) but that some revisions (balance, regularity, predictability) are needed:

##### Balance

In the original definition, *balance* means providing an equal weight of elements on each side of the horizontal or vertical axis. There are two aspects that can change the way we define Balance when moving to non-rectangular displays: (1) the symmetrical axes of the display and (2) the definition of a regular shape. For the symmetrical axes, examples produced by our designers suggest that balance should follow the vertical axis. Most element shapes in the productions were rectangular except for the designs shown in Figure 4.5-right. In these two designs the elements shapes were directly related to the shape of the display (the designed cut the elements in circle or triangle).

##### Regularity

*Regularity* means providing consistently spaced horizontal and vertical alignment points and spacing, as well as using similar element sizes, shapes and colors overall. Similar to Balance, there are two aspects that can potentially change the way we define Regularity when moving to non-rectangular displays: (1) the alignment axes which can be more than just horizontal/vertical and (2) the definition of a regular shape. Concerning the alignment axes, it is possible to imagine different layouts and to deviate from the rectangular grid: in fact only 10/20 productions used a rectangular grid, 4 used a radial alignment and 6 used a tangential alignment (aligned with one or more edges of the screen). Concerning the regularity of element shapes we observed that designers reshaped the elements for two main purposes: to fit the shape and for aesthetics.



### Predictability

*Predictability* means providing conventional orders or arrangements to ensure that viewing a screen enables one to predict how another will look. Predictability is also enhanced through design Regularity. This guideline rather links the way several pages are designed and thus we have no reason to think that this should change with the shape of the device. To a large extent we can also couple this definition with some of Nielsen's guidelines of "consistency across platform", e.g. the fact that headers and menus are always at the top of a webpage. While all designers decided to keep the regular menu position at the top of the shape for the Circles, most of them inversed the position for the Triangles. This change is rather surprising, given that it goes against traditional web page layouts.

### Other composition principles

Our analysis suggested that the other composition principles of Galitz should not change on freeform displays: *Proportion* (using aesthetically pleasing proportions for components of the screen); *Simplicity* (optimizing the number of elements on a screen as well as minimizing the alignment points); *Sequentiality* (arranging elements to guide the eye through the screen); *Economy* (providing few styles to deliver the message as simply as possible) and *Unity* (using similar sizes or shapes for related information). We had no reasons to believe these would change so we did not investigate these principles further.

### 4.4.3 Experimental studies

As in our previous work on text mapping, we suggested hypotheses on how these three existing composition principles (balance, regularity and predictability) would generalize to freeform interfaces, and we evaluated them in a set of online surveys.

#### Surveys design overview

To evaluate the differences between the composition principles we opted for paired comparison experiments which consist in asking participants to choose between two conditions, here two layout visualizations. The experiment is designed so that each participant rates each pair of visualizations. Pairwise comparison ratings have been proven to produce more realistic results than asking for individual rankings (e.g. using a Likert scale). We asked participants to compare pairs of layout visualizations and say which one was nicer (i.e. visually pleasing), clearer (i.e. not confusing) and more symmetric (aesthetics terms proposed in [Lavie and Tractinsky, 2004]).

We used the same shapes as in our previous studies, i.e. a circle and triangle, with and without a hole, and we compared visual compositions among shapes, but not between shapes. In a follow-up experiment, we systematically explored shapes with increasing number of edges: triangle (3), trapezoid (4), pentagon (5) and hexagon (6). We also included an inversed triangle to see if the orientation of the shape had any effects. We gave four surveys to the participants matching different hypotheses:

- Survey 1: Balance and symmetry (Figure 4.6). We studied 3 symmetry axes (vertical, shape and all) and 2 element shapes (rectangular or matching display shape).
- Survey 2: Regularity (Figure 4.7). We studied four grid layouts (regular, radial, oriented and random).
- Survey 3: Regularity (Figure 4.8). We tested whether it was better to follow the regularity but have elements cut by the display shape or to break the regularity by having the elements fit the shape. We tested 2 conditions (elements out or in).
- Survey 4: Predictability (Figure 4.9). We changed the position of the menu. We tested 3 positions for the menu (top, bottom and following the shape).

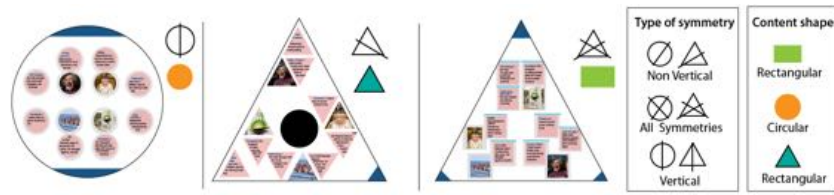


Figure 4.6: Survey 1 conditions.

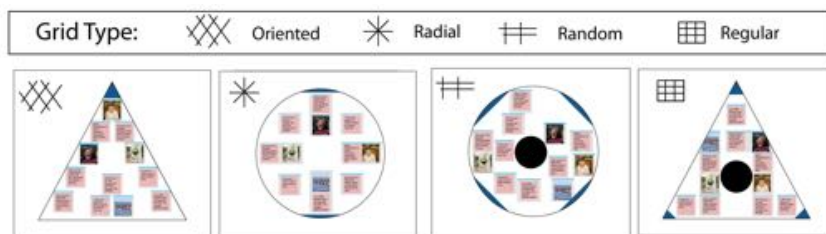


Figure 4.7: Survey 2 conditions.



Figure 4.8: Survey 3 conditions.



Figure 4.9: Survey 4 conditions.

### Summary of studies results

Based on our findings we propose a set of guidelines for laying out content on non-rectangular displays. Some of these design guidelines contradict current conventions on rectangular displays.

- **Symmetry axis:** The symmetry axis should be vertical to ensure that the final design is nice, clear and symmetric.
- **Content shape:** Instead of using the traditional rectangular boxes for text or images, designers can reshape the content to fit the display (circular on circle, triangular on triangles, etc.). This reshaping will have different effects depending on the display shape: it will look nicer with circular content, or more symmetric with triangular content. However designers should be aware that sometimes reshaping content might make it appear less clear (such as in our triangle condition).
- **Grid layout:** While using the traditional regular grid works well for certain shapes (regular and inversed triangles), using a grid with the same shape as the display shape can make the overall design look pleasing, clear and symmetric (as with radial grids in circle, pentagon and hexagon displays). A non-regular grid can benefit from non-rectangular content, as it better fits the shape of the grid (triangular content in oriented grid for instance).
- **Breaking content:** To solve the problem of content not fitting exactly on the display, designers should favor breaking the regularity of the grid and making all content fit, rather than cutting elements by trimming the edges.
- **Menu position:** While placing the menu at the traditional position on top of the interface works best for triangle and hexagon displays, designers could place it at the bottom in certain cases: this is a position that is nicer, clear and symmetric for a circular display, and that is equivalent to the top position for certain shapes (pentagon and trapezoid).

### Generalization to other shapes

Since this work is the first exploration on how visual composition principles apply to non-rectangular displays, we decided to adopt a context-independent approach. We chose to study the generic properties of layout design instead of focusing on a given interface for a given application. The reason is that we wanted to provide generalizable findings rather than narrow in on specific guidelines that would be only valid for a specific case. Our choice of shapes was based on the usage scenarios envisioned for non-rectangular displays. Some of our results seem to be consistent across shapes, such as the fact that shape-like content looks better than a rectangular content, suggesting that they are probably valid for other shapes. Other results seem to depend on the display shape, such as the layout grid: while a radial grid is best for most display shapes (circle, hexagon, pentagon and trapezoid displays), a regular grid is better on triangular displays.

## 4.5 Chapter conclusion

In this chapter we presented our contributions to tackle the challenge of how to display information on freeform interfaces. The introduction of such displays creates unprecedented challenges for designers who have to rethink new ways of creating user interfaces. The foremost concern is how to legibly present textual content. Our results agree with and extend upon other findings in the existing literature on text legibility, but they also uncover unique instances in which different rules need to be applied for non-rectangular displays. In a follow-up exploration, we studied how traditional content layouts can be adapted to fit

different non-rectangular displays. Based on our findings we propose a set of guidelines, some of which contradict current conventions on rectangular displays. Our work is a first step toward defining new guidelines for the design of free-form displays. We detail some perspectives to this research challenge in the last chapter of this manuscript.



## Chapter 5

# Research methods for ubiquitous interaction

### Chapter content

---

5.1	Design methods . . . . .	53
5.2	Prototyping methods . . . . .	58
5.3	Evaluation methods . . . . .	62

---

The most common approach for building UIs is to adopt a user-centered iterative design process. This process involves four main activities [Preece *et al.*, 2001]: identifying needs and requirements; developing alternative designs; building interactive versions of the designs; and evaluating these interactive versions with users. While this process relies on relatively well known methods when building desktop-based UIs, the characteristics of ubiquitous interaction call for rethinking these traditional methods. Without being exhaustive, in this chapter we sum up the methods we used to create the ubiquitous interactions presented in the previous chapters. For some of them, we proposed novel approaches which represented a contribution by themselves (marked with an \* in the title of the section).

### 5.1 Design methods

The first two activities of interaction design, i.e. gathering needs and developing design alternatives, can be challenging when considering ubiquitous interaction. First, the activity takes place in external environments, requiring **in-the-wild observations (1)**. Second, the set of potential interactions and mappings to interactive tasks is usually quite large, leading to the use of novel design methods, such as **design through device composition (3)** or **user-elicitation (or guessability) studies (3)**. Finally, some ubiquitous interaction scenarios are prospective, i.e. exploring future possible usages, making it difficult to ground the design on existing practices. In this case, it can be interesting to gather **design probes (4)** from designers to generate novel usage scenarios. Hereafter we describe how we applied each one of these three methods in our work.

#### 5.1.1 In-the-wild observations

Observing users early in the design process can help designers to understand the users' context, task and goals [Preece *et al.*, 2001]. Observations may take place in controlled environments, such as a laboratory, or in the field, where people are observed performing

their tasks in a natural setting. Such observations can have different degrees of participation, from a passive observation ('outsider') to an active observation ('insider'), where the observer is part of the group under study. As mobile technologies are now common in public spaces, it becomes easier to carry in the wild 'outsider' observations, which can be very valuable to unveil unanticipated usages, difficult to observe in a more active observation. Such unpredicted usages are important to consider early in the design process of ubiquitous interactions, as they usually reveal how people gets around current technology, opening new research paths.

To illustrate how in-the-wild observations can be used to unveil unexpected usages, we report on informal observations made during a workshop on Interaction Techniques for Mobile Collocation that I co-organized at MobileHCI 2016 [Lucero *et al.*, 2016], on September 6, 2016 in Florence, Italy. During the workshop we discussed F-formations: the term F-formation (or facing formation) was originally coined to describe the spatial arrangement of people in social encounters. Later, proxemics prototypes have been developed that exploit F-formations to support mobile collocated interactions. Inspired by the work presented at the workshop on dynamic F-formations in non-traditional settings, the workshops organizers together with the workshop participants decided to go out and make exploratory F-formation observations in the wild.

During this in-the-wild observation session, participants (n=12) were to observe and make annotations of anything that would seem unusual or that had not been previously reported in previous studies of F-formations in controlled settings, some of which were discussed during the workshop. Such observations could include information on group sizes, how groups move in an open space, physical distance between people, or their potential use of devices. Participants were split into three groups of four and were asked to observe formations of tourists around the Dome of Florence Cathedral (i.e., Il Duomo) pedestrian area.

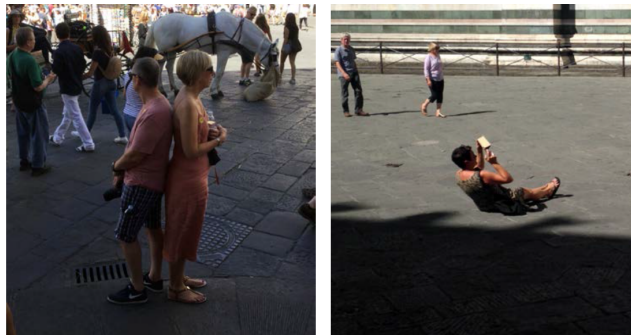


Figure 5.1: Examples of an unusual F-formation (left, back to back) and phone usage (right) captured during in-situ observations.

The results are fully reported in [Lucero and Serrano, 2017]. We observed a mix of tourists and locals doing different activities in this space: tourists walking alone or as small groups, shoppers carrying bags, persons resting by sitting on or lying down on benches, families carrying suitcases or with strollers, people riding or walking next to their bikes, persons walking dogs, mobile street artists selling their work. While we observed activities that are closely related to tourism and thus one would expect to encounter in such a context, we also observed some unusual formations that have not been previously reported in F-formation studies conducted in controlled settings, for instance a standing back-to-back formation (Figure 5.1-left). Another unusual situation consisted of a lady who broke away from a group to capture a picture of the Duomo (Figure 5.1-right). As she wanted to capture as much of the Duomo as possible, or perhaps from a particular angle, she decided

to sit on the ground to take a low-angle shot (Figure 6, right). While this picture does not allow us to say much about F-formations themselves, it does again help us make a point about what we gained by going into the wild to make such observations, as these have not been reported in F-formation studies in controlled settings.

### 5.1.2 Research through design by device composition (\*contribution)

Ubiquitous interaction often implies rethinking existing interaction devices. When coming up with a novel interaction device, HCI practitioners often follow an empirical design approach resulting in the development of ad-hoc solutions, usually combining existing devices, without relying on a systematic or structured process.

To overcome this limitation, we investigated the concept of device composition to promote the potential of combining existing devices to create new ones, and to leverage performing this combination in a more systematic manner. Device composition consists in physically putting together several existing devices to create a new one, hereafter referred to as a compound device. To develop this concept, we defined, illustrated and evaluated a design space for the physical combination of interaction devices, DECO [Perelman *et al.*, 2016].

Our design space, DECO (Figure 5.2), is structured along two dimensions: 1) the physical arrangement, i.e. how the different devices are physically composed and 2) the physical manipulation, i.e. the way a user will manipulate each element of the compound device.

		Physical arrangement				
		Enclosed	Glued			Separated
Topology		Enclosed	A-B	A<-B	A<->B	Separated
Fusion type			A-B	A<-B	A<->B	
Temporal usage	Sequential	GlobeMouse	Hexagonal bezel-less ★ FlipKeyboard Codex	Trackball Keyboard LensMouse	Ink222	Hexagonal bezel-less ★
	Parallel	Lexip 3D Inflatable Mouse	Razer naga Dualscreen HybridTouch	PadMouse CubicMouse RPM2 ★	Mouse 2.0	Siftable HoverPad RPM2 ★

★ = Dynamic compound device

Figure 5.2: Different compound devices from the literature classified using DECO.

To validate our design space, we used the approach proposed by Beaudouin-Lafon [Beaudouin-Lafon, 2004] to evaluate design models. This approach is based on three properties characterizing the ability of a model to describe existing solutions (descriptive power), compare existing solutions (evaluative power), and generate novel solutions (generative power). Using DECO, we classified and compared existing devices to illustrate how our design space helps in describing and comparing different solutions. We also used this design space to elaborate a novel compound device supporting multi-dimensional interaction: this device results from the combination of a spherical mouse, the Roly-Poly Mouse [Perelman *et al.*, 2015] with a traditional laser mouse. Through our 3-part exploration, we demonstrated that DECO is a useful design space that can be used to describe, compare and generate novel compound devices.

### 5.1.3 Elicitation studies: exploiting similarity over agreement (\*contribution)

To explore the breadth of potential interaction gestures and their mapping to interactive tasks, one approach is to elicit user input through an elicitation (or guessability) study. For



exploring potentially rich and vast gesture sets, user elicitation or guessability studies have shown favourable results. Wobbrock et al. found that eliciting gestures from users resulted in over 40% more gestures than if asked by expert designers [Wobbrock *et al.*, 2009]. This motivated the use of such an approach to identify gestures for multitouch tabletops, mobile motion gestures or for foot interaction among others.

During our exploration of Hand-To-Face gestures [Serrano *et al.*, 2014a], we asked participants to suggest suitable gestures on the face (above the neck) and on the HWD. We put aside any recognizer issues and asked users to perform gestures at their will without worrying about the underlying sensing technology. We asked participants to include gestures on the entire face, i.e. any region on or above the neck. This allows for a larger set of potential gestures. We also assessed users' preference for interacting with either the face or areas of the HWD, for each of the tasks.

We analyzed the agreement between participants for the set of gestures produced for each task using Wobbrock's approach [Wobbrock *et al.*, 2009]. The agreement value ranges between 0 (no agreement) and 1 (total agreement) and indicates whether users agreed on using a specific gesture for a given task. We group gestures which are of the same type (swipe, tap, etc.) and occur in the same area (cheek, chin, etc.). The mean value for the agreement score was 0.14 (SD 0.06), with 36% of the tasks having an agreement value higher than 0.2%. While this score seems low, it is on par with that from other previous guessability studies. This low agreement score is mainly due to the variety of areas used for zooming (ear, hair, nose, neck, mouth, jaw, forehead, chin and cheek). Swiping the cheek is considered different than swiping the forehead. To solve this issue, we suggested **using similarity instead of agreement**.

To this end, we first proposed a taxonomy to describe hand-to-face gesture properties to evaluate gesture similarity. Our taxonomy includes five properties to describe the gesture mapping and its physical characteristics (nature, temporal, pose, number of fingers and area). Based on the previous taxonomy, we defined a formula to calculate the similarity score, which indicates whether different gestures share common properties. The similarity score  $S_t$  of a task  $t$  is the average of the agreement for every property  $P_i$  of our taxonomy, from the set of properties  $P_t$ . To calculate this value we use the formula of the agreement score:  $G_i$  is the subset of gestures with identical value for the property  $P_i$  from the set of gestures  $G_t$ .

$$S_t = \frac{\sum_{P_i} (\sum_{G_i} (\frac{G_i}{G_t})^2)}{P_t} \quad (5.1)$$

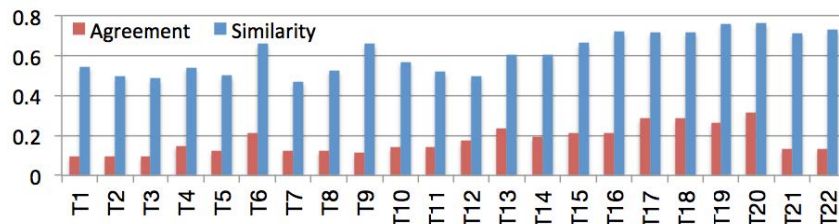


Figure 5.3: Comparison between agreement and similarity scores.

Using this formula, we gathered some more informative results on the proposed gestures. The mean value for the overall similarity score (figure 5.3) is 0.61 (SD 0.1). The Nature of zooming gestures (0.86) is mainly based on the metaphor of pinching with two fingers. The Pose for zooming is always dynamic (1.0) and most subjects used two fingers for zooming (0.80). Actually, our similarity analysis allows also to describe gestures more precisely: we can describe panning gestures using our taxonomy as <abstract, continuous, dynamic,

one finger, cheek/forehead> and zooming gestures as <abstract discrete/continuous, dynamic, two-fingers, cheek/chin/forehead>.

Our approach only addressed one issue of guessability studies, i.e. the limitation of using the agreement as unique score. This problem was also pointed out in a recent work [Tsandilas, 2018] that identifies other problems of the analyses methods for these studies (such as the interpretation of the agreement values or the statistical inferences) and suggests some alternative solutions. As practitioners in HCI, i.e. a relatively recent research field, it is our responsibility to question and update existing research methods, particularly when they are as widely employed as guessability studies.

#### 5.1.4 Design probes

The first activity in interaction design is to identify the needs and establish requirements. However, one of the key challenges when designing prospective ubiquitous interfaces is to identify the needs that will emerge from implementation of novel ubiquitous services. In such cases, working with designers can help generating novel usage scenarios, leading to design probes that will feed the analysis on needs and requirements.

We adopted this approach when working on non-rectangular displays [Serrano *et al.*, 2017]. To begin our exploration on the visual composition of graphical elements on non-rectangular displays, we first captured how graphical designers tacitly organize visual layouts. We gathered qualitative probes that we could use to generate new hypotheses for the visual composition of elements on free-form displays.

We asked five graphic designers to compose webpages on non-rectangular shapes (Figure 5.4). This task relies heavily on creativity and thus we designed this study to be in the form of homework. Designers had a week to do the task wherever they wanted, thus avoiding us to interfere with any creative processes that might emerge from their environment. We then analyzed the productions to better understand which choices they made. We analyzed how their designs can be generalized to non-rectangular displays and proposed a set of hypotheses, later evaluated through a user study.



Figure 5.4: Design probes on non-rectangular interfaces.

The previous design probe was driven by a concrete research question: designers were given a precise task to complete. Another approach when conducting prospective explorations is to give designer the freedom to propose novel usage scenarios and designs based on a general concept. We adopted this approach during a 1-week workshop in January 2018 with design students from the Design Master DTCT (Design Transdisciplinaire, Culture et Territoires) of the University Toulouse Jean Jaures. The workshop, entitled "#Stranger-Screens", was preceded by a one-day seminar, where researchers and practitioners from different fields (HCI, design, hardware) presented their work, as an inspiration for the design students. Then, the students were given three days to generate ideas and craft one design proposal (Figure 5.5). These designs were presented the last day of the week, opening different perspectives on the usage scenarios, contexts and shapes of non-rectangular displays. The quality, diversity and depth of the proposed scenarios were beyond what

usual brainstormings would produce, showing the value of carrying design probes with designers.

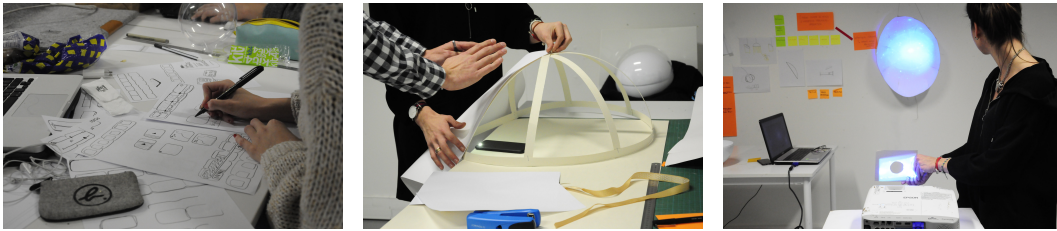


Figure 5.5: Design workshop on non-rectangular interfaces.

## 5.2 Prototyping methods

Prototyping ubiquitous interaction often requires combining novel interaction modalities, which can be difficult in a rapid iterative design process. Rapidly designing and creating new devices is also complex, in particular due to the lack of proper design and ideation methods. Finally, research on ubiquitous interaction often requires to experimentally validate interactions which can not be easily implemented at the current time (such as on-body gestures for instance). To solve these three issues, in this section we present our work on rapid prototyping tools and the methods we used to develop proof-of-concept pervasive prototypes.

### 5.2.1 Rapid prototyping tools (\*contribution)

During my PhD, we were specifically interested in multimodal interactive systems that include several interaction modalities used by the user (input modalities) in an independent or combined way. My doctoral research was dedicated to conceptual and software tools for efficiently exploring the huge set of possibilities during the design phase of input multimodal interfaces [Serrano, 2010]. We presented a component-based approach for designing and prototyping functional and/or simulated input multimodal interfaces. Our approach relied on the data-flow, from input devices to the user-tasks, depicted by an assembly of components.

Our work was both conceptual and practical. Our conceptual model consisted of a characterization space for describing components. This space helps interaction designers to define component assemblies and to consequently consider different design alternatives for input multimodality. Our software tool is a component-based prototyping tool that implements the characterization space (i.e., our conceptual model). We implemented several multimodal prototypes using this tool, involving a large variety of interaction modalities including 2D gestures, 3D pointing techniques, speech commands, head and body movements, geolocation techniques as well as tangible techniques.

### 5.2.2 Proof-of-concept prototypes

A proof-of-concept (PoC) is the realization of a certain method or idea to demonstrate its feasibility. In HCI, PoCs are used to fill an interactive technology gap, with the idea that if the PoC validates the concept (for instance in terms of usability), future work will ultimately fill the gap. In this section we describe how we implemented and applied proof-of-concept prototypes to evaluate ubiquitous interactions that could not be easily implemented otherwise.

### On-body input

We explored hand-to-face gestures for HWDs without emphasizing the technology that would ultimately support this style of input. Several experimental options exist to implement this on-body input, such as a camera mounted on the HWD, body-implanted sensors or instruments worn on the finger.

To evaluate our interaction (see Chapter 2.3), we implemented a proof-of-concept prototype using a Vicon T20 infrared optical tracking system with six cameras positioned around the user (front, front-right and right side at different heights). We placed infrared markers on the participant's index finger (Figure 5.6). To detect skin contact, we used a proximity sensor connected to a micro-controller. The sensor set was connected through a USB cable to a desktop computer. To make the sensor set unobtrusive, we integrated it into a glove worn by the user. During our studies we had a negligible number of tracking errors (0.15% of all trials). The system had no perceivable latency: all input was merged to the same program and sent to the HWD through USB. Optical tracking ran at 690 Hz; contact sensor at 600 Hz; merged data were sent to the HWD at a measured rate of 142 Hz. While this solution would obviously be impractical for real use in mobile context, our solution was sufficient for carrying in-the-lab studies and demonstrating the utility of on-body interaction.

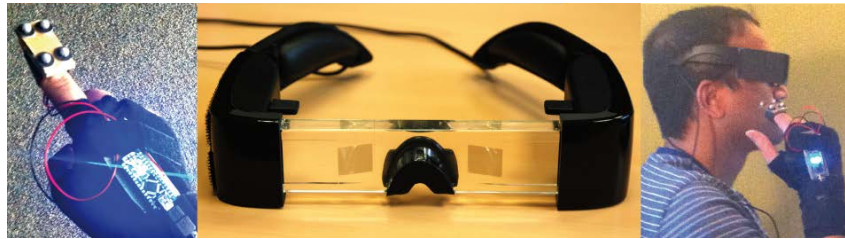


Figure 5.6: Experimental implementation of hand-to-face gestures: IR markers and micro-controller on hand; Epson Moverio HWD and subject wearing HWD.

### Finger identification

Detecting fingers individually is a difficult task, and no touch technology efficiently supports finger identification yet. During our work on how visually impaired (VI) people explore raised-line diagrams [Bardot *et al.*, 2017], we needed to accurately track hands and fingers movements, since previous studies showed that VI people use both hands and different fingers. Our approach needed to be unobtrusive, since we did not want to limit or influence how participants used their hands and fingers. We adopted a colour tracking camera-based approach: we used a Logitech C270 webcam (1280x720 px) located above the touch screen in order to track the ten fingers according to coloured markers placed on each nail (Figure 5.7). The acquisition rates were 50 Hz for the camera and 100 Hz for the touchscreen. While this approach has some limitations, such as the number of possible colors that can be used without reducing the tracking accuracy, it proved to be sufficient for our study in a light-controlled environment.

### Holographic displays

Holographic or true-3D refers to any 3D digital display capable of producing mid-air, full-depth-cue (or volumetric), multi-angle and/or multi-user images without the need for user instrumentation. During our exploration of true-3D interaction around a mobile device [Serrano *et al.*, 2014b], we needed a way to emulate such a display for a single user, which

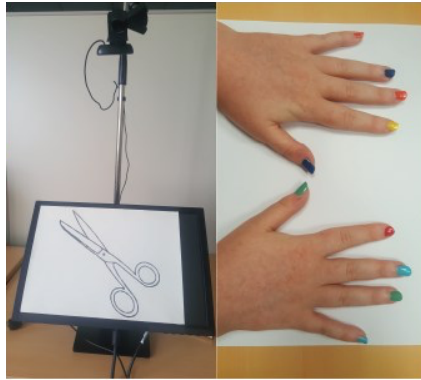


Figure 5.7: Experimental setup: a raised-line drawing is placed over the touchscreen. A camera located above the drawing tracks fingers movements.

could be instrumented for our study purposes. Our implementation was based on a stereoscopic display coupled with head tracking, on the VisCube platform, an immersive environment composed of a projection wall and floor (Figure 5.8). In this system the user has to wear polarized glasses with IR markers to allow visual head-tracking. 3D content was developed using GLUT. The position of the mobile device in the environment was tracked using a Vicon IR motion tracking system and IR markers.

The results of our experiments are influenced by the technology we used to emulate the true-3D displays. The obvious differences between this technology and the final true-3D display, in terms of color, brightness or 3D perception, may alter the results from our experiments. However, most of those results are strongly influenced by human physiological limitations on wrist-based rotation and arm reach. Thus we think these technical differences do not have a fundamental impact on our findings. Moreover, researchers have used such platforms for developing and testing novel technologies.

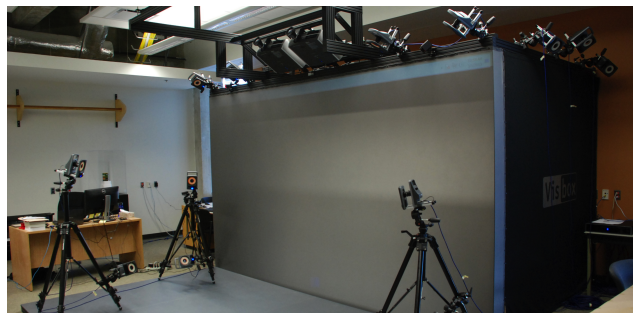


Figure 5.8: We built our true-3D prototype in a VisCube 3D immersive system, tracking the mobile device using optical tracking.

### 5D Mouse prototype

Building a working version of our spherical 5D mouse, the Roly-Poly Mouse (RPM) [Perelman *et al.*, 2015], presented several technical challenges to properly track its position, orientation, and user input (i.e. a selection mechanism). To track RPM **position** several embedded and non-embedded solutions are possible. Embedded solutions include using magnetic sensors, although these are not precise enough to allow mouse-like pointing.



Not-embedded solutions include using IR cameras (similar to the Vicon system used for on-body input) or an underlying sensitive surface, similar to Wacom's tablets. All of these solutions would allow detecting the z-dimension at a certain level. To track the **orientation** of the device, we could use an embedded inertial measurement unit (IMU).

There are two main options to integrate an always-available **selection mechanism** invariant to rotation: an all-around ring button and a capacitive surface. The ring button, situated all around the device, permits multi-touch input with several fingers. A more elegant solution would be to use a multi-touch surface. However, this solution would lack of haptic feedback, like on a regular button. This solution has proved to be useful for extending a traditional mouse such as Apple's Magic Mouse in which it is combined with a mechanical switch. The main challenge with this solution would be to distinguish a finger touch from a palm contact or with a grasping/squeezing gesture.

For our initial exploration of the Roly-Poly Mouse (RPM) gestures [Perelman *et al.*, 2015], we used infrared optical markers tracked by 12 OptiTrack cameras (1mm precision) to track the translation, rotation and roll of the device. The system senses the position ( $x, y, z$ ) and orientation (yaw, pitch, roll) of RPM at 100Hz. We placed IR markers on the RPM to allow the cameras detect the device without impeding the user's ability to grab the device with different hand postures. Informal tests had also confirmed that the marker did not limit the amplitude of comfortable rolls: the maximum possible roll of RPM given these physical markers was  $70^\circ$  in the marker support directions. Our tracking setup did not register contact with the underneath surface, thus we could not use clutching in our experiments. To demonstrate the feasibility of such a device, we created a working an integrated version using a Polhemus Patriot Wireless tracker, which is however less precise and robust than the infrared solution.

In our next evolution of RPM, called TDome [Saidi *et al.*, 2017], the device holds an x-IMU of x-io Technologies to detect the Roll and Rotation of the device in 3D. The IMU is composed of a triple-axis gyroscope, accelerometer and magnetometer. The refresh rate of the sensors goes up to 512Hz and we used Bluetooth to connect the IMU with the computer. The IMU offered an angular precision of  $1^\circ$ . We 3D printed a holder to fit the IMU in a horizontal position inside the TDome. To detect the displacement of the device, we used an infrared bezel that generated TUIO events. We implemented a filtering process to discard touch events that were detected when fingers touched the surface around the device.

In our latest work, which explored the use of RPM to facilitate selecting toolbar items [Dubois *et al.*, 2018], we enhanced the original RPM. As the original one, the version consists of a sphere with a diameter of 8 cm, which includes a Bluetooth enabled Inertial Measurement Unit (the same xIMU by xIO Tech). In comparison to the original RPM, our version was placed on a Wacom Intuos 3D tablet (216x135 mm, resolution: 2540 lpi). As the tablet is multitouch, it can detect the translation of RPM and finger taps. We therefore covered the RPM surface with a graphite lacquer to give the device a conductive coating. We tried to insert various forms of button on top of RPM. However our pre-tests showed that using a physical button on RPM altered the device handling gesture and brought a number of technical issues (button position, etc.). Instead, we considered the use of a tactile surface underneath RPM to detect a user's finger tap: the user can employ any finger of the same hand that manipulates RPM to tap on the surface, although participants seemed to prefer the thumb. An algorithm associates the first touch on the tablet to the RPM position, and triggers a tap event only when detecting a second touch. Alternatively, the user can press a key on the keyboard with the non-dominant hand: as the user's main task is probably involving keyboard input, this bimanual setting offers a fluid interaction compatible with regular keyboard input (the keyboard is only used as a validation).

In conclusion, we implemented various versions of the Roly-Poly Mouse (Figure 5.9), which illustrates the interest of using in-the-lab prototypes to evaluate interactions. Each iteration demonstrated the benefits of the device for a different task or context, which motivated us to pursue our research agenda on this device and to improve its implementation.

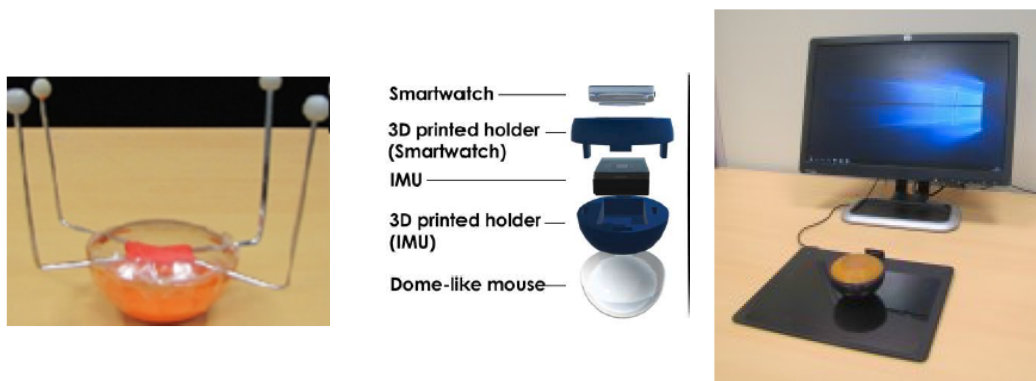


Figure 5.9: From left to right: first version of RPM using optical tracking; second version integrating an IMU to track rotation and orientation; last version additionally used a Wacom tablet to track its position.

### 5.3 Evaluation methods

In this section we sum up two evaluation approaches that we carried. The first one concerns the evaluation of interaction modalities that are still not implemented using a component-based wizard-of-oz approach. The second consists on a type of evaluation that we believe has not yet got full attention from the HCI community, i.e. subjective feedback gathering through pairwise comparisons.

#### 5.3.1 Component-based Wizard-of-Oz (\*contribution)

To accelerate the prototype-test cycle at the early stage of the design, one solution is to adopt a Wizard-of-Oz approach. As explained in [Davis *et al.*, 2007], "Wizard of Oz prototype is an incomplete system that a designer can simulate behind curtain while observing the reactions of real end users". WoZ studies have been shown to allow fast prototyping of interactive applications by allowing the evaluator (wizard) to simulate missing functions.

Following our work on the OpenInterface framework, while the rapid development of multimodal prototypes is possible, testing such prototypes raised further technical and experimental issues. To overcome these limitations, we proposed OpenWizard [Serrano and Nigay, 2010], a component-based approach for the rapid prototyping and testing of input multimodal interaction. OpenWizard is based on a simple idea: the designer and the developer should be able to work both on the multimodal prototype and on the WoZ experiment at the same time.

We illustrate this approach using the puzzle metaphor (Figure 5.10 - left). A multimodal interaction prototype can be seen as a puzzle, where each piece corresponds to a physical device, a transformation algorithm, a fusion algorithm or a task. The designer and the developer assemble those puzzle pieces in order to create the complete puzzle, their multimodal application prototype, that can then be evaluated. The aim is to avoid losing time and effort in waiting for missing pieces and in creating and implementing pieces that will then be discarded during the evaluation.

If we translate this idea into our component-based approach (Figure 5.10 - right), we simply have to replace puzzle pieces by software components. The designer and the developer can then build multimodal prototypes easily by assembling functional components (devices, transformation, composition, task) with WoZ (non-functional) components. To do so, we define WoZ components as parts of the component-based approach for rapidly developing multimodal prototypes. WoZ components are characterized according to the

roles that a WoZ component can play in the data-flow of input multimodal interaction, from devices to tasks.

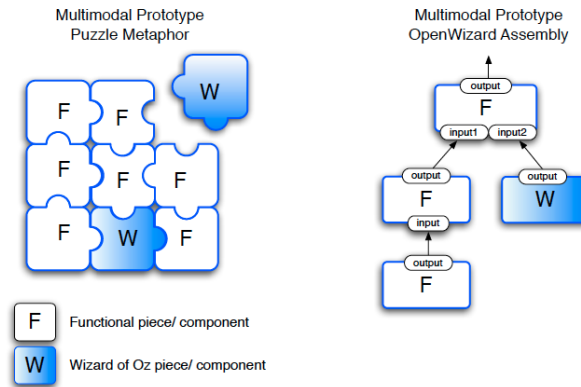


Figure 5.10: Definition of a multimodal prototype using the component-based WoZ approach. Left: concept illustration through the puzzle metaphor. Right: its implementation on our component-based approach.

### 5.3.2 Pairwise comparisons

Quantitative controlled experiments are assumed to be the best tool to demonstrate UI efficiency. But they are only useful when it is possible to test counterbalanced variables without introducing confounds. In some situations, it is not possible to find an experimental setup following this rule. For instance, in readability studies, the issue is that text presentations affect readability. Thus any effects observed could also be a result of the text presentations that change according to the conditions (confound variables).

A way to get participants' input is to use subjective judgement. Estimating preferences based on subjective judgements is a critical step in psychological experiments with applications in many fields such as marketing, environmental sciences and health economics. In particular pairwise experiments have been widely used. In such studies, two conditions are presented to participants who then indicate one alternative over the other. Pairwise comparison ratings have been proven to produce more realistic results than asking for individual rankings (e.g. using a Likert scale).

During our exploration of visual layouts on non-rectangular interfaces [Serrano *et al.*, 2017], we asked 57 participants to compare pairs of layout visualizations and say which one was nicer (i.e. visually pleasing), clearer (i.e. not confusing) and more symmetric. Participants could give three answers for each question: Visualization-1, Visualization-2 or Both. Our analysis of the results consisted in three steps:

**Step 1: Individual consistency checking.** We computed the Transitivity Satisfaction Rate (TSR), which quantifies the consistency of a participant's judgments over multiple questions. E.g. if A is found more restrictive than B, and B more than C, then we should have A more restrictive than C. We removed 4 participants whose TSR was below 0.8. The mean TSR for all other users was 0.92 (SD = 0.05) and at least over 0.8 for all of them, thus denoting that they paid full attention to the study.

**Step 2: Overall consistency checking.** To test the overall consistency across participants we checked the stochastic transitivity properties or computed Kendall's-coefficient. For each participant, we computed a list of rankings of visualisations and used the kendall-tau Python library to produce a coefficient for each pair of participants, computed as a per-



centage (100% means all participants perfectly agreed, 0% they perfectly disagreed). Our results show that the mean Kendall's-coefficient is above 50% for all conditions.

**Step 3: Model the data.** The individual and overall consistencies were confirmed, so we proceeded to model the data. We used the Bradley-Terry-Luce model, which associates an ability metric to each condition that have been paired-compared as well as the p-value for each pair comparisons. Note that the Bradley-Terry-Luce model computes a p-value that express how the visualizations compare to one specific visualization only, which serves as reference and is a parameter of the formula. We thus performed several tests to compute the significant level for each comparison. To counteract the problem of making multiple comparisons tests we used a Bonferroni correction for each result.

To report the results, we represented the metric of each visualization computed via the Bradley-Terry-Luce model. The metric gives a value between 0 and 1, where the lowest condition equals 0 and the highest 1. We also indicated the standard error values given by the model.

In conclusion, using a pairwise comparison study allowed us to evaluate the subjective value of layout visualizations, which would have been very difficult otherwise. This shows the importance of looking for evaluation methods from other fields (here in psychological experiments). The main challenges in this case are to find the appropriate methods, apply them properly and explain to HCI practitioners and reviewers the value and validity of such methods.

### 5.3.3 Conclusion

Our work on designing novel interactions for ubiquitous computing led to some advances on design, prototyping or evaluation methods. Each project implied not only thinking about our contribution to HCI, but also about our research procedures. Research on HCI is a relatively young domain, at the intersection of different fields which have well known methods (such as AB experiment design on psychology). As HCI matures, some of our methods will become established and hopefully applied more methodically, improving the overall quality of our research.

# Chapter 6

## Perspectives

### Chapter content

---

6.1	From digital to physical ubiquitous displays . . . . .	65
6.2	Immersive data analytics to explore spatio-temporal data . . . . .	67
6.3	Pervasive freeform interfaces . . . . .	68

---

Our work has explored different types of ubiquitous interfaces (mobile, multi-display and freeform) with the general goal to facilitate the exploration and manipulation of complex data (large number of commands, spatial, 3D and volumetric data). As these types of interfaces slowly populate our surroundings, they reshape our city environments. While the term "smart city" englobes a variety of concepts, technology is one of its core factors. According to [Chourabi *et al.*, 2012], "smart computing refers to a new generation of integrated technologies that provide IT systems with real-time awareness of the real world and advanced analytics to help people make more intelligent decisions about alternatives and actions".

As these cities get richer in terms of collected data, it becomes crucial to better interlace interactive data exploration in our everyday spatial environment. This general idea drives our three perspectives. To this end, we propose to investigate three complementary approaches: tying data to physical objects (1), facilitating immersive data analytics (2) and supporting the rise of pervasive freeform interfaces (3).

### 6.1 From digital to physical ubiquitous displays

In our work we have investigated the exploration of virtual content, such as spatial data, 3D environments or very large graphs. The next evolution in ubiquitous data analytics could consist in taking advantage of data physicalization. Data physicalization is an emerging area of research [Jansen *et al.*, 2015], which can have obvious advantages for visually impaired users, but could also leverage embedded data analysis, i.e. accessing data in the spatial environment where it belongs or where it was collected.

#### Physical interactive maps for visually impaired users

Providing a visually impaired user with a way to independently explore or construct physical representations of visual graphics could be invaluable. Tangible interfaces are, in this sense, particularly relevant, as they provide a way to interact with digital information via the manipulation of phicons (i.e. physical icons [Ishii and Ullmer, 1997]), and therefore also provide a way to translate digital information into a physical form.

To this end, previous work has mostly focused on rendering punctual symbols tangible (i.e. physical and associated to a digital content). For instance, [McGookin *et al.*, 2010] developed a tangible interface for the exploration of line graphs and bar charts by visually impaired people: phicons are placed in a physical grid to represent the top of a bar or the turning point of a linear function. Authors observed that the objects were regularly knocked over during the exploration and hence provided a few recommendations concerning the design of phicons for visually impaired users, which need to be stable.

Such an approach is limited for the construction of most graphics that include lines and not only points. As a first step to overcome this limitation, our group at the University of Toulouse designed and constructed a novel type of physical icon called Tangible Reels, presented at CHI 2016 [Ducasse *et al.*, 2016]. A Tangible Reel is composed of a sucker pad that ensures stability, and of a retractable reel that is used to create physical links between phicons. We also developed a non-visual tabletop interface that enables a visually impaired user to build a tangible map using Tangible Reels and to interact with it.

This contribution is a first step and still has limitations: it only allows to render graphics composed of a limited number of points and lines. There is a need to design an interface that is accessible to visually impaired users without any assistance, and that allows them to render tangible any digital graphical content that is composed of points, lines, surfaces and volumes, as well as dynamically explore and annotate them. As fabrication methods (3D printing, laser cutting, etc.) become more affordable and easier to put into service, they could represent an interesting perspective to fill this gap. The main challenge then is how to make these fabricated graphics interactive and accessible for the visually impaired. Two major approaches seem to emerge: either integrating the interactive mechanisms in the physical object itself, or have users wear the interaction devices, as in our previous work with smartwatches.

### Public interactive physical models

Besides the interest for visually impaired people, embedding data into physical objects and into our surroundings is becoming more important as we move toward an interlaced reliance on data for our daily activities [Piper *et al.*, 2002] [Raskar *et al.*, 2004]. Such embeddings allow for in-place or in-situ analytics [Elmqvist and Irani, 2013], but also constitute effective communication tools when presenting information to non-expert users [Jansen *et al.*, 2013]. While technical capabilities, including spatial augmented reality [Bimber and Raskar, 2005] or on-object interactions [Zhang *et al.*, 2017] are maturing, there exists little discussion on the granularity levels for interacting with such datasets in public places.

We can illustrate this need for embedded data interaction with a specific use-case: architectural models. Digital architectural models are growing in complexity as architects use these for recording detailed building specifications (material types, 3D representations, structural monitoring sensors), all of which are contained in standardized documents, such as the Building Information Model (BIM). Such large volumes of data often require specialized expertise to comprehend and are inaccessible to non-experts.

An alternative for exploring this data is to embed a selected subset of relevant information only, into their physical counterpart in which unnecessary details are hidden. It thus involves a degraded version of the data and digital model enclosed into a traditional BIM approach. Such physical models are typically on display to showcase a newly constructed building or at a museum as a cultural exhibit. Within this use-case, there is a need to propose and explore embedded data interaction with such architectural models.

There exist numerous challenges to embed interactive data into physical models. First, constructing the physical model needs to be quick to support flexibility and allow for a rapid update on the dataset; second, linking the physical model to its associated digital information must be easy to allow a range of end-users to use such a system; third, interacting with the embedded information needs to account for the environmental constraints

(for example, models that are encased in a vitrine should still be interactive); finally, the presentation of the information should be easily visible and should clearly link the data to the physical model. Our approach to tackle this challenge will be first to focus on the tools for the design and prototyping of interactive physical models; then we will explore interaction techniques to access this physical data in a public setting, with a specific focus on tactile glass cases. This work will be carried by a PhD starting in 2018, funded by the neOCampus project and the Occitanie region.

## 6.2 Immersive data analytics to explore spatio-temporal data

A complementary approach to the physicalization of data consists in exploring virtual data in an augmented reality environment, a paradigm called "immersive data analytics" [Chandler *et al.*, 2015]. Recent technological advancements in augmented reality devices open up new opportunities for users to interact in their environments. Systems like the HoloLens, MetaVision or Moverio allow the user to display numerical data and visualisations directly on the physical world by attaching them to a fixed physical anchor. These technologies offer the user interaction opportunities that are to this day insufficiently explored. We do not have an implicit design rules to guide the developer towards solutions for these environments. This results in a compilation of partially satisfactory solutions for interaction.

There is a need to provide users of augmented reality environments with rich solutions to explore complex data. Previous solutions for mixed reality mostly considered a limited number of degrees of freedom. For instance, in our work on spatial data exploration using hand-to-face input, we only used 2 degrees of freedom to pan or zoom. This approach no longer stands when exploring complex data such as spatio-temporal datasets. Spatio-temporal datasets are generated daily, whether to collect disease movements, migration patterns, or other movement data. Their visualization usually includes animated representations of space-time cubes [Bach *et al.*, 2014]. When facing such rich and complex data environments, the exploration and interpretation of such data is crucial for acting on the systems, such as, for example, to limit the spread of a disease or to study traffic patterns. For such three-dimensional datasets, finding a coherent and efficient interaction to perform a set of operations to better explore these data is a major challenge.

Solutions explored so far include the mouse, tangible user interfaces and the mid-air interaction [Cordeil *et al.*, 2017]. Using a 2D device, such as the mouse, for a 3D task greatly limits the interaction. Although tangible interfaces add physicality, a lack of feedback raises some issues in terms of the number of physical bricks. Finally, the mid-air interaction supports direct data manipulation, but may be limiting in terms of gesture detection and 3D perception (related to depth detection). Our goal to maintain the freedom of movement of mid-air interactions, the degrees of freedom of tangible interactions and the accuracy of the mouse to provide a flexible and precise solution for interaction with immersive visualizations.

A perspective to solve this challenge is to use our multi-DoF mouse (Roly-Poly Mouse) in this new interaction context. We will study the concept of on-body tangible interactions using the user as a physical support for interaction to give him the freedom of movement necessary to explore immersive visualization anchored in the real world. This solution will be based on the combination of 1) the use of a multi-DOF mouse-type wireless device, combining the precision inherent to a mouse, the flexibility of tangibles and the large control capabilities of multi-degrees of freedom mice and 2) the use of the body to guide the physical manipulations of the device by exploiting the proprioception of the user (the body's natural capacity to sense its own body parts and perceive their movements and localizations) while limiting muscle fatigue inherent to mid-air interactions.

### 6.3 Pervasive freeform interfaces

Our work on freeform interfaces started to address some of the challenges of these types of pervasive interfaces, in particular in terms of displaying text and laying out content. However, our initial work did not look into the cognitive aspects of looking for information on non-rectangular displays and only considered a limited set of simple shapes (such as circles and triangles). Hence our perspectives include studying these cognitive aspects and generalizing our approach to design content for freeform interfaces using computer graphics methods.

#### Cognitive aspects of non-rectangular interfaces

Although our previous work showed essential differences between rectangular and non-rectangular interfaces when reading text, these prior studies have not examined visual search strategies employed by users. Revealing these patterns is key in identifying how best to place and structure content [Buscher *et al.*, 2009] on non-rectangular displays. Designers could use such information to place relevant information at strategic locations for rapid access, key knowledge for identifying where to place menus or attention-grabbing banners on websites [Buscher *et al.*, 2009].

Our future work will fill this gap by investigating gaze patterns when visually searching information on non-rectangular interfaces. We will build on previous work and explore this question using the visual layout and display shapes investigated in our previous studies. Since gaze tracking has been widely used to understand how users search information on traditional screens, we will use it to log the gaze patterns and highlight particular strategies on non-rectangular interfaces.

#### Mapping content and input interaction on freeform interfaces

We will develop an interaction paradigm for non-rectangular user interfaces with a focus on two main objectives: exploring alternate mappings of content onto non-rectangular surfaces (1) and proposing novel input interaction techniques for non-rectangular displays (2).

Mapping content onto non-rectangular surfaces poses organisation, adaptation and deformation challenges. Organisation consists in geometrically arranging items on the shape to fit the available space (for instance, laying out icons). Adaptation consists in defining how the content changes when it is displaced (for instance, when text is scrolled). Deformation occurs when the display surface shape changes over time (for instance when an occluding physical object, such as a cup, is displaced on the surface). We will propose and study different mappings of content onto freeform displays to organize, adapt and deform content. The first step will be to geometrically define the coordinate system of the non-rectangular surface: for this, we will use computer graphics methods. We propose to rely on Computer Aided Design trimming technics in which free-form parametric shapes are defined over a support surface (in our case the full projection). Distortions will be controlled with local parametrizations and implicit shape representation may be considered for eventual in/out display shape tests. We will then first focus on simple content, such as text, menus, windows and/or GUI items among others. We will finally move to more complex visual content, such as time series or geospatial data visualisation.

We will explore two approaches to interact with such displays, either direct touch on the surface itself or indirect input on a distant touch surface (e.g. on a handheld touchscreen, a common approach to interact with public displays). For both input types, our objective is to provide pointing and content manipulation techniques (such as scrolling or 2D panning). For indirect input, we will also design and evaluate different mappings between the rectangular input coordinates (touchscreen) and the non-rectangular projection.

# Bibliography

- [Ashbrook, 2010] Daniel L Ashbrook. *Enabling mobile microinteractions*. Georgia Institute of Technology, 2010.
- [Bach *et al.*, 2014] Benjamin Bach, Pierre Dragicevic, Daniel Archambault, Christophe Hurter, and Sheelagh Carpendale. *A Review of Temporal Data Visualizations Based on Space-Time Cube Operations*, pages 23–41. Eurographics Conference on Visualization (EuroVis 2014), 2014.
- [Bailly *et al.*, 2012] Gilles Bailly, Jörg Müller, Michael Rohs, Daniel Wigdor, and Sven Kratz. Shoesense: A new perspective on gestural interaction and wearable applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12*, pages 1239–1248, New York, NY, USA, 2012. ACM.
- [Balakrishnan and Hinckley, 2000] Ravin Balakrishnan and Ken Hinckley. Symmetric bi-manual interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '00*, pages 33–40, New York, NY, USA, 2000. ACM.
- [Bardot *et al.*, 2016] Sandra Bardot, Marcos Serrano, and Christophe Jouffrais. From tactile to virtual: Using a smartwatch to improve spatial map exploration for visually impaired users. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '16*, pages 100–111, New York, NY, USA, 2016. ACM.
- [Bardot *et al.*, 2017] Sandra Bardot, Marcos Serrano, Bernard Oriola, and Christophe Jouffrais. Identifying how visually impaired people explore raised-line diagrams to improve the design of touch interfaces. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17*, pages 550–555, New York, NY, USA, 2017. ACM.
- [Beaudouin-Lafon, 2004] Michel Beaudouin-Lafon. Designing interaction, not interfaces. In *AVI '04: Proceedings of the working conference on Advanced visual interfaces*, pages 15–22, New York, NY, USA, 2004. ACM.
- [Bederson, 2004] Benjamin B. Bederson. Interfaces for staying in the flow. *Ubiquity*, 2004(September):1–1, September 2004.
- [Benko and Feiner, 2005] Hrvoje Benko and Steven Feiner. Multi-monitor mouse. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems, CHI EA '05*, pages 1208–1211, New York, NY, USA, 2005. ACM.
- [Bergé *et al.*, 2014] Louis-Pierre Bergé, Marcos Serrano, Gary Perelman, and Emmanuel Dubois. Exploring smartphone-based interaction with overview+detail interfaces on 3d public displays. In *Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices & Services, MobileHCI '14*, pages 125–134, New York, NY, USA, 2014. ACM.
- [Bimber and Raskar, 2005] Oliver Bimber and Ramesh Raskar. *Spatial Augmented Reality: Merging Real and Virtual Worlds*. A. K. Peters, Ltd., Natick, MA, USA, 2005.

- [Boring *et al.*, 2009] Sebastian Boring, Marko Jurmu, and Andreas Butz. Scroll, tilt or move it: Using mobile phones to continuously control pointers on large public displays. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7, OZCHI '09*, pages 161–168, New York, NY, USA, 2009. ACM.
- [Boring *et al.*, 2010] Sebastian Boring, Dominikus Baur, Andreas Butz, Sean Gustafson, and Patrick Baudisch. Touch projector: Mobile interaction through video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10*, pages 2287–2296, New York, NY, USA, 2010. ACM.
- [Boring, 2011] Sebastian Boring. Interacting in multi-display environments. *DC Ubicomp*, 05 2011.
- [Bortolaso *et al.*, 2013] Christophe Bortolaso, Matthew Oskamp, T.C. Nicholas Graham, and Doug Brown. Ormis: A tabletop interface for simulation-based training. In *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces, ITS '13*, pages 145–154, New York, NY, USA, 2013. ACM.
- [Bragdon *et al.*, 2011] Andrew Bragdon, Eugene Nelson, Yang Li, and Ken Hinckley. Experimental analysis of touch-screen gesture designs in mobile environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11*, pages 403–412, New York, NY, USA, 2011. ACM.
- [Buscher *et al.*, 2009] Georg Buscher, Edward Cutrell, and Meredith Ringel Morris. What do you see when you're surfing?: Using eye tracking to predict salient regions of web pages. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '09*, pages 21–30, New York, NY, USA, 2009. ACM.
- [Buxton, 2008] Bill Buxton. The long nose of innovation. *BusinessWeek*, January 2008.
- [Cauchard *et al.*, 2011] Jessica R. Cauchard, Markus Löchtefeld, Pourang Irani, Johannes Schoening, Antonio Krüger, Mike Fraser, and Sriram Subramanian. Visual separation in mobile multi-display environments. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, UIST '11*, pages 451–460, New York, NY, USA, 2011. ACM.
- [Cauchard *et al.*, 2012] Jessica R. Cauchard, Mike Fraser, Teng Han, and Sriram Subramanian. Steerable projection: Exploring alignment in interactive mobile displays. *Personal Ubiquitous Comput.*, 16(1):27–37, January 2012.
- [Chandler *et al.*, 2015] T. Chandler, M. Cordeil, T. Czauderna, T. Dwyer, J. Glowacki, C. Goncu, M. Klapperstueck, K. Klein, K. Marriott, F. Schreiber, and E. Wilson. Immersive analytics. In *2015 Big Data Visual Analytics (BDVA)*, pages 1–8, Sept 2015.
- [Chokshi *et al.*, 2014] Apoorve Chokshi, Teddy Seyed, Francisco Marinho Rodrigues, and Frank Maurer. eplan multi-surface: A multi-surface environment for emergency response planning exercises. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces, ITS '14*, pages 219–228, New York, NY, USA, 2014. ACM.
- [Chourabi *et al.*, 2012] Hamed Chourabi, Taewoo Nam, Shawn Walker, J. Ramon Garcia, Sehl Mellouli, Karine Nahon, Theresa Pardo, and Hans Scholl. Understanding smart cities: An integrative framework. *45th Hawaii International Conference on System Sciences*, pages 2289–2297, 01 2012.
- [Cockburn *et al.*, 2009] Andy Cockburn, Amy Karlson, and Benjamin B. Bederson. A review of overview+detail, zooming, and focus+context interfaces. *ACM Comput. Surv.*, 41(1):2:1–2:31, January 2009.

- [Cordeil *et al.*, 2017] Maxime Cordeil, Benjamin Bach, Yongchao Li, Elliott Wilson, and Tim Dwyer. *A Design Space for Spatio-Data Coordination: Tangible Interaction Devices for Immersive Information Visualisation*, pages 46–50. IEEE Computer Society Press, 9 2017.
- [Cotting and Gross, 2006] Daniel Cotting and Markus Gross. Interactive environment-aware display bubbles. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology*, UIST '06, pages 245–254, New York, NY, USA, 2006. ACM.
- [Davis *et al.*, 2007] Richard C. Davis, T. Scott Saponas, Michael Shilman, and James A. Landay. Sketchwizard: Wizard of oz prototyping of pen-based user interfaces. In *Proceedings of UIST'07*, pages 119–128. ACM Press, 2007.
- [Dix *et al.*, 2003] Alan Dix, Janet E. Finlay, Gregory D. Abowd, and Russell Beale. *Human-Computer Interaction (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2003.
- [Dubois *et al.*, 2018] Emmanuel Dubois, Marcos Serrano, and Mathieu Raynal. Rolling-menu: Rapid command selection in toolbars using roll gestures with a multi-dof mouse. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 367:1–367:12, New York, NY, USA, 2018. ACM.
- [Ducasse *et al.*, 2016] Julie Ducasse, Marc J-M Macé, Marcos Serrano, and Christophe Jouffrais. Tangible reels: Construction and exploration of tangible maps by visually impaired users. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 2186–2197, New York, NY, USA, 2016. ACM.
- [Elmqvist and Irani, 2013] Niklas Elmqvist and Pourang Irani. Ubiquitous analytics: Interacting with big data anywhere, anytime. *Computer*, 46(4):86–89, April 2013.
- [Elmqvist *et al.*, 2008] Niklas Elmqvist, Nathalie Henry, Yann Riche, and Jean-Daniel Fekete. Melange: Space folding for multi-focus interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1333–1342, New York, NY, USA, 2008. ACM.
- [Galitz, 2007] Wilbert O. Galitz. *The Essential Guide to User Interface Design: An Introduction to GUI Design Principles and Techniques*. John Wiley & Sons, Inc., New York, NY, USA, 2007.
- [Geyer *et al.*, 2012] Florian Geyer, Anita Höchtl, and Harald Reiterer. Harnessing the benefits of bimanual and multi-finger input for supporting grouping tasks on interactive tabletops. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, NordiCHI '12, pages 496–499, New York, NY, USA, 2012. ACM.
- [Gotsch *et al.*, 2016] Daniel Gotsch, Xujing Zhang, Jesse Burstyn, and Roel Vertegaal. Holoflex: A flexible holographic smartphone with bend input. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 3675–3678, New York, NY, USA, 2016. ACM.
- [Granell and Leiva, 2017] Emilio Granell and Luis A. Leiva. btap: Back-of-device tap input with built-in sensors. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '17, pages 52:1–52:6, New York, NY, USA, 2017. ACM.
- [Hachet *et al.*, 2009] Martin Hachet, Fabrice Declé, Sebastian Knödel, and Pascal Guitton. Navidget for 3d interaction: Camera positioning and further uses. *Int. J. Hum.-Comput. Stud.*, 67(3):225–236, March 2009.
- [Han Kim *et al.*, 2010] K Han Kim, Matthew Reed, and Bernard J Martin. A model of head movement contribution for gaze transitions. *Ergonomics*, 53:447–57, 04 2010.



- [Harrison *et al.*, 2010] Chris Harrison, Desney Tan, and Dan Morris. Skinput: Appropriating the body as an input surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 453–462, New York, NY, USA, 2010. ACM.
- [Hilliges *et al.*, 2012] Otmar Hilliges, David Kim, Shahram Izadi, Malte Weiss, and Andrew Wilson. Holodesk: Direct 3d interactions with a situated see-through display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 2421–2430, New York, NY, USA, 2012. ACM.
- [Hinckley, 2003] Ken Hinckley. The human-computer interaction handbook. In Julie A. Jacko and Andrew Sears, editors, *The Human-computer Interaction Handbook*, chapter Input Technologies and Techniques, pages 151–168. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 2003.
- [Houben and Marquardt, 2015] Steven Houben and Nicolai Marquardt. Watchconnect: A toolkit for prototyping smartwatch-centric cross-device applications. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1247–1256, New York, NY, USA, 2015. ACM.
- [Ishii and Ullmer, 1997] Hiroshi Ishii and Brygg Ullmer. Tangible bits: Towards seamless interfaces between people, bits and atoms. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, pages 234–241, New York, NY, USA, 1997. ACM.
- [Jankowski *et al.*, 2010] Jacek Jankowski, Krystian Samp, Izabela Irzynska, Marek Jozwowski, and Stefan Decker. Integrating text with video and 3d graphics: The effects of text drawing styles on text readability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 1321–1330, New York, NY, USA, 2010. ACM.
- [Jansen *et al.*, 2013] Yvonne Jansen, Pierre Dragicevic, and Jean-Daniel Fekete. Evaluating the efficiency of physical visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 2593–2602, New York, NY, USA, 2013. ACM.
- [Jansen *et al.*, 2015] Yvonne Jansen, Pierre Dragicevic, Petra Isenberg, Jason Alexander, Abhijit Karnik, Johan Kildal, Sriram Subramanian, and Kasper Hornbæk. Opportunities and challenges for data physicalization. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 3227–3236, New York, NY, USA, 2015. ACM.
- [Javed *et al.*, 2012] Waqas Javed, Sohaib Ghani, and Niklas Elmqvist. Polyzoom: Multiscale and multifocus exploration in 2d visual spaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 287–296, New York, NY, USA, 2012. ACM.
- [Jones *et al.*, 2012] Brett Jones, Rajinder Sodhi, David Forsyth, Brian Bailey, and Giuliano Maciocci. Around device interaction for multiscale navigation. In *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services*, MobileHCI '12, pages 83–92, New York, NY, USA, 2012. ACM.
- [Kaizu *et al.*, 2010] Kazunari Kaizu, Samik Ghosh, Yukiko Matsuoka, Hisao Moriya, Yuki Shimizu-Yoshida, and Hiroaki Kitano. A comprehensive molecular interaction map of the budding yeast cell cycle. *Molecular systems biology*, 6:415, 09 2010.
- [Kikuchi *et al.*, 2017] Takashi Kikuchi, Yuta Sugiura, Katsutoshi Masai, Maki Sugimoto, and Bruce H. Thomas. Eartouch: Turning the ear into an input surface. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '17, pages 27:1–27:6, New York, NY, USA, 2017. ACM.

- [Kimura *et al.*, 2011] Hidei Kimura, Akira Asano, Issei Fujishiro, Ayaka Nakatani, and Hayato Watanabe. True 3d display. In *ACM SIGGRAPH 2011 Emerging Technologies*, SIGGRAPH '11, pages 20:1–20:1, New York, NY, USA, 2011. ACM.
- [Kin *et al.*, 2011] Kenrick Kin, Björn Hartmann, and Maneesh Agrawala. Two-handed marking menus for multitouch devices. *ACM Trans. Comput.-Hum. Interact.*, 18(3):16:1–16:23, August 2011.
- [Kohn, 1999] Kurt W. Kohn. Molecular interaction map of the mammalian cell cycle control and dna repair systems. *Molecular biology of the cell*, 10 8:2703–34, 1999.
- [Kratz *et al.*, 2012] Sven Kratz, Michael Rohs, Dennis Guse, Jörg Müller, Gilles Bailly, and Michael Nischt. PalmSpace: Continuous around-device gestures vs. multitouch for 3d rotation tasks on mobile devices. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '12, pages 181–188, New York, NY, USA, 2012. ACM.
- [Kurtenbach and Buxton, 1991] Gordon Kurtenbach and William Buxton. Issues in combining marking and direct manipulation techniques. In *Proceedings of the 4th Annual ACM Symposium on User Interface Software and Technology*, UIST '91, pages 137–144, New York, NY, USA, 1991. ACM.
- [Lavie and Tractinsky, 2004] Talia Lavie and Noam Tractinsky. Assessing dimensions of perceived visual aesthetics of web sites. *Int. J. Hum.-Comput. Stud.*, 60(3):269–298, March 2004.
- [Lee *et al.*, 2017] Juyoung Lee, Hui-Shyong Yeo, Murtaza Dhuliawala, Jedidiah Akano, Junichi Shimizu, Thad Starner, Aaron Quigley, Woontack Woo, and Kai Kunze. Itchy nose: Discreet gesture interaction using eog sensors in smart eyewear. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers*, ISWC '17, pages 94–97, New York, NY, USA, 2017. ACM.
- [Lucero and Serrano, 2017] Andres Lucero and Marcos Serrano. Towards proxemic mobile collocated interactions. *International Journal of Mobile Human Computer Interaction (IJMHCI)*, 9(4):10, 2017.
- [Lucero *et al.*, 2016] Andrés Lucero, Aaron Quigley, Jun Rekimoto, Anne Roudaut, Martin Porcheron, and Marcos Serrano. Interaction techniques for mobile collocation. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, MobileHCI '16, pages 1117–1120, New York, NY, USA, 2016. ACM.
- [Malacria *et al.*, 2010] Sylvain Malacria, Eric Lecolinet, and Yves Guiard. Clutch-free panning and integrated pan-zoom control on touch-sensitive surfaces: The cyclostar approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 2615–2624, New York, NY, USA, 2010. ACM.
- [McGookin *et al.*, 2010] David McGookin, Euan Robertson, and Stephen Brewster. Clutching at straws: Using tangible interaction to provide non-visual access to graphs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 1715–1724, New York, NY, USA, 2010. ACM.
- [Moscovich and Hughes, 2008] Tomer Moscovich and John F. Hughes. Indirect mappings of multi-touch input using one and two hands. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1275–1284, New York, NY, USA, 2008. ACM.

- [Nancel *et al.*, 2011] Mathieu Nancel, Julie Wagner, Emmanuel Pietriga, Olivier Chapuis, and Wendy Mackay. Mid-air pan-and-zoom on wall-sized displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 177–186, New York, NY, USA, 2011. ACM.
- [Nancel *et al.*, 2013] Mathieu Nancel, Olivier Chapuis, Emmanuel Pietriga, Xing-Dong Yang, Pourang P. Irani, and Michel Beaudouin-Lafon. High-precision pointing on large wall displays using small handheld devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 831–840, New York, NY, USA, 2013. ACM.
- [Nii *et al.*, 2012] Hideaki Nii, Kening Zhu, Hiromi Yoshikawa, Nyan Lin Htat, Roland Aigner, and Ryohei Nakatsu. Fuwa-vision: An auto-stereoscopic floating-image display. In *SIGGRAPH Asia 2012 Emerging Technologies*, SA '12, pages 13:1–13:4, New York, NY, USA, 2012. ACM.
- [Oakley *et al.*, 2015] Ian Oakley, DoYoung Lee, MD. Rasel Islam, and Augusto Esteves. Beats: Tapping gestures for smart watches. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1237–1246, New York, NY, USA, 2015. ACM.
- [Perelman *et al.*, 2015] Gary Perelman, Marcos Serrano, Mathieu Raynal, Celia Picard, Mustapha Derras, and Emmanuel Dubois. The roly-poly mouse: Designing a rolling input device unifying 2d and 3d interaction. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 327–336, New York, NY, USA, 2015. ACM.
- [Perelman *et al.*, 2016] Gary Perelman, Marcos Serrano, Mathieu Raynal, Celia Picard, Mustapha Derras, and Emmanuel Dubois. Deco: A design space for device composition. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, DIS '16, pages 435–446, New York, NY, USA, 2016. ACM.
- [Pinhanez, 2001] Claudio S. Pinhanez. The everywhere displays projector: A device to create ubiquitous graphical interfaces. In *Proceedings of the 3rd International Conference on Ubiquitous Computing*, UbiComp '01, pages 315–331, London, UK, UK, 2001. Springer-Verlag.
- [Piper *et al.*, 2002] Ben Piper, Carlo Ratti, and Hiroshi Ishii. Illuminating clay: A 3-d tangible interface for landscape analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '02, pages 355–362, New York, NY, USA, 2002. ACM.
- [Poupyrev *et al.*, 2006] Ivan Poupyrev, Henry Newton-Dunn, and Olivier Bau. D20: Interaction with multifaceted display devices. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '06, pages 1241–1246, New York, NY, USA, 2006. ACM.
- [Preece *et al.*, 2001] Jenny Preece, Yvonne Rogers, and Helen Sharp. *Beyond Interaction Design: Beyond Human-Computer Interaction*. John Wiley & Sons, Inc., New York, NY, USA, 2001.
- [Raskar *et al.*, 1998] Ramesh Raskar, Greg Welch, Matt Cutts, Adam Lake, Lev Stesin, and Henry Fuchs. The office of the future: A unified approach to image-based modeling and spatially immersive displays. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '98, pages 179–188, New York, NY, USA, 1998. ACM.
- [Raskar *et al.*, 2004] Ramesh Raskar, Paul Beardsley, Jeroen van Baar, Yao Wang, Paul Dietz, Johnny Lee, Darren Leigh, and Thomas Willwacher. Rfig lamps: Interacting with a self-describing world via photosensing wireless tags and projectors. In *ACM SIGGRAPH 2004 Papers*, SIGGRAPH '04, pages 406–415, New York, NY, USA, 2004. ACM.

- [Rico and Brewster, 2010] Julie Rico and Stephen Brewster. Usable gestures for mobile interfaces: Evaluating social acceptability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 887–896, New York, NY, USA, 2010. ACM.
- [Roudaut *et al.*, 2014] Anne Roudaut, Rebecca Reed, Tianbo Hao, and Sriram Subramanian. Changibles: Analyzing and designing shape changing constructive assembly. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 2593–2596, New York, NY, USA, 2014. ACM.
- [Ruiz *et al.*, 2011] Jaime Ruiz, Yang Li, and Edward Lank. User-defined motion gestures for mobile interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 197–206, New York, NY, USA, 2011. ACM.
- [Saidi *et al.*, 2016] Housseem Saidi, Marcos Serrano, and Emmanuel Dubois. Investigating the effects of splitting detailed views in overview+detail interfaces. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '16, pages 180–184, New York, NY, USA, 2016. ACM.
- [Saidi *et al.*, 2017] Housseem Saidi, Marcos Serrano, Pourang Irani, and Emmanuel Dubois. Tdome: A touch-enabled 6dof interactive device for multi-display environments. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 5892–5904, New York, NY, USA, 2017. ACM.
- [Serrano and Nigay, 2010] Marcos Serrano and Laurence Nigay. A wizard of oz component-based approach for rapidly prototyping and testing input multimodal interfaces. *Journal on Multimodal User Interfaces*, 3(3):215–225, Apr 2010.
- [Serrano *et al.*, 2013] Marcos Serrano, Eric Lecolinet, and Yves Guiard. Bezel-tap gestures: Quick activation of commands from sleep mode on tablets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 3027–3036, New York, NY, USA, 2013. ACM.
- [Serrano *et al.*, 2014a] Marcos Serrano, Barrett M. Ens, and Pourang P. Irani. Exploring the use of hand-to-face input for interacting with head-worn displays. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 3181–3190, New York, NY, USA, 2014. ACM.
- [Serrano *et al.*, 2014b] Marcos Serrano, Dale Hildebrandt, Sriram Subramanian, and Pourang Irani. Identifying suitable projection parameters and display configurations for mobile true-3d displays. In *Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices & Services*, MobileHCI '14, pages 135–143, New York, NY, USA, 2014. ACM.
- [Serrano *et al.*, 2015] Marcos Serrano, Barrett Ens, Xing-Dong Yang, and Pourang Irani. Gluey: Developing a head-worn display interface to unify the interaction experience in distributed display environments. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '15, pages 161–171, New York, NY, USA, 2015. ACM.
- [Serrano *et al.*, 2016] Marcos Serrano, Anne Roudaut, and Pourang Irani. Investigating text legibility on non-rectangular displays. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 498–508, New York, NY, USA, 2016. ACM.
- [Serrano *et al.*, 2017] Marcos Serrano, Anne Roudaut, and Pourang Irani. Visual composition of graphical elements on non-rectangular displays. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 4405–4416, New York, NY, USA, 2017. ACM.

- [Serrano, 2010] Marcos Serrano. *Interaction multimodale en entrée : Conception et Prototypage*. Thèses et habilitations, University of Grenoble, 2010. Thèse de doctorat Informatique préparée au Laboratoire d’Informatique de Grenoble (LIG), Université Joseph Fourier. 283 pages.
- [Shin *et al.*, 2012] Choonsung Shin, Jin-Hyuk Hong, and Anind K. Dey. Understanding and prediction of mobile application usage for smart phones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing, UbiComp ’12*, pages 173–182, New York, NY, USA, 2012. ACM.
- [Song *et al.*, 2011] Peng Song, Wooi Boon Goh, Chi-Wing Fu, Qiang Meng, and Pheng-Ann Heng. Wysiwyf: Exploring and annotating volume data with a tangible handheld device. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’11*, pages 1333–1342, New York, NY, USA, 2011. ACM.
- [Su *et al.*, 2010] Jing Su, Alyssa Rosenzweig, Ashvin Goel, Eyal de Lara, and Khai N. Truong. Timbremap: Enabling the visually-impaired to use maps on touch-enabled devices. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services, MobileHCI ’10*, pages 17–26, New York, NY, USA, 2010. ACM.
- [Ten Koppel *et al.*, 2012] Maurice Ten Koppel, Gilles Bailly, Jörg Müller, and Robert Walter. Chained displays: Configurations of public displays can be used to influence actor-, audience-, and passer-by behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’12*, pages 317–326, New York, NY, USA, 2012. ACM.
- [Tsandilas, 2018] Theophanis Tsandilas. Fallacies of agreement: A critical review of consensus assessment methods for gesture elicitation. *ACM Trans. Comput.-Hum. Interact.*, 25(3):18:1–18:49, June 2018.
- [Walter *et al.*, 2013] Robert Walter, Gilles Bailly, and Jörg Müller. Strikeapose: Revealing mid-air gestures on public displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’13*, pages 841–850, New York, NY, USA, 2013. ACM.
- [Wang Baldonado *et al.*, 2000] Michelle Q. Wang Baldonado, Allison Woodruff, and Allan Kuchinsky. Guidelines for using multiple views in information visualization. In *Proceedings of the Working Conference on Advanced Visual Interfaces, AVI ’00*, pages 110–119, New York, NY, USA, 2000. ACM.
- [Weiser, 1991] Mark Weiser. The computer for the 21st century. *Scientific American*, 265(3):94–104, 1991.
- [Wobbrock *et al.*, 2009] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’09*, pages 1083–1092, New York, NY, USA, 2009. ACM.
- [Yamashita *et al.*, 2017] Koki Yamashita, Takashi Kikuchi, Katsutoshi Masai, Maki Sugimoto, Bruce H. Thomas, and Yuta Sugiura. Cheekinput: Turning your cheek into an input surface by embedded optical sensors on a head-mounted display. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology, VRST ’17*, pages 19:1–19:8, New York, NY, USA, 2017. ACM.
- [Ye *et al.*, 2014] Hanlu Ye, Meethu Malu, Uran Oh, and Leah Findlater. Current and future mobile and wearable device use by people with visual impairments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’14*, pages 3123–3132, New York, NY, USA, 2014. ACM.

- [Zhang *et al.*, 2017] Yang Zhang, Gierad Laput, and Chris Harrison. Electrick: Low-cost touch sensing using electric field tomography. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 1–14, New York, NY, USA, 2017. ACM.
- [Zhao *et al.*, 2008] Haixia Zhao, Catherine Plaisant, Ben Shneiderman, and Jonathan Lazar. Data sonification for users with visual impairment: A case study with georeferenced data. *ACM Trans. Comput.-Hum. Interact.*, 15(1):4:1–4:28, May 2008.



# Selected publications

The following pages contain fifteen selected publications that give an overview of my work since the end of my PhD. I have organized the publications according to the chapters of this document. Papers resulting from the work of PhD students under my supervision (Sandra Bardot, Gary Perelman and Housseem Saidi) are marked with the symbol ◁ and papers in A\* conferences with the symbol ★.

## Publications on mobile interactions with large data spaces (chapter 2)

**P1 (★).** Marcos Serrano, Eric Lecolinet, and Yves Guiard. 2013. Bezel-Tap gestures: quick activation of commands from sleep mode on tablets. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13). ACM, New York, NY, USA, 3027-3036. DOI: <https://doi.org/10.1145/2470654.2481421>

**P2 (★).** Marcos Serrano, Barrett M. Ens, and Pourang P. Irani. 2014. Exploring the use of hand-to-face input for interacting with head-worn displays. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14). ACM, New York, NY, USA, 3181-3190. DOI: <https://doi.org/10.1145/2556288.2556984>

**P3.** Marcos Serrano, Dale Hildebrandt, Sriram Subramanian, and Pourang Irani. 2014. Identifying suitable projection parameters and display configurations for mobile true-3D displays. In Proceedings of the 16th international conference on Human-computer interaction with mobile devices and services (MobileHCI '14). ACM, New York, NY, USA, 135-143. DOI: <https://doi.org/10.1145/2628363.2628375>

**P4 (◁).** Sandra Bardot, Marcos Serrano, and Christophe Jouffrais. 2016. From tactile to virtual: using a smartwatch to improve spatial map exploration for visually impaired users. In Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '16). ACM, New York, NY, USA, 100-111. DOI: <https://doi.org/10.1145/2935334.2935342>

## Publications on multi-display tools and interfaces (chapter 3)

**P5.** Marcos Serrano, Barrett Ens, Xing-Dong Yang, and Pourang Irani. 2015. Gluey: Developing a Head-Worn Display Interface to Unify the Interaction Experience in Distributed Display Environments. In Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '15). ACM, New York, NY, USA, 161-171. DOI: <https://doi.org/10.1145/2785830.2785838>

**P6 (★, ◁).** Housseem Saidi, Marcos Serrano, Pourang Irani, and Emmanuel Dubois. 2017. TDome: A Touch-Enabled 6DOF Interactive Device for Multi-Display Environments. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17). ACM, New York, NY, USA, 5892-5904. DOI: <https://doi.org/10.1145/3025453.3025661>



**P7.** Louis-Pierre Berge, Marcos Serrano, Gary Perelman, and Emmanuel Dubois. 2014. Exploring smartphone-based interaction with overview+detail interfaces on 3D public displays. In Proceedings of the 16th international conference on Human-computer interaction with mobile devices and services (MobileHCI '14). ACM, New York, NY, USA, 125-134. DOI: <https://doi.org/10.1145/2628363.2628374>

**P8** (◁). Houssein Saidi, Marcos Serrano, and Emmanuel Dubois. 2016. Investigating the effects of splitting detailed views in Overview+Detail interfaces. In Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '16). ACM, New York, NY, USA, 180-184. DOI: <https://doi.org/10.1145/2935334.2935341>

#### Publications on freeform interfaces (chapter 4)

**P9** (★). Marcos Serrano, Anne Roudaut, and Pourang Irani. 2016. Investigating Text Legibility on Non-Rectangular Displays. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16). ACM, New York, NY, USA, 498-508. DOI: <https://doi.org/10.1145/2858036.2858057>

**P10** (★). Marcos Serrano, Anne Roudaut, and Pourang Irani. 2017. Visual Composition of Graphical Elements on Non-Rectangular Displays. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17). ACM, New York, NY, USA, 4405-4416. DOI: <https://doi.org/10.1145/3025453.3025677>

#### Other relevant publications cited in this manuscript

**P11** (◁). Gary Perelman, Marcos Serrano, Mathieu Raynal, Celia Picard, Mustapha Derras, and Emmanuel Dubois. 2016. DECO: A Design Space for Device Composition. In Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16). ACM, New York, NY, USA, 435-446. DOI: <https://doi.org/10.1145/2901790.2901893>

**P12** (★, ▷). Gary Perelman, Marcos Serrano, Mathieu Raynal, Celia Picard, Mustapha Derras, and Emmanuel Dubois. 2015. The Roly-Poly Mouse: Designing a Rolling Input Device Unifying 2D and 3D Interaction. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15). ACM, New York, NY, USA, 327-336. DOI: <https://doi.org/10.1145/2702123.2702244>

**P13** (★). Emmanuel Dubois, Marcos Serrano, and Mathieu Raynal. 2018. Rolling-Menu: Rapid Command Selection in Toolbars Using Roll Gestures with a Multi-DoF Mouse. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). ACM, New York, NY, USA, Paper 367, 12 pages. DOI: <https://doi.org/10.1145/3173574.3173941>

**P14** (★, ▷). Sandra Bardot, Marcos Serrano, Bernard Oriola, and Christophe Jouffrais. 2017. Identifying how Visually Impaired People Explore Raised-line Diagrams to Improve the Design of Touch Interfaces. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17). ACM, New York, NY, USA, 550-555. DOI: <https://doi.org/10.1145/3025453.3025582>

**P15** (★). Julie Ducasse, Marc J-M Mace, Marcos Serrano, and Christophe Jouffrais. 2016. Tangible Reels: Construction and Exploration of Tangible Maps by Visually Impaired Users. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16). ACM, New York, NY, USA, 2186-2197. DOI: <https://doi.org/10.1145/2858036.2858058>

# Bezel-Tap Gestures: Quick Activation of Commands from Sleep Mode on Tablets

Marcos Serrano      Eric Lecolinet      Yves Guiard

Telecom ParisTech - CNRS LTCI UMR 5141

46 Rue Barrault, 75013 Paris, France

{marcos.serrano, eric.lecolinet, yves.guiard}@telecom-paristech.fr

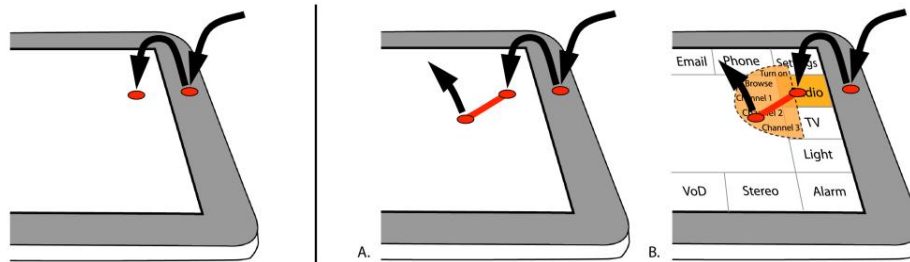


Figure 1: Left: A basic Bezel-Tap gesture involves a tap on the bezel immediately followed by a tap on the screen. Right: Bezel-Tap Slide, a hierarchical extension of Bezel-Tap supporting up to 64 commands. A) Expert mode. B) Novice mode.

**ABSTRACT**

We present Bezel-Tap Gestures, a novel family of interaction techniques for immediate interaction on handheld tablets regardless of whether the device is alive or in sleep mode. The technique rests on the close succession of two input events: first a bezel tap, whose detection by accelerometers will awake an idle tablet almost instantly, then a screen contact. Field studies confirmed that the probability of this input sequence occurring by chance is very low, excluding the accidental activation concern. One experiment examined the optimal size of the vocabulary of commands for all four regions of the bezel (top, bottom, left, right). Another experiment evaluated two variants of the technique which both allow two-level selection in a hierarchy of commands, the initial bezel tap being followed by either two screen taps or a screen slide. The data suggests that Bezel-Tap Gestures may serve to design large vocabularies of micro-interactions with a sleeping tablet.

**Author Keywords**

Interaction techniques; Mobile devices; Bezel Gestures; Accelerometers; Micro-Interaction; Marking Menus.

**ACM Classification Keywords**

H.5.2. Information interfaces and presentation: Interaction.

**General Terms**

Design; Human Factors; Performance; Experimentation.

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.

CHI 2013, April 27–May 2, 2013, Paris, France.

Copyright © 2013 ACM 978-1-4503-1899-0/13/04...\$15.00.

**INTRODUCTION**

A limitation of mobile devices is that they provide little support for quick commands. Micro-interactions, which rest on unconditional availability and fast access [3,4], are especially desirable for frequent actions such as, for instance, controlling a media player, checking emails and SMS, calling friends and family, getting the local map or the weather forecast, and more basic actions like copying, pasting, and application switching. Since mobile devices constantly switch to sleep mode to save energy, interaction is hampered by the need to reactivate them whenever they have gone to sleep, typically by pressing a physical button and sliding a widget on the screen. This problem is exacerbated when mobile devices are used to control multimedia devices (TV, set-top box, etc.) and home equipment (home automation, domotics systems), a usage likely to become common in the near future [10, 32]. In this scenario remote commands expand the large set of mobile applications. The challenge, then, is to allow always available and rapid access to a relatively large number of commands.

This paper presents Bezel-Tap Gestures, a novel interaction technique allowing micro-interactions on mobile devices. Not only can the technique serve to open an application and launch a favorite command rapidly, it can also wake the device if asleep. Using a combination of input modalities, it is based on the close succession of two input events: a tap on the bezel, detected by accelerometers, and a press or a sliding gesture on the touchscreen. While primarily designed for tablets, which have quite large bezels, Bezel-Tap Gestures can be adapted to smaller devices such as smartphones. Little visual attention is required and eyes-free operation is possible.

We present a field experiment indicating that the technique is robust to the risk of accidental activations (false detections). A second experiment investigated the optimal size of the vocabulary of commands, while exploring the usability of the different regions of the bezel (top, bottom, left, right) with the device in idle mode. We will explain how Bezel-Tap Gestures (and their extensions) can be used for activating a relatively large number of commands in both novice and expert mode, with a seamless transition from the former to the latter. Finally we will report performance data in a third experiment that illustrates the merits of an extension of Bezel Gestures [6] that allows selecting more commands than the original technique, in comparison with Bezel-Tap Gestures, it being understood that the techniques can be combined if desirable.

#### RELATED WORK

Gestural interaction provides an efficient means for activating commands rapidly, especially when hotkeys are not available as is generally the case on mobile devices. Marking menus [22] are a well-known technique relying on 2D gestures. They have inspired many HCI studies, some of them dedicated to mobile devices [11, 21]. One major merit of these techniques is to make it easy to discover and learn the gestures: all commands are visible in novice mode and gestures are learned incidentally from repetitive use. However, they interfere with common interactions on the screen (e.g. drag to pan), especially on mobile devices, which lack mouse buttons or an equivalent mechanism to differentiate interaction states [7]. The spatial shape or the space-time kinematics of certain gestures can serve as mode delimiters (e.g., pigtailed [16], rubbing [26] or rolling gestures [29]). These techniques are well suited for triggering a few dedicated commands at the application level, but we are looking for a more global shortcut mechanism that will not interfere with application shortcuts. Finally, we have the problem that gestural interaction does not work with the device in sleep mode, and unfortunately that is often the case due to the high power consumption of capacitive sensors.

Motion of the device in 3D space is another resource that can be exploited for triggering shortcuts, but the interpretation of 3D gestures poses an even trickier delimiter problem, the system having to distinguish intentional from background motion. DoubleFlip [30] solves this problem with a specific gesture that precedes the actual gestural commands and serves as an input delimiter. JerkTilts [5] extends this idea by merging the gesture delimiter and the command, yielding a vocabulary of up to eight commands. But, again, these interaction techniques seem more well suited for activating a small set of application-specific commands. And, unlike Marking menus, they fail to support a fluid transition from novice to expert mode. They may also be more error prone, especially when the delimiter is merged with the command because 3D gestures are inherently difficult to interpret. Finally, 3D gestures are more appropriate for smartphones than for

tablets because their size and inertial properties make them relatively difficult to move abruptly in 3D space.

Other approaches leverage extra input modalities. For instance, tap input through the pocket [27] has been proposed for embedded interaction on mobile phones without the need to access the keypad. Two other techniques based on acceleration sensing are ForceTap and Whack Gestures. ForceTap [14] distinguishes strong from gentle taps on the screen of a mobile phone. Whack Gestures [18] work by striking the mobile device, worn at the waist, with an open palm. Concerning tap techniques on tablets, TapSense [13] uses a microphone to detect different types of finger taps on the screen (nail, tip or knuckle).

Some techniques take advantage of the bezel to perform specific actions. In the seminal work of Hinckley et al. [15] on sensing techniques for mobile devices, touch on the bezel was used to initiate scrolling. Bezel tap has recently been used in a soft keyboard as a complementary input modality to insert a space [23]. Bezel Gestures, which are gestures on the touchscreen that start from the bezel, have been investigated for one-handed interaction on mobile devices in [6]. Results showed mark-based bezel gestures are faster and more accurate than free-form bezel gestures [6]. In [28], bezel swipes were proposed for scrolling and multiple selection on mobile touchscreens. Bezel gestures have been used for opening menus in [17] and in [12], which compares the performance of several menu layouts for mark-based bezel gestures, focusing on eyes-free interaction on small mobile devices and text entry. Today Bezel gestures are present by default on Android and iOS, to make a status panel appear, and on Windows 8 for switching between apps and bringing up items.

#### BEZEL-TAP GESTURES

The technique we introduce in this paper offers a supplementary input resource: it does not interfere with common interaction techniques, including Bezel gestures [6], and so it is usable without changing user habits on main mobile platforms. Most importantly, the technique makes it possible to both wake up the device and activate a command without the risk of battery over-draining, as explained in detail in a later section.

In its basic form a Bezel-Tap gesture involves two successive events recorded in different input channels, a tap on the bezel immediately followed by a tap (or a slide) on the screen. The first tap is detected by accelerometers and the second tap (or slide) by the touchscreen (Figure 1).

As shown in our false positive study below, there is very little risk of inadvertent activation. The fast succession of two events from different input channels (a tap *not concomitant* with a screen contact followed *within a short time interval* by a screen touch) is a low probability event that can serve as a unique signature. For instance, no Bezel-Tap gesture is recognized if the user double taps the screen or double taps the bezel. Moreover, a tap on the bezel

induces a high instant acceleration compared to background tablet movements and it is preceded by a small amount of time when the device moves very little (in normal usage, the user is holding the device, not shaking it when interacting with it).

This unique succession of input events acts as a mode delimiter, hence avoiding Bezel-Tap gestures to interfere with common interaction techniques, which typically rely on just the touchscreen. It also allows selecting a given command by considering the location of the contact on the touchscreen. Bezel-Tap gestures thus merge a gesture delimiter and a command selection. They can serve as a substitute for keyboard shortcuts on mobile devices, where a physical keyboard is generally absent.

As we will see in Experiment 2, selection performance is fast and accurate up to a fairly large number of commands, even in the absence of any screen feedback. Moreover, we will later present extensions (Bezel-Tap Slide and Bezel-Tap3) that allow selecting even more commands (Figure 1). We will also propose Bezel-Tap menus, a form of Marking menus [22] that are compatible with Bezel-Tap gestures. As with Marking menus Bezel-Tap menus make it possible to interact in the same way in novice and expert mode and are thus expected to provide a seamless novice-to-expert transition. This feature is indeed important for discovering and learning shortcuts.

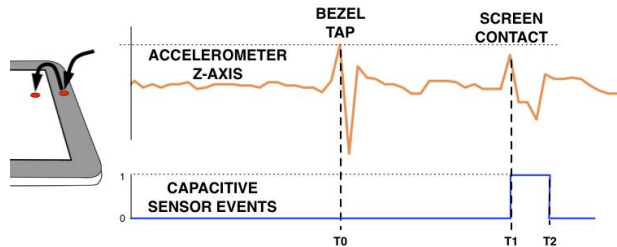
Bezel-Tap gestures were primarily designed for tablets, which indeed provide quite large bezels (their width is for instance between 1.8cm and 2cm on the iPad and the Galaxy Tab). Designed for holding the device, bezels favor bimanual interaction [34] and incidentally offer a large surface for tapping. Bezel-Tap gestures can also be used on the top and bottom bezels of smartphones, which are generally large enough. They could also be adapted to devices with very thin bezels by detecting taps on the edge or close to the edge on the back of the device.

**Tap detection**

To detect taps accurately using the accelerometer, the sampling must be done at least at a frequency of 100 Hz (1 sample every 10ms). The difference between three consecutive accelerometer samples along the Z axis is calculated in real-time, this giving the instant acceleration within a frame of 20ms (Figure 2). Then we use three conditions to detect a tap on the bezel:

- **Threshold:** The instant acceleration must be higher than a given threshold. This threshold was determined in a study on which we report at the beginning of Experiment 1.
- **Sign:** The sign of the acceleration along the Z axis (positive or negative, depending on the orientation of the accelerometer) indicates whether the user taps on the front or on the back of the device. We can thus dismiss all the taps on the back.

- **Little acceleration:** As said before, this instant acceleration must be preceded by a small amount of time, 30ms from our experiments, with very little acceleration. This property allows making the difference between noise and bezel taps, because normally the user is holding the device before tapping.



**Figure 2. Accelerometer and touchscreen signals elicited by the first two events of a Bezel-Tap gesture.**

A Bezel-Tap gesture involves a subsequent tap or slide on the touchscreen that must occur within a short time interval (Figure 2). So we have two temporal constraints:

- The time delay between the two events must be greater than a few milliseconds (50ms with our hardware) so that a Bezel-Tap gesture cannot start with a screen tap (this could otherwise happen in the case of a double tap on the screen if the user taps hard).
- The delay must not exceed 600ms to avoid taking unrelated events into account (i.e., a screen contact occurring long after a tap). This value was obtained experimentally from Exp. 1, described below, where mean inter-tap time was 437ms (SD 143ms).

Incidentally, it is worth noticing that Bezel-Tap gestures allows identifying a location on the device (that serves to select a command) but only rely on the touchscreen to do so. While accelerometers could theoretically serve to detect tap locations, hardware currently available on mobiles devices would make this hardly feasible. Using two input modalities (the accelerometer and the touchscreen) solves this problem: the touchscreen provides information the accelerometer is unable to provide while avoiding interfering with normal use of the touchscreen.

**Waking-Up from Sleep Mode: Power Considerations**

Bezel-Tap Gestures make it possible to trigger commands quickly, even with the device in sleep mode. This property is useful even if the device is password protected as many commands are not security threatening, this being especially true when the device serves to control home equipment. And recent surveys reveal anyway that more than 60% of mobile users do not use passwords [19].

This reactivation feature requires the accelerometer to remain permanently powered. The small power consumption of accelerometers makes this feasible without considerably reducing the battery life. This is an interesting property of accelerometers, which contrasts with most other input devices, especially capacitive screens, which require

about  $3000\mu\text{A}$  in idle mode and  $5000\mu\text{A}$  in active mode. This is why techniques only relying on the touchscreen, for instance Bezel Gestures, would not be appropriate for reactivating the device. Tapping the bezel is also more convenient than pressing a button, a very low consumption input device, simply because it represents a very large target that can hardly be missed.

Some components of a mobile device, such as the GSM and the CPU, remain powered in sleep mode in order to receive calls or SMS. As shown in [8] the power consumption is about  $69\text{mW}$  during sleep mode for a 2.5G Android-based smartphone (hence allowing about 45h of battery life). The power needed by the accelerometers used in our prototype ( $1.2\text{mW}$ ) should thus only reduce the battery life of this device by about three quarters of an hour. This figure would be much less using recent, more power-efficient models such as the LIS3DH [24], which may use as little as  $7.2\mu\text{W}$ . And performance is permanently improving thanks to research on the continuous sensing of human activity through mobile sensors [25, 9]. The solution we are examining in this paper is hence indeed viable in terms of power consumption, especially in the case of tablets, which usually have a fairly large battery (e.g., the battery of the Samsung Galaxy Tab 10.1 supplies  $7000\text{mAh}$ ).

Tapping on the bezel will power up the capacitive touchscreen of the device for a few milliseconds. While, as seen above, the consumption of a capacitive screen is far from negligible, this is not a problem because such an event will only occur rarely, even in mobile context, as we will see in the false positive study presented below.

The inter-tap time (the amount of time elapsed from the bezel tap to the screen touch) is larger than the time needed to reactivate the touchscreen. In theory, touchscreens have very small reactivation latencies: less than  $10\text{ms}$  for an Atmel maXTouch on a Samsung Galaxy Tab tablet. Using a camera, we approximately measured how much time was needed for reactivating an iPhone and an iPad. According to our measurements wake-up takes about  $240\text{ms}$  (6 frames in a 25 fps image) after pressing the physical button of these devices. This duration is about half of the inter-tap time we measured in Experiment 2. Our technique is thus already compatible with common commercial devices.

#### Hardware

We first tested our tap algorithm with a Samsung Galaxy tablet Tab 10.1 (display:  $10.1''$ , resolution  $1280 \times 800$  pixels, dimensions:  $256.7 \times 175.3 \times 8.6$  mm, original weight:  $565\text{g}$ ) running Android 3.2, containing one built-in three-axis accelerometer situated on the top-right corner of the device (position found from a device tear down). Pre-tests showed that this accelerometer permits to detect taps with sufficient accuracy on the top and right bezel regions ( $95\%$  of taps detected, this resulting from its location), with a general success rate of  $84\%$  over all bezel regions.

The Bezel-Tap technique can hence already work with existing equipment, but not on all bezel regions. This led us to build a prototype to perform a more general experiment. To do so we fixed an external accelerometer on the back of the bottom-left corner of the device (Figure 3), an easy solution from a manufacturing point of view as accelerometers are cheap, light and small objects (about  $2 \times 2 \times 1\text{mm}$ ).



**Figure 3.** Test prototype with an external accelerometer.

We used an ADXL335 3-axis  $\pm 3\text{g}$  accelerometer [1], available in a small ( $4 \times 4 \times 1.45\text{mm}$ ) plastic chip package. The accelerometer was plugged to a micro-controller Arduino Nano 3.0 [2], connected through a USB wire to a PC. The wire was fixed together with the tablet USB power cable in order to minimize its detrimental effect on the manipulability of the prototype. As the power cable had to be plugged anyway during tests, having a wired prototype entailed no significant extra cost. A Java program was written in order to parse values from the external accelerometer and dispatch them to the tablet through the WiFi network. The back of the prototype was shielded in order to resist extensive experimentation and to preserve a handy manipulation.

#### EXPERIMENT 1: TRUE AND FALSE POSITIVES STUDY

The goal of this experiment was first to optimize the threshold for tap detection (true positives), and second to evaluate the probability of false positives when using the tablet in the field. For practical reasons, we only used the built-in accelerometer in this experiment since the external accelerometer needed to be plugged to a computer. Hence, we only considered taps that the built-in accelerometer could detect with sufficient accuracy, that is to say taps on the top and right bezel regions.

#### Threshold setting

We asked six users to perform 40 taps each, 20 on each of the top and right regions of the bezel. The logging software was running on a PC and there was no visual feedback. The participants were first explained what a bezel tap was and the difference between a tap and a touch. We asked users to tap, in separate blocks of trials, on the two regions of the bezel while holding the tablet in their hands. Region order was counterbalanced.



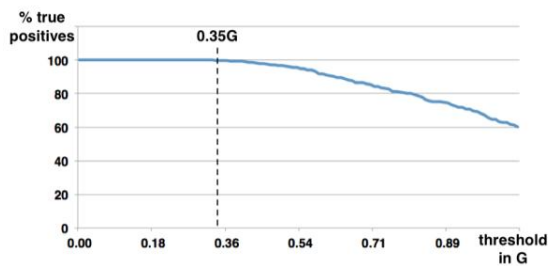


Figure 4. True positive rate versus tap threshold value in G.

**Results**

We analyzed logs studying the relationship between tap threshold and the percentage of taps recognized. The results (Figure 4) show that true positive rate starts decreasing above a threshold of 0.35G (3.4 m.s<sup>-2</sup>). We thus decided to use this value for tap detection.

**False Positives: A Field Study**

In order to evaluate the probability of inadvertent activations of Bezel-Tap gestures (false positives), we completed a field study with 12 people using a Samsung Galaxy Tab 10.1 equipped with a web browser implementing Bezel-Tap Gestures. We chose a web browser for two reasons: it is one of the most common applications used in tablets [33], and it can be used for displaying various kinds of data. We gave the participants a tablet for 24 hours, asking them to use the web browser for at least one hour in total wherever they wanted.

**Data collection**

We collected 25 hours and 26 minutes of tablet use from the 12 participants (median 100 min per participant, SD 70 min). We logged the data from the internal accelerometer and the touchscreen input.

**Results**

All users performed the experiment at home. There were no false positives at all, considering both Bezel-Tap Gestures and the variants we present in the second part of this paper. A few taps were detected by the accelerometer (on average 7.7 taps per hour of use) but none of them were followed by a screen contact in less than 600ms, so that no false positive was detected. This field study suggests that our technique is robust to common usage.

**False Positives: Taps in Mobility**

We also investigated the effect of carrying the device in mobile context, in public transportation or while riding a bike to see how many taps would be detected by the accelerometer just because of the motion of the device. A high number of detections would increase the probability of inadvertent activations of Bezel-Tap gestures when the user uses the device and drain the battery because taps reactivate the capacitive sensor of the screen. We hence conducted an experiment carrying a Galaxy Tab tablet inside a backpack during subway, bus and bike journeys.

**Data collection**

The author collected 8 hours and 37 minutes of tablet’s internal accelerometer log in mobile context (3h14min of bus, 1h1min of subway and 4h22min of bike).

**Results**

On average 6.15 taps per hour were detected: 9.5 taps/hour on bus, 4.9 taps/hour on subway and 3.8 taps/hour on bike. This small number let us drop our concerns. Inadvertent activations should almost be as rare as in the previous study and the added power consumption is negligible.

**Conclusion**

These studies show there is very little risk of an accidental activation of the bezel tap technique. We only used the built-in accelerometer (and taps on the top and right bezel regions), thus showing that our technique can be implemented using current commercial devices. These results suggest that the risk of false positives would remain very low if a second accelerometer was used for detecting taps on all four bezel regions, as proposed above.

**EXPERIMENT 2: BEZEL-TAP PERFORMANCE**

The goal of this experiment was to evaluate the performance of Bezel-Tap gestures in terms of speed and precision depending on the size of the command set. Additionally, we also wanted to compare performance for the different regions of the bezel (top, bottom, left, right). Basic Bezel-Tap gestures were used for this experiment (a tap on the bezel followed by a tap on the screen). No help was provided to the participants (the screen of the tablet was full black).

**Task and Instructions**

The participants had to use, in separate conditions, the four bezel regions located at the top, bottom, left, or right of the device, and the two-tap sequence served to select within the specified region one item among 4, 5, 6, or 7. The items consisted of squared areas, the same size, laying on the periphery of the touchscreen as shown in Figure 5.

We decided to deliver to our participants visual stimuli that fully specified what they had to do, using a laptop screen placed just in front of them (Figure 5), thus simulating the case of a highly practiced user. Our task instructions emphasized accuracy, asking our participants to minimize their error rate (primary demand) while wasting no time (secondary demand).

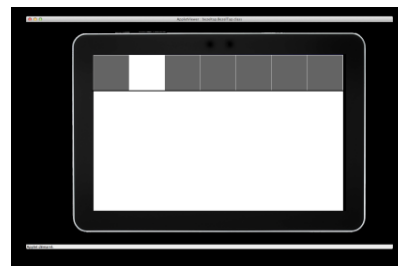


Figure 5. An example, with N = 7, of a stimulus of Exp. 2, as displayed on the laptop screen.

The participants were provided with knowledge of results: Following each gestural response, the stimulus display changed on the laptop screen, with the blue target rectangle either turning green in case of a correct response or the wrongly selected rectangle being highlighted in red in case of an error.

### Participants

Ten men and two women, aged 26.5 years on average, volunteered for the experiment. Eight were familiar with tablets.

### Apparatus

We used the Samsung Galaxy Tab prototype with two accelerometers (the built-in and the external accelerometer, as described in previous section). We managed to remove the status bar that the Android system displays by default at the bottom of the screen, and so we had at our disposal 100% of the screen surface area. Next, we covered the bezel with black tape, hiding the tablet logo and the camera objective, so that the bezel surface was homogeneous all around. The final weight of the experimental device was 650g.

### Procedure

This experiment followed a 4 x 4 within-participant design with *menu region* (top, bottom, left, or right) and *number of menu items* ( $N = 4, 5, 6, \text{ or } 7$ ) as factors. Four blocks of trials were run for each menu location, the menu-location factor being balanced by means of a 4x4 Latin square. Each block of trials required  $4+5+6+7 = 22$  selections,  $N$  being increased gradually: the first four stimuli (presented in randomized order) asked for the selection of one item among four, and so on for  $N=5, 6$  and  $7$ .

The sequence of events in a trial was as follows: stimulus onset, then gestural response, then knowledge of results display, then time out. The experiment, in which we recorded a total of  $22 \times 16 \times 12 = 4,224$  double-tap gestures, lasted about 30 min per participant.

### Data Collection

We recorded signals from our two accelerometers (the built-in one plus the external) as well as touch events from the screen tablet. Beside success rates, we measured total trial completion time, from stimulus onset to screen release. The experiment was videotaped from beginning to end.

### Results and Discussion

Our experimental manipulations had little or no influence on the speed of performance. It took our participants a pretty constant 1.5s to complete the various gestures, regardless of  $N$  and regardless of the menu location. In contrast, as shown in Figure 6, the success rate declined monotonically with the increase in set size. It was on average 96.9% at  $N = 4$ , 96.5% at  $N = 5$ , 90.4% at  $N = 6$  and 90.6% at  $N = 7$  ( $F_{3,33} = 15.87, p < .0001$ ). Although performance accuracy seems somewhat poorer for the top location, this reflects an effect present in essentially two participants: in fact there was no consistent effect of menu

location ( $F_{3,33} = 1.05, p = .38$ ), and no consistent interaction between menu location and  $N$  ( $F_{9,99} < 1$ ).

Notice in Figure 6 that the slope of the error-rate curve tends to increase from odd to even but tends to stagnate from even to odd, suggesting that odd-numbered menus, which are symmetric about their central item, allow more accurate performance than even-numbered menus.

An inspection of the confusion matrix in the case  $N = 5$  revealed a better accuracy for odd-numbered than even-numbered items: the error rate was respectively 3.5%, 0.9% and 7.6% for all, odd and even items ( $t_{11} = 3.85, p = .001$  one-tailed). Likewise, performance was more accurate for external items, situated at the corners of the screen, than internal items: we recorded 0.3% errors on average over items #1 and #5 vs. 5.7% errors on average over items #2, #3, and #4 ( $t_{11} = 4.31, p = .006$ ). Those results are consistent with previous research pointing out the fact that corners and physical edges are useful landmarks for both blind and sighted people [20].

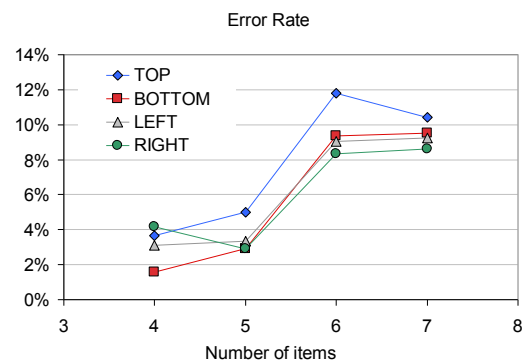


Figure 6. Error rate as a function of  $N$ , for each menu location.

To sum up, this experiment confirmed the usability of the Bezel-Tap technique in all four regions of the device in sleep mode. Selection times, on the order of 1.5 sec (with the reaction time included), were compatible with the micro-interaction concept [3]. Performance accuracy was not lower for  $N = 5$  (96.5%) than for  $N = 4$  (96.9%). One practical suggestion that arises from these results is that a set size of five items is optimal for the technique. The high accuracy for  $N = 5$  in little-trained participants, confirmed that the technique is worthy of consideration in relation with shortcut commands.

### EXTENSIONS AND BEZEL-TAP MENU

Bezel-Tap Gestures can be extended to allow selecting a large number of items. We propose two new techniques that rest on a hierarchical organization (Figure 1-right). Bezel-Tap3 (BT3) involves three taps: one on the bezel and two on the screen. The second tap selects a group of items and the third tap an item in this group. Bezel-Tap Slide (BTSlide) involves a tap on the bezel followed by a slide on the screen. It works in the same way as BT3 except that the

starting and ending points of the slide play the role of the second and third taps.

Both techniques rely on Bezel-Tap menus (Figure 7) for the novice mode. Due to the nature of BT3 and BTSlide gestures, Bezel-Tap menus are hierarchical. Their first level consists of five rectangular items for all four bezel regions (Figure 7). Normal item size is 256x160 px, but according to Experiment 2 results we decided to expand even items and reduce corner items by 15% in order to increase even-numbered items success rate.

The second level of the menu rests on 180° radial menus. The second tap (resp. the starting point of a slide) selects one of these radial menu, for instance the "Radio" menu in Figure 7. The third tap (resp. the ending point of a slide) selects an item in this menu. The selection of the radial menu hence depends on the location of the tap on the screen (which must lay in one of the rectangles on the periphery of the screen) and the selection of the item on the direction of the segment between the this tap and the following one (resp. the direction of the slide).

Menus only appear if the user waits more than 600ms between the second and the third tap. A complete novice user just needs to tap on the bezel then anywhere on the screen and wait for 600ms. The bezel menu then appears and the user can select the proper radial menu and the proper item in this menu. A more knowledgeable user will select the proper radial menu by tapping on the appropriate rectangle, then wait the radial menu to appear to select the desired item. An expert user will bypass these stages by performing all taps (or slides) without any visual feedback. In any case, user interaction is similar in totally novice, partially novice and fully expert modes, a property expected to provide a seamless novice-to-expert transition.

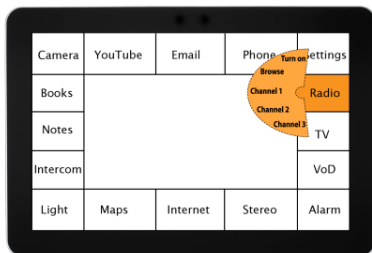


Figure 7. Two-level Bezel-Tap Menus.

To make interaction easier, we did not assign menus to corner items (which can serve as one-step shortcuts for very frequent commands). This design allows for a total of 12 radial menus (3 per region). Each radial menu offers five items over 180° to avoid moves back to the bezel. A Bezel-Tap menu can thus comprise a total of 64 items: 4 one-step shortcuts and 60 menu items (4 regions x 3 menus x 5 items). While this number may seem large for a vocabulary of shortcuts for a standard user, this is useful for the novice mode as the menus can also contain commands that will not

be triggered in expert mode. Having related commands grouped together in menus is convenient and allows incidental learning.

In the next section we will compare the performance in expert mode of BT3, BTSlide and an extension of Bezel Gestures [6]. Interestingly, these three techniques can coexist: Bezel Gestures do not rely on the same initial event as Bezel-Tap techniques, which themselves do not rely on the same second event. Quite a large number of commands (3x64=192) can hence theoretically be provided if these techniques are used together.

**EXPERIMENT 3: TWO-LEVEL COMMAND HIERARCHY**

The goal of this study was to evaluate the usability of Bezel-Tap Gestures for command selection in expert mode. Participants were asked to perform the full gestural sequence illustrated in Figure 8, using either BT3, BTSlide or an extension of Bezel Gestures [6] that allows selecting more commands than the original technique. In our implementation of this technique, the finger starts to slide from the bezel, reaches the selected menu and finally slides out of that menu in one of the five possible directions.

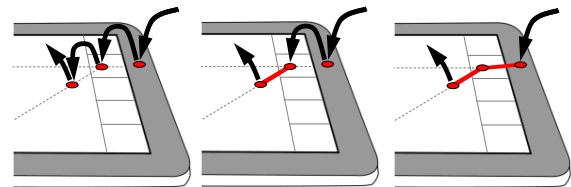


Figure 8. The hierarchical gesture techniques tested in Exp. 3, BT3 (left), BTSlide (center) and Bezel Gestures (right).

This experiment also investigated the impact of vision on performance. During a first session, which formed the main part of the experiment, the participants were allowed to watch the tablet (whose screen was black, hence providing no feedback) while carrying out the tasks. But during a second session, run with a subset of the participants, the tablet was hidden, as a test of whether the gestures could be successfully performed in eyes-free mode (Fig. 9).



Figure 9. A participant performing a gesture under an apron in the eyes-free session.

The right hand (of right-handers) usually staying by default near the bottom-right corner of the tablet, the right and bottom regions of the bezel are most easily accessible for tapping. In this experiment we decided to focus on these two comfortable regions, leaving aside the top and left regions.



We again asked participants to minimize their error rate with no waste of time and provided them with knowledge of result. The classification of responses for the first selection (menu) was 2D: if the user touched beyond the menu rectangle, the gesture was judged false. The classification of responses for the second selection (item) in the radial menu was just angular, with no distance limit.

**Participants and Apparatus**

Nine men and three women performed the experiment, aged 27.5 years on average, all right-handed? Six of them also participated in the second session with the tablet hidden. We used the same apparatus as in Exp. 2 and again the experiment was videotaped from beginning to end.

**Design**

We used a 2x3 within-participant design with *menu region* (right, bottom) and *interaction technique* (BT3, BTSlide and Bezel gestures) as factors. In the visible-tablet session, three blocks of trials were run for each interaction technique, the interaction technique factor being balanced by means of a 3x3 Latin square. Each block was composed of 33 trials: 3 shortcuts (corner items) plus 6 menus x 5 items/menu. The Menu Position factor was sorted randomly. The experiment involved 33 x 9 = 297 trials per participant (3,564 in total) and lasted about 30 minutes. In the shorter invisible-tablet session, which lasted about 20 minutes, two blocks of trials were run for each interaction technique, using the same 33 trial-blocks. 33 x 6 = 198 trials were ran per user, 1188 in total.

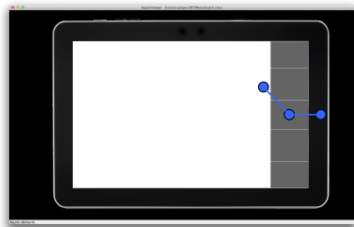


Figure 10. An example of a stimulus of Exp. 3, as displayed on a laptop screen.

The sequence of events in a trial was as follows: stimulus display, as shown in Figure 10, then gestural response, then knowledge of results display, then time out.

**Dependent Measures**

We will report data computed on average over participants: the dependent measures are the success rate as well as the total duration of the gestural sequence measured from the stimulus onset to the final release (in practice, we computed within-participant medians rather than means because the time distributions were slightly skewed).

**Results and Discussion**

Let us start with the data of the main session, in which the tablet was visible. Concerning accuracy, there was a statistically significant effect of the technique factor on error rate ( $F_{2,22} = 5.5, p = .01$ ), with an average error percentage of 5.2% for BT3, 4.5% for BTSlides, and 8.7%

for Bezel Gestures (Figure 11). While a Tukey post-hoc test indicated no significant difference between BT3 and BTSlide, the difference between BTSlide and Bezel Gesture was significant ( $p < .05$ ). Thus the selection was less error prone with BTSlides than Bezel Gestures, with a mean advantage of 4.2%. There was also a clear-cut advantage for the bottom relative to the right location of the menu ( $F_{1,11} = 12.6, p = .004$ ). There was no significant interaction between the two factors ( $F_{2,22} = 2.15, p = .14$ ).

The large difference in accuracy for the right and bottom bezels mainly comes from outliers (12.5%, 3 std. dev. from mean). Using the median instead of the mean, error rates for BT3 are 4.5% (right) vs. 4.5% (bottom), for BT-Slide 5.5% (right) vs. 1% (bottom) and for Bezel Gestures 9% (right) vs. 5.5% (bottom). Since most users held the left side of the tablet, the difference between the right and bottom error rates may be due to the fact that the tablet is more likely to move on its perpendicular axis when the user interacts on its right side.

With respect to performance speed, we found a significant effect of the technique factor on trial completion time ( $F_{2,22} = 50.6, p < .0001$ ). As visible in Figure 12 and as confirmed by Tukey tests, performance was faster with Bezel Gestures than BT3 and BTSlide (on average a 483ms and a 467ms difference, respectively), with the last two techniques not differing consistently between each other. The effect of the region factor was marginally significant, the difference being now in favor of the right region ( $F_{1,11} = 4.76, p = .052$ ). There was no significant interaction between the two factors ( $F_{2,22} = 1.12, p > .3$ ).

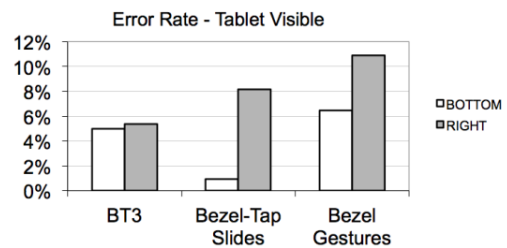


Figure 11. Error rate for the three techniques and the two menu regions, with the tablet visible.

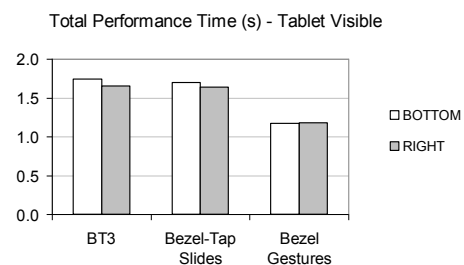


Figure 12. Performance speed for the three techniques and the two menu regions, tablet visible.

Combining the accuracy and speed information, the data from this experiment thus suggest that the two variants of the Bezel-Tap technique allow more accurate but slower performance than Bezel Gestures [6].

For the shorter eyes-free session, error rates were unsurprisingly higher, now with a grand average of 14.5% (to be compared with the 6.1% error rate observed in the visible-tablet session). We found more accurate performance with BT3 and BTSlides, but the effect of the technique was not significant ( $F_{2,10} = 1.28, p > .3$ ). However, the effect of the menu region was significant ( $F_{1,5} = 8.81, p = .03$ ), with consistently more errors recorded in the right rather than bottom region of the tablet (on average 20.4% vs. 8.7%, respectively). The interaction was significant ( $F_{2,10} = 4.69, p = .03$ ), reflecting the fact that the region effect was less marked with the BT3 than the other two techniques.

Turning to performance speed, we found of course that hiding the device under an opaque shield slowed down performance (from 1.47s to 1.75s, on average). The technique factor had a highly consistent effect on trial completion time ( $F_{2,10} = 40.3, p < .0001$ ), reflecting the speed superiority of Bezel Gestures. The effect of the region factor on speed was marginally significant ( $F_{1,5} = 5.5, p = .06$ ), reflecting faster performance on the right region.

This extra session shows that Bezel-Tap Gestures is quite a robust technique: The success rate remains around 86% and the two-level selection time around 2s in the case of a complete deprivation of visual monitoring.

### Conclusion

In sum, it appears that after little training all three techniques allow accurate selection in a two-level hierarchical menu. Bezel-Tap techniques (BT3 and BTSlide) were more accurate than Bezel Gestures (error rates were 5.2%, 4.5% and 8.7%, respectively), but slower. Anyway, the speed was pretty good for Bezel-Tap techniques (1.6s on average), this making them appropriate for micro-interactions, defined in [3] as “interactions with a device that take less than four seconds to initiate and complete”. This result is even more interesting considering that Bezel-Tap gestures can not only select a command but also reactivate the device in this short amount of time.

### EXPECTED APPLICATIONS

As said before, Bezel-Tap Gestures are especially useful for accessing frequent commands when the tablet is in sleep mode. Mobile devices constantly switch to sleep mode in order to save energy. A recent survey shows that Android users have an average of 177 applications on their phones, using 73 of them occasionally and 32 very often [31]. The study shows some usage patterns that would benefit most from shortcuts, such as re-accessing the last application. Examples are visualizing the weather, map or calendar; controlling music or video player; opening the camera or one of the various video or photo-taker applications (such as Instagram); checking one’s email, SMS, twitter or

Facebook messages. Allowing always-available low-consuming access to most frequent commands will improve mobile interaction.

Novel usages of tablets require even more commands than needed by this already over-populated software environment. In particular, tablets start being used as remote controls for interacting with home equipment and multimedia devices [10, 32]. In this scenario, the challenge is to allow always available and rapid access to commands (e.g., turning on the light, changing the temperature, selecting one’s favorite TV or radio channels, one’s favorite Web applications or music streaming services, etc.) Users need methods to organize and access this large set of commands. Bezel-Tap Gestures are especially well suited for this task because more often than not the tablet will be in sleep mode when the user wants to control home equipment.

The extended version of Bezel Gestures can be used as an extra resource if more than 64 commands are needed. As already mentioned, both techniques can be used together because they do not rely on the same initial event. However, Bezel Gestures should be preferred for commands that are unlikely to be performed when the tablet is asleep (typically, commands that are only used once a given application is opened, or commands for copying, pasting, application switching, etc.).

### CONCLUSION AND FUTURE WORK

We have presented Bezel-Tap Gestures, a technique that allows selecting commands quickly on mobile devices. Bezel-Tap Gestures do not interfere with common interaction techniques when the device is in active mode. They also work in sleep mode, both waking up the device and triggering a command. We conducted a false positive study that validated the robustness of the technique. Another experiment confirmed the viability of basic Bezel-Tap Gestures with no visual help from the screen. A success rate of 96.5% and a completion time of 1.4s was for instance obtained for a set of 25 items (5 per Bezel region, a set size that seems optimal).

We then proposed two extensions which can accommodate more commands, Bezel-Tap3 (BT3) and Bezel-Tap Slide (BTSlide). These techniques rest on Bezel-Tap menus, which are hierarchical and can contain up to 64 items. The first level (menu selection), arranged all around the screen, contains 4 shortcuts and 12 menus. The second level (item selection within the menu) consists of 180° radial menus containing five item. We performed an experiment to compare the expert-mode performance of BT3, BTSlide and an extension of Bezel Gestures. All three techniques allowed accurate selection, Bezel-Tap techniques being more accurate (around 95% of correct recognition) but slower than Bezel Gestures. The time performance was pretty good for all techniques, making them very appropriate for designing micro-interactions.

How Bezel Tap gestures might help the design of immediate interaction on devices with very thin bezels is a question for future research. As said before, taps could be detected the edge or close to the edge on the back of the device. The size of the device also matters. We implemented a smartphone prototype using a HTC Hero under Android. According to pre-tests, we reduced the size of the hierarchical menu, using 3 items per bezel region for the first level and 3 items for the radial menu in the second level. Preliminary experiments showed the feasibility of the technique using two hands, one for holding the device and the other one for interacting with the bezel and the screen. However, more work is necessary to estimate false positives rates in real usage and to check whether the technique could be used with a single hand.

#### ACKNOWLEDGMENTS

We acknowledge the financial support of the Quatro2 project funded by Systematic Paris-Region. We thank the anonymous reviewers of this article for their relevant critiques and recommendations. Many thanks to S.T. Perrault and D.-B. Vo for their many suggestions.

#### REFERENCES

- ADXL335 Accelerometer, Analog Devices. <http://www.analog.com/>
- Arduino Nano board. <http://arduino.cc/en/Main/ArduinoBoardNano>
- Ashbrook, D. Enabling mobile microinteractions. PhD Thesis, Georgia Institute of Technology (2009).
- Asbrook, D., Baudisch, P., White, S. Ninya: Subtle and Eyes-Free Mobile Input with a Magnetically-Tracked Finger Ring. *In Proc. of CHI 2011*, ACM, 2043-2046.
- Baglioni, M., Lecolinet, E. and Guiard, Y. Jerktilts: Using accelerometers for eight-choice selection on mobile devices. *In Proc. of ICMI 2011*, ACM, 121-128.
- Bragdon, A. et al. Experimental Analysis of Touch-Screen Gesture Designs in Mobile Environments. *In Proc. of CHI 2011*, ACM, 403-412.
- Buxton, W. A three-state model of graphical input. *In Proc. of Interact 1990*, Elsevier, 449-456.
- Carroll, A. and Heiser, G. An analysis of power consumption in a smartphone. *In Proc. of USENIXATC 2010*, USENIX, 21-35.
- Consolvo, S. et al. Activity sensing in the wild: a field trial of ubifit garden. *In Proc. of CHI 2008*, ACM, 1797-1806.
- Control 4, Smart Home products. <http://www.control4.com/>
- Francone J. et al. Wavelet Menus on Handheld Devices: Stacking Metaphor for Novice Mode & Eyes-Free Selection for Expert Mode. *In Proc. of AVI 2010*, ACM, 173-180.
- Jain, M. and Balakrishnan, R. User Learning and Performance with Bezel Menus. *In Proc. of CHI 2012*, ACM, 2221-2230.
- Harrison, C., Schwarz, J., Hudson, S. TapSense: Enhancing Finger Interaction on Touch Surfaces. *In Proc. of UIST 2011*, ACM, 627-634.
- Heo, S., Lee, G. ForceTap: Extending the Input Vocabulary of Mobile Touch Screens by adding Tap Gestures. *In Proc. of MobileHCI 2011*, ACM, 113-122.
- Hinckley, K. et al. Sensing techniques for mobile interaction. *In Proc. of UIST 2000*, ACM, 91-100.
- Hinckley, K. et al. Design and Analysis of Delimiters for Selection-Action Pen Input Phrases in Scriboli. *In Proc. of CHI 2005*, ACM, 451-460.
- Hinckley, K., Yatani, K. Radial Menus With Bezel Gestures. Publication date: 08/25/2011. Patent publication number 20110209093.
- Hudson, S. et al. Whack Gestures: Inexact and Inattentive Interaction with Mobile Devices. *In Proc. of TEI 2010*, ACM, 109-112.
- Identity Fraud Report. Javelin Strategy (2012).
- Kane, S. et al. Usable gestures for blind people: understanding preference and performance. *In Proc. of CHI 2011*, ACM, 413-422.
- Kin K., Hartmann B., Agrawala M. Two-handed marking menus for multitouch devices. *ACM Trans. Comput.-Hum. Interact.* 18(3): 16 (2011)
- Kurtenbach, G. and Buxton, W. 1991. Issues in combining marking and direct manipulation techniques. *In Proc. of UIST 1991*, ACM, 137-144.
- Li, F., Guy, R., Yatani, K., Truong, K. The 1Line Keyboard: A QWERTY Layout in a Single Line. *In Proc. of UIST 2011*, ACM, 461-470.
- LIS3DH 3-axis accelerometer, STMicroelectronics. <http://www.st.com/internet/analog/product/250725.jsp>
- Lu, H. et al. The jigsaw continuous sensing engine for mobile phone applications. *In Proc. of SenSys 2010*, ACM, 71-84.
- Olwal, A., Feiner, S., Heyman, S. (2008). Rubbing and Tapping for precise and rapid selection on touch-screen displays. *In Proc. of CHI 2008*, ACM, 295-304.
- Raikonen, S. et al. Tap Input as an Embedded Interaction Method for Mobile Devices. *In Proc. of TEI 2007*, ACM, 263-270.
- Roth, V. Turner, T. Bezel Swipe: Conflict-Free Scrolling and multiple selection on mobile screen devices. *In Proc. of CHI 2009*, ACM, 1523-1526
- Roudaut, A., Lecolinet, E. and Guiard, Y. Microrolls: Expanding Touch-Screen Input Vocabulary by Distinguishing Rolls vs. Slides of the Thumb. *In Proc. of CHI 2009*, ACM, 927-936.
- Ruiz, J. and Li, Y. Doubleflip: a motion gesture delimiter for mobile interaction. *In Proc. of CHI 2011*, ACM, 2717-2720.
- Shin, C., Hong, J-H., and Dey, A. Understanding and prediction of mobile application usage for smart phones. *In Proc. of UbiComp 2012*, ACM, 173-182.
- Tsekelves, E. et al. Interacting with Digital Media at Home via a Second Screen. In Ninth IEEE International Symposium on Multimedia Workshops (2007).
- Understanding Tablet Device Users. AdMob/Google, US (March 2011).
- Wagner, J., Huot, S., Mackay, W. BiTouch and BiPad: Designing Bimanual Interaction for Hand-held Tablets. *In Proc. of CHI 2012*, ACM, 2317-2326.

# Exploring the Use of Hand-To-Face Input for Interacting with Head-Worn Displays

Marcos Serrano<sup>1,2</sup>, Barrett Ens<sup>1</sup>, Pourang Irani<sup>1</sup>

<sup>1</sup>University of Manitoba

Winnipeg, MB, Canada, R3T 2N2

{marcos, bens, irani}@cs.umanitoba.ca

<sup>2</sup>IRIT - University of Toulouse

Toulouse, France, 31062

marcos.serrano@irit.fr

## ABSTRACT

We propose the use of *Hand-to-Face* input, a method to interact with head-worn displays (HWDs) that involves contact with the face. We explore Hand-to-Face interaction to find suitable techniques for common mobile tasks. We evaluate this form of interaction with document navigation tasks and examine its social acceptability. In a first study, users identify the cheek and forehead as predominant areas for interaction and agree on gestures for tasks involving continuous input, such as document navigation. These results guide the design of several Hand-to-Face navigation techniques and reveal that gestures performed on the cheek are more efficient and less tiring than interactions directly on the HWD. Initial results on the social acceptability of Hand-to-Face input allow us to further refine our design choices, and reveal unforeseen results: some gestures are considered culturally inappropriate and gender plays a role in selection of specific Hand-to-Face interactions. From our overall results, we provide a set of guidelines for developing effective Hand-to-Face interaction techniques.

## Author Keywords

HMD; HWD; Head-worn display; mobile interfaces; body interaction; input techniques

## ACM Classification Keywords

H.5.2. Information interfaces and presentation: Interaction.

## INTRODUCTION

Head-mounted devices are becoming available for widespread, daily use through lighter form factors and with transparent displays. We refer to these modern accessories as head-worn displays (HWDs). As consumers may soon get affordable access to HWDs [6, 30, 31], ways in which they interact with content on such devices is being actively investigated [17].

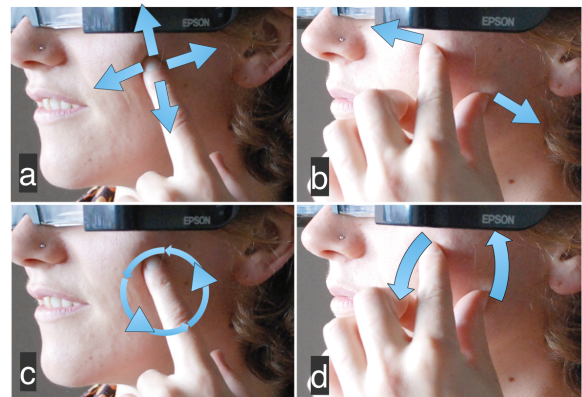
Currently, HWDs provide onboard microphones and small capacitive sensors for user input. Voice recognition is useful for command-based tasks such as for search queries but has limited use in certain settings (i.e. noisy environments). The capacitive surface on the temple of

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

CHI 2014, April 26– May 01, 2014, Toronto, ON, Canada.

Copyright 2014 ACM 978-1-4503-2473-1/14/04...\$15.00.

HWDs presents a viable on-device method for input, but it has limited input space. Other less self-contained options, such as a wearable device or an auxiliary smartphone, can also allow for input [21,26]. However, these require carrying and may be occluded by the HWD content. Natural user interfaces [29] can overcome the above limitations. However, mid-air input [2,14] suffers from the lack of tactile feedback and on-body gestures [8, 9] such as making contact with the arm skin [8], are often coupled with on-body projection for output.



**Figure 1. Hand-To-Face input for navigation. a) Panning, b) Pinch zooming, c) Cyclo zooming, d) Rotation zooming. Our studies show that Cyclo was not socially acceptable while Rotation was not efficient.**

We propose *hand-to-face input* as a novel, alternative method for interacting with HWDs. We define hand-to-face input as any gesture that involves touching, rubbing, scratching or caressing the face. This approach is especially well-suited for interaction with HWDs for many compelling reasons: (i) the face is often touched [18, 20] making it a promising area for subtle interactions; (ii) it offers a relatively large surface area for interaction, but not normally clothed as are other areas; (iii) it facilitates eyes-free, single-handed input, which can be invaluable in mobile settings (e.g. riding a bike, holding on in a bus); and (iv) is in close proximity to the HWD, making it likely to accommodate device-borne sensors and creating a natural extension of the device temple.

We first explore the design space of hand-to-face input by eliciting from users the range of gestures for various mobile tasks, such as navigation and action selection. Our study

participants generally identified the cheeks and forehead as good surfaces for gestures. Based on these results, we designed hand-to-face navigation techniques (Figure 1) and found these to be most effective and least tiresome when carried out on the cheek. Given this non-ordinary form of interaction, we also examined whether hand-to-face interaction was perceived to be socially suitable. In most cases, participants found this form of input acceptable.

Our contributions include 1) an elicitation of potential hand-to-face gestures for mobile tasks with HWDs; 2) a design of hand-to-face input techniques for document navigation; and, 3) a validation of the suitability of such interactions for use in public settings.

#### STATE OF THE ART

Our research is inspired by recent results on interaction methods for wearable displays, in particular work on mid-air gestures and on-body input. We also discuss the implications of casually touching the face in daily settings.

#### Mid-air gestures

Mid-air gestures have gained significant interest recently due to the development of readily available gesture tracking systems. Mid-air gestures, using one or two hands, often require additional equipment, for example on the shoes [2], or on the wrist [14] to capture user input. Other forms of gestures such as foot input [1] or head movement [28], are possible but would affect viewing stability on a HWD. For the most part, mid-air gestural input suffers from a lack of haptic feedback.

#### On-body interaction

A large body of work has considered appropriating the body as an interaction surface. On-body interaction offers an input surface, the human skin, with the advantage of leveraging human proprioception as an extra feedback mechanism. This can overcome some of the limitations with mid-air interactions. Much of this prior work has considered coupling on-body input with on-body projection using wearable pico-projectors, such as with Skininput [8], Brainy Hand [29] and OmniTouch [9]. These projection-based techniques are also adaptable to HWDs.

Gustafson et al. [7] investigated using palm-based imaginary interfaces, i.e. interfaces without any visual feedback. Their results show that tactile sensing on the palm allowed users to orient themselves more effectively than cues sensed by the pointing finger. Likewise for hand-to-face interaction, cues sensed by the facial skin could actually help orient the user. Recent work by Wagner et al. [32] proposed a body-centric design space to describe multi-surface and on-body interaction. This study however omitted exploring interactions on the face.

Only a few studies have explored the use of specific face areas, such as the ear [16] or the tongue [25] for input. We explore the range of different face areas, focusing on the design and social suitability of hand-to-face input, as well as its impact on effective document navigation.

#### Hand-to-face casual gestures

Hand to face contact is frequent and can lead to an average of 15.7 contacts/hour in work settings [20]. As contact with the mouth, eyes and/or nostrils can lead to the transmission of diseases, hand-to-face input should not include such health-sensitive areas. More recently, researchers studied hand-over-face postures in communication involving different emotions (frustration, surprise...): they found that 13% of the gestures were on the upper vs. 89% on the lower regions of the face (some covered both regions) [18]. These results further suggest that implicit face contact can be subtle, frequent and natural. We seek to make face contact explicit for common mobile tasks.

#### DESIGNING HAND-TO-FACE INTERACTIONS

Factors influencing the design of hand-to-face interaction include: facial area of interaction, number of fingers used, gesture style, physical demand, social acceptability and distinguishing explicit from implicit input

#### Area of interaction

Prior research suggests which facial areas are commonly used (lower regions [18]), as well as those to avoid (mouth or eyes [20]). In this first exploration, we investigate the range of gestures possible on defined areas of the frontal part of the face, such as the chin, cheek or forehead. We explore the feasibility of input on areas of the HWD such as the temple, hinge, bridge, frame and glass. We identify which regions are more effective and less prone to fatigue.

#### Number of fingers

The use of multiple fingers can affect hand-to-face input as certain regions on the face or the wearable display have limited surface area. However, using multiple fingers can be natural for mobile tasks such as zooming in/out of a map. HWDs offer a rather limited surface (Figure 2), which can make multi-finger interaction difficult. The face seems more appropriate for this type of interaction due to its large dimensions. We study this aspect in our second experiment.

#### 1D vs. 2D Gestures

The temple of several contending HWDs consist of a long, narrow surface (Figure 2). As a result, 1D gestures, such as flicking are better suited than 2D ones, such as pinching. Moreover, we expect large areas of the face such as the cheek to be better at supporting 2D gestures, such as panning on a map. We elicit from study participants potential gestures for 1D and 2D tasks.

#### Physical demand

Lifting the arm to touch parts of the face, such as the neck or the chin, can be physically demanding. However, touching the upper areas of the face, such as the hair or the forehead, requires more effort as the user has to perform larger arm movements. We investigate the physical demand of interacting with different areas of the face in our studies.

#### Social acceptability

Since casual hand-to-face gestures are common, they can be perceived as a non-intrusive and subtle input method.



Conversely, users may be reluctant to perform on-body interaction, particularly on areas that garner social attention, such as the face. It could also interfere with users' facial cosmetic products. To address these issues, we examine the social acceptability of such input after identifying the most effective hand-to-face techniques.



**Figure 2. HWDs ordered by increasing temple size: Telepathy [30], Google Glass [6], Vuzix M100 [31] and Moverio [4].**

### Implicit and Explicit Gestures

Given the frequency at which face contact occurs, hand-to-face gestures need to be invoked explicitly to avoid the Midas touch problem [12], i.e. inadvertently issuing a command to the device. This can be avoided using a delimiter at the start of the input, such as through voice (e.g. "HWD-navigate"), by pressing a button on the display or by pressing harder for explicit input. This concern is not unique to face input, as interacting with HWDs also requires that the user enter a specific mode. We leave out the examination of this factor in this first exploration.

### GUESSABILITY STUDY

To explore the breadth of potential hand-to-face gestures and their mapping to interactive tasks, we elicited user input through a guessability study [1,24,27,33]. For a set of common mobile tasks, as in [24], we asked participants to suggest suitable gestures on the face (above the neck) and on the HWD.

### Overview and rationale

For exploring potentially rich and vast gesture sets, user-elicitation or guessability studies have shown favourable results [1, 24, 27, 33]. Wobbrock et al. [33] found that eliciting gestures from users resulted in over 40% more gestures than if asked by expert designers. This motivated the use of such an approach to identify gestures for a multi-touch tabletop [33], for mobile motion gestures [24], and for foot interaction [1]. As in these previous works, we focus primarily on human capabilities. Consequently, we put aside any recognizer issues and asked users to perform gestures at their will without worrying about the underlying sensing technology. We asked participants to include gestures on the entire face, i.e. any region on or above the neck. This allows for a larger set of potential gestures. We also assessed users' preference for interacting with either the face or areas of the HWD, for each of the tasks.

### Mobile tasks considered

Most HWDs possess similar applications to those currently available on smartphones. For example, demonstrations of Google Glass [6] show users taking photos and videos with a camera, navigating a map, texting, selecting phone calls and even browsing the internet [6]. In some cases, HWDs

are designed as an auxiliary companion to mobile devices, such as the Vuzix M100, a wearable wireless display for Android smartphones [31]. For these reasons in this first exploration, we elicit gestures for common mobile tasks. Inspired by Ruiz et al. [24] we divide tasks into two categories: action (tasks 1 to 9) and navigation (tasks 10 to 22). Action includes tasks such as answering a call, taking a picture or selection. Navigation includes tasks such as opening the home screen, discrete 2D navigation or continuous 2D panning and zooming.

### Participants

Fourteen (6 female) volunteers participated, of 27.8 years on average. We did not filter out participants who had experience with natural user interfaces. As a result, of our participants 7 were familiar with mid-air gestures, 1 had seen on-body interaction (in a video) and one had used a HWD. Users were all unfamiliar with our proposed hand-to-face input. We rewarded them with a gift card.

### Apparatus

We mocked up a HWD using a plastic set of sunglasses (Fig. 3) with a larger craft board temple as the interactive surface. To define the size of the interactive temple we calculated the average width of that used on five commercial HWDs. The size of this interactive temple was 11×2 cms. The weight of the added piece was negligible.



**Figure 3. Mockup of the HWD used in the study.**

### Procedure

We familiarized our participants with HWD capabilities by showing them a demonstration video of Google Glass [6]. The video shows a first-person view of the display features of the device (such as image or map browsing) without showing the interaction modalities. We showed this particular video to not bias participants to current HWD interactions. We asked participants to design gestures for a HWD for each of the given tasks. Participants had to generate one gesture for hand-to-face input and another for use on any part of the HWD surface. Unconstrained by any time limit, users wrote down the most suitable gesture for each task on a schematic front and profile face image (similar to Figure 4). After sketching the gestures for a set of tasks, we asked them to perform their gestures, which we video recorded. Finally, we asked users to select and justify, for each task, which gesture they preferred, i.e. the one on the HWD or on the face. This took about 60 minutes in total. Participants wore the mock HWD for both conditions.

### Collected data

Every user generated one sheet with a sketch of the gesture on the face and on the display for every task. We video recorded the entire experiment and users' verbal comments.

### RESULTS

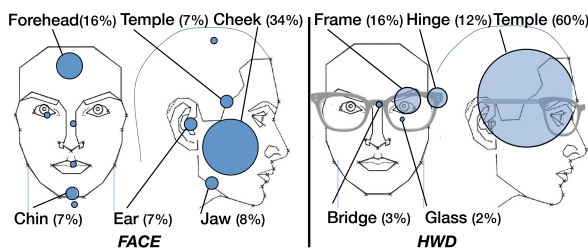
We collected gestures for 22 tasks × 14 users × 2 conditions (face and HWD) for a total for 616 gestures. The complete

gesture set is available online (<http://hci.cs.umanitoba.ca/face-input>).

**Comparing interaction for the face and the HWD**

*Areas of interaction*

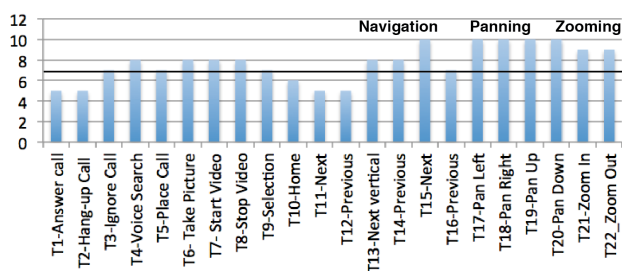
On the face, participants produced gestures for a total of 11 different areas, such as the cheek, forehead, ear, chin or jaw. The results reveal a distribution (Figure 4) with gestures concentrated on the cheek (34%) and then on the forehead (16%). Other areas saw an equal distribution of gestures: jaw (8.7%), ear (7.7%), temple (7.4%), and chin (7.4%). Areas such as the eyes, nose, lips, neck and hair were least used. On the HWD, participants used 5 different interaction areas: temple, hinge, frame, bridge and glass. Most of the gestures (60%) were situated on the temple.



**Figure 4.** Main areas identified by participants as suitable for input on the face (left) and on the HWD (right). The circles' size is proportional to the percentage of gestures (in brackets).

*User preference*

Overall users preferred interaction on the face for the navigation tasks (T10-T22), while opinions were mixed for the actions tasks (T1-T9) (Figure 5). Users particularly preferred using the face for panning (10/14 preferred the face) and zooming (9/14). Users indicated that "the face provides a larger area" [P2], which is perceived as a benefit for panning and zooming. This is particularly true when using the cheek, since it is "the best part to interact with the face" [P9] and it is "like a touchpad" [P1].



**Figure 5.** Number of users preferring face to HWD for each task. We emphasize the number 7 (50% of the 14 users).

*Benefits and limitations of face and HWD input*

We summarize users' perceived benefits and limitations for interacting on the face and the HWD in Table 1. These results are influenced by the form factor of the mockup used in the study, whose general characteristics (weight, inertia or grip on head) differ from popular HWDs. For instance, physical discomfort could be improved in a

commercial HWD, although continuous input on the temple may still be irritating. Interestingly, two female users preferred the HWD for all tasks; one suggested that face input would interfere with facial cosmetics and the other felt that it would be socially awkward. We examine these issues in our final study.

HWD	Face
+ <i>Metaphor</i> : "for some tasks, the HWD make sense, such as taking a picture, it's like a camera" [P2]	+ <i>Large surface</i> : "the face has a larger surface" [P14]
+ <i>Rapid activation</i> : "HWD is useful for single taps and for smaller gestures" [P5]	+ <i>Natural interaction</i> : "the face makes you feel more natural" [P2]
+ <i>Physical gestures</i> : "holding and moving the HWD is nice" [P9]	+ <i>Symbolic gestures</i> : "mapping gestures is easy, for instance tapping near the ear" [P3],
- <i>Limited space</i> : "HWDs have limited space, specially for panning and zooming" [P9]	+ <i>Memorization</i> : "Face is more meaningful, easier to remember" [P8]
- <i>Physical demand</i> : "HWDs are higher up on the face" [P8]	+ <i>Flat surfaces</i> : "cheek is like a touchpad" [P1], "forehead is flat and colinear to the field of view of the HWDs" [P4]
- <i>Instability</i> : "HWDs move when touched" [P7], "HWDs shake" [P6], "touching the HWD may move the camera" [P10]	- <i>Makeup</i> : "did not want to touch the face due to makeup" [P11]
- <i>Occluded vision</i> : "occlusion of field of view when using the HWDs" [P4]	- <i>Dirt</i> : "finger skin oil will remain on forehead" [P12]
- <i>Physical discomfort</i> : "tapping the HWD hurts my nose" [P9].	- <i>Sensitive surface</i> : "the face is too sensitive, touching often may hurt" [P3]

**Table 1.** User comments on benefits (+) and limitations (-) of HWD and Face interaction.

**Participant agreement on hand-to-face gestures**

Participants produced a large variety of hand-to-face gestures (Figure 6) which we analyze in two steps. We use prior methods [33] for the first analysis while the second one is based on a taxonomy we infer from users' responses.



**Figure 6.** Examples of gestures identified during the study by participants. a) Touching the mouth to place a call; b) touching the hair to ignore a call; c) rotating two fingers, or d) touching the chin to do a voice search; e) swiping the forehead to start video recording; f) pinching the frame to take a picture (on the HWD); g) tapping on the HWD bridge to go to the home screen; and h) grabbing the ear to ignore a call.

*First analysis*

We analysed the agreement between participants for the set of gestures produced for each task using Wobbrock's approach [33]. The agreement value ranges between 0 (no

agreement) and 1 (total agreement) and indicates whether users agreed on using a specific gesture for a given task. We group gestures which are of the same type (swipe, tap, etc.) and occur in the same area (cheek, chin, etc.).

The mean value for the agreement score is 0.14 (SD 0.06), with 36% of the tasks having an agreement value higher than 0.2% (Figure 7). While this score seems low, it is on par with that from other previous guessability studies [1]. The gestures for panning (0.28) showed the higher agreement score. The two main panning gestures were “cheek swipe” (Fig. 1-a) and “forehead swipe” (same but on forehead). The gestures for zooming (the others preferred by users) have an agreement score of 0.13.

This low agreement score is mainly due to the variety of areas used for zooming (ear, hair, nose, neck, mouth, jaw, forehead, chin and cheek). Swiping the cheek is considered different than swiping the forehead. We propose a taxonomy to describe hand-to-face gesture properties to evaluate gesture similarity in our second analysis

#### Hand-to-face gestures taxonomy

Through an open coding analysis, we infer a hand-to-face gesture taxonomy from the results of the study above, from the state-of-the-art on implicit hand-to-face gestures [18] and from previous guessability studies on motion gestures [24]. While this taxonomy could include many properties, such as the number of hands, we describe only the properties sufficient to characterize our results. Our taxonomy includes five properties to describe the gesture mapping and its physical characteristics: **1) Nature:** the nature of the gesture can be a *metaphor* (scratching the face or covering the mouth, 56% of total), *physical* (grabbing the ear, 4%), *symbolic* (drawing a letter or a symbol such as a circle, 2%) or *abstract* (38%). **2) Temporal:** the gesture can be *discrete* (48%) or *continuous* (52%). **3) Pose:** the gesture can be a static (36%) or a dynamic pose (64%). **4) Number of fingers:** gestures can use one (79%), two (17%) or several fingers (4%). **5) Area:** we divide the face into the following areas: cheek, forehead, temple, ear, eye, chin, mouth, neck, jaw, hair and nose (see percentages on Fig. 4).

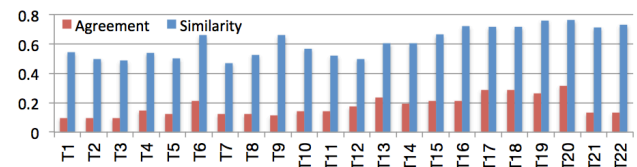
For instance, the gesture in Figure 6-a, consisting of touching the mouth to place a call, can be described as *<metaphor, discrete, static, two-fingers, mouth>*.

#### Second analysis

Based on the previous taxonomy, we define a formula to calculate the similarity score, which indicates whether different gestures share common properties. The similarity score  $S_t$  of a task  $t$  is the average of the agreement for every property  $P_i$  of our taxonomy, from the set of properties  $P_t$ . To calculate this value we use the formula of the agreement score [33]:  $G_i$  is the subset of gestures with identical value for the property  $P_i$  from the set of gestures  $G_t$ .

$$S_t = \frac{\sum_{P_i} \left( \sum_{G_i} \left( \frac{G_i}{G_t} \right)^2 \right)}{P_t}$$

The mean value for the overall similarity score (Figure 7) is 0.61 (SD 0.1) with panning (0.73) and zooming (0.72) having the highest agreement scores. The *Nature* of zooming gestures (0.86) is mainly based on the metaphor of pinching with two fingers. The *Pose* for zooming is always dynamic (1.0) and most subjects used two fingers for zooming (0.80).



**Figure 7. Agreement and similarity scores of hand-to-face gestures for each task. Panning (T17-T20) shows the highest overall scores while zooming (T21-T22) has a high similarity.**

#### Summary

The guessability study reveals that users find hand-to-face gestures appealing for navigation tasks (panning and zooming). We can describe panning gestures using our taxonomy as *<abstract, continuous, dynamic, one finger, cheek/forehead>* and zooming gestures as *<abstract, discrete/continuous, dynamic, two-fingers, cheek/chin/forehead>*. Given the high prevalence for document navigation in mobile contexts (a recent survey reveals Google Maps is the most used mobile application worldwide [5]), we further investigate the performance, fatigue and user preference of hand-to-face input for panning and zooming.

#### HAND-TO-FACE NAVIGATION STUDY 1: PANNING

In this study we evaluate the performance, physical effort and user preference of hand-to-face gestures for panning.

#### Participants

Twelve students (2 female) from a local university volunteered with an average of 25.9 years and were rewarded with a gift card.

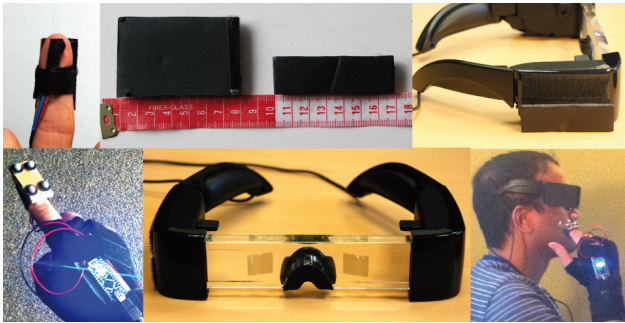
#### Apparatus

We used the Epson Moverio BT-100 a commercially available HWD. The device weighs 220g, has a resolution of 960×540 pixels and a refresh rate of 60Hz. The device measures 17.8×20.5×4.7cm and runs Android 2.2. We included two different temple sizes (Size factor). The Oversized temple corresponds to the actual device’s temple, which measures 7×5cm (larger than most HWDs to be released). The Regular temple is based on the average temple size used in our guessability study, measuring 7×2cm. We used craft board to mock up the temples.

We implemented face gestures using a Vicon T20 infrared optical tracking system with six cameras positioned around the user (front, front-right and right side at different heights). We placed infrared markers on the participant’s index finger (Figure 8). To detect skin contact, we used a proximity sensor connected to a micro-controller. The sensor set was connected through a USB cable to a desktop computer. To make the sensor set unobtrusive, we



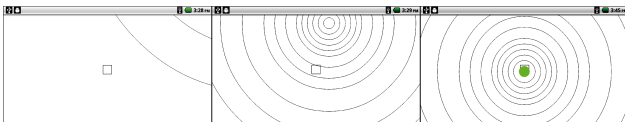
integrated it into a glove worn by the user. During our studies we had a negligible number of tracking errors (0.15% of all trials). The system had no perceivable latency: all input was merged to the same program and sent to the HWD through USB. Optical tracking ran at 690 Hz; contact sensor at 600 Hz; merged data were sent to the HWD at a measured rate of 142 Hz.



**Figure 8. Top: Sensor used to detect skin contact; Oversized and Regular temples measures (cms); Regular temple attached to HWD. Bottom: IR markers and microcontroller on hand; Epson Moverio HWD and subject wearing HWD.**

#### Task and mappings

Users were asked to pan until the target, represented by the smallest central circle in a set of concentric rings, is in the screen center (Figure 9). The target is placed in one of 6 different directions (2 vertical, 2 horizontal and 2 diagonals), at 3 screen widths distance. Participants used one of four different interactive surfaces, two on the face (cheek and forehead) and two on the display (Oversized and Regular temple). We include two different temple sizes in order to study the impact of its size on navigation.



**Figure 9. Visual feedback of the panning task displayed on the HWD from initial state (left) to target reached (right).**

We use three panning interactions: displacement-based (D), flick (F) and rate-based (R). With Pan-D, finger movement is directly mapped to the movement of the map. In Pan-F, the user flicks to pan, mimicking the iOS flick behaviour [10]. In Pan-R, the distance panned from the initial touch position is mapped to the finger velocity movement.

#### Design

The experiment used a 4×3×3 design with Area (Cheek, Forehead, Oversized temple, Regular temple), Gesture Mapping (Pan-D, Pan-F, Pan-R), and Direction (Horizontal, Vertical and Diagonal) as factors. The experiment is divided into 12 blocks, each corresponding to a specific area. Each block was repeated three times. Order of blocks is counterbalanced across participants. Trials in the block were grouped by mapping (i.e. all trials of flicking were performed together) always in the same order (Pan-D, then

Pan-F, then Pan-R). Every direction is performed two times per block. Every condition is repeated 6 times in total.

#### Procedure

Participants were allowed to get used to the apparatus and were given sufficient training, first on the desktop display, then on the HWD. The experiment was divided into blocks. The user starts a block by pressing a key. In each block, there is a 2 sec interval between trials. We informed users that they can take a break between blocks and ask them to be as fast as possible during the trials. Participants wore the HWD during the entire experiment.

#### Collected data

We logged all tracking data and measured time to complete the task from stimulus onset and first touch. We measured physical demand by using the Borg CR10 scale of perceived exertion, specifically adapted to physical demand [3,11]. We also measured user preference using a 5-point Likert scale to rate the four areas and the three techniques. We collected 216 trials per user (4 areas × 3 techniques × 3 directions × 6 repetitions) × 12 users = 2592 trials in total.

#### Results

##### Time Performance

Trial completion time (Figure 10-left) which was measured from first touch until target reached, we found a significant effect of Area ( $F_{3,33} = 12.7$ ,  $p < .0001$ ), Technique ( $F_{2,22} = 52.7$ ,  $p < .0001$ ) and Orientation ( $F_{2,22} = 13.9$ ,  $p < .0001$ ). There is an interaction effect between Area and Technique ( $F_{6,66} = 8.2$ ,  $p < .0001$ ) and between Technique and Orientation ( $F_{4,44} = 7.1$ ,  $p < .0001$ ). The interaction effect between Area and Technique is mostly due to the significant time taken to complete the task with the Pan-R (Rate) technique on the regular temple. In this particular case users' finger often accidentally slid out of the temple due to its small size, forcing users to clutch often.

For Area, a post-hoc test reveals that Cheek (mean 6.5s) is significantly faster than Forehead (mean 8.2s) and Regular Temple (mean 8.9s). We found no significant difference between Cheek and Oversized Temple (mean 7s). This can be expected as they have a similar effective surface: the cheek is generally larger but participants tended to use its central area. With Technique, a post-hoc test reveals that Rate (mean 9.8s) is significantly slower than Displacement (mean 6.6s) and Flick (mean 6.4s). While Rate was particularly slow for the Regular temple, it also showed worst performance on Cheek and Oversized temple. As expected, Horizontal Orientation (mean 6.5s) is significantly faster than Vertical (8.1s) and Diagonal (8.3s).

##### Fatigue and User preference

To analyze the results from the Borg questionnaire (Figure 10-right) we perform an Anova test (Borg being a scalar value). There is a significant effect of both Area ( $F_{3,33} = 9.2$ ,  $p < .0001$ ) and Technique ( $F_{2,22} = 6.9$ ,  $p = .005$ ) on effort. A one-way Anova reveals that Cheek and Oversized Temple are less tiring than Regular HWD Temple and Forehead.

While we expected this result for Forehead, which forces the user to raise her arm, the score for Regular Temple is surprising. One reason might be that the small surface requires more clutching, thus requiring the user to keep her arm suspended in air longer. The results of the Likert questionnaire reveal that users largely prefer the Oversized Temple (50% Strongly agree) and the Cheek (55% Agree). Forehead (60% Strongly disagree) and Regular temple (20% Strongly disagree and 40% disagree) were the least preferred areas by users.

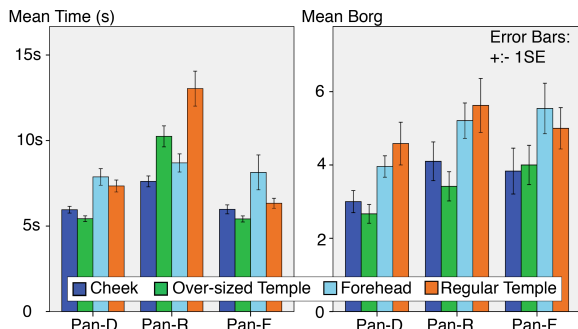


Figure 10. Mean time in s. (left) and Mean Borg value (right) for every panning technique and interaction area.

Summary

This first study on using hand-to-face gestures for panning revealed that the best facial area for input is the Cheek. The Forehead and the Regular HWD Temple not only showed worse performance, but also result in higher fatigue. Overall there was no difference between the Cheek and the Oversized temple, but both were favored over the Regular temple. The Oversized temple, however, is far larger than most HWDs, suggesting that the Cheek is a preferred interaction region.

HAND-TO-FACE NAVIGATION STUDY 2: ZOOMING

The goal of this study is to evaluate the performance of hand-to-face gestures, for full document navigation which includes zooming and panning. From the previous study we dismiss Forehead due to its low overall results.

Tasks and mappings

The task was the same as in the previous study, but with a much more distant target that required zooming to reach effectively. We selected three different zooming techniques, based on a combination of prior known methods and from the guessability study: Linear, Rotation and Cyclo. Linear zooming, by pinching with two fingers, is the classical touchscreen technique. Circular zooming with two fingers (using the angle of rotation) is based on the metaphor of adjusting an optical lens. Cyclo is a one finger zooming technique proposed by Malacria et al. [19] as a way to avoid clutching when zooming. It consists of doing a circular movement with one finger. The orientation of the rotation is mapped to the zoom direction (in or out). We use the same implementation as in [19].

Participants

We recruited 15 students (2 females) from a local university with an average of 26 years. 10 of them assisted Study 1.

Design

The experiment followed a 3x3 design with Area (Cheek, Oversized HWD temple, Regular HWD temple) and Zoom Technique (Linear, Rotation, Cyclo) as factors. We chose the Pan-D technique from the first study. We counterbalanced the Area factor, while the technique was presented by increasing level of difficulty.

Apparatus, procedure and collected data

We used the same apparatus as in the previous experiment, adding an extra sensor and IR marker to detect thumb touch and position. We follow the same procedure and measured time, physical demand and user preference as in Study 1.

Results

We collected 162 trials per user (3 Areas x 3 Techniques x 3 Directions x 6 Repetitions) x 12 users = 1944 trials in total.

Time Performance

We found a main effect of our two factors, Area ( $F_{2,22} = 3.7, p = .038$ ) and Technique ( $F_{2,22} = 5.5, p = .009$ ) for completion time (Figure 11). We found no interaction between Area and Technique. Post Hoc tests reveal that Cheek (mean 23.8s) is significantly faster than Regular Temple (mean 27.0s) and Oversized Temple (mean 26.7s). Concerning the Technique, 2 Finger Rotation (mean 29.7s) is significantly slower than 2 Finger Linear (mean 24.7s) and 1 Finger Cyclo (mean 23.1s). There is no significant difference between the last two techniques.

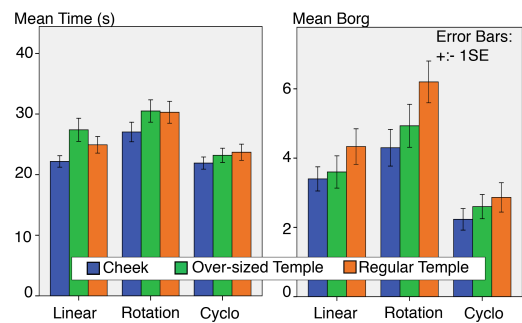


Figure 11. Mean time in seconds (left) and mean Borg value (right) for Technique and interaction Area.

Fatigue and User preference

The results of the Borg questionnaire reveal a main effect of both Area ( $F_{2,22} = 8.0, p = .002$ ) and Technique ( $F_{2,22} = 20.1, p < .001$ ) on fatigue. Post-hocs reveal that the Regular Temple was perceived as more fatiguing than the Cheek or the Oversized Temple (Figure 11). Concerning techniques, Cyclo was least fatiguing technique, followed by Linear and Rotation. The results of the Likert questionnaire are similar to that of study 1: overall, users preferred the Cheek (26% Strongly agree, 33% Agree) and the Oversized Temple (20% Strongly agree, 53% Agree) over the Regular Temple (26% strongly disagreed, 46% disagreed).

### Summary

These results extend further our exploration of Study 1, providing insight into hand-to-face interaction for document navigation. The main finding is that the Cheek is more efficient than both the Oversized and Regular temples for zooming. While the Oversized temple was efficient in Study 1 for one finger panning, it becomes inefficient with a two-finger gesture. Both the classical Pinch and the single-finger Cyclo are equally efficient in our study. However, to access an intermediate zoom level Cyclo could lose precision on the Regular temple, as smaller circles cause faster zooming, which may lead to overshooting.

### SOCIAL ACCEPTANCE OF HAND-TO-FACE GESTURES

While we demonstrated that hand-to-face interaction techniques improve navigation performance, we know little on how comfortable users would feel in different social contexts. We therefore carry a controlled exploration of the social acceptance of our hand-to-face gestures.

#### Questionnaire and gestures

Participants were asked to rank on a 5-point Likert scale the social acceptability of hand-to-face gestures. For every gesture and social context we asked the same question: "Would you be willing to perform this gesture in context X?" We grouped two social aspects into a single factor we call Social context. This factor encompasses audience (who you are with) and location (where you are), both inspired from previous studies [2, 23]. We focus on the panning and zooming gestures from the earlier studies. For panning, we include *Displacement* (Pan-D) and *Flick* (Pan-F). For zooming, we include *Linear* (L) and *Cyclo* (C).

#### Participants

We recruited 12 students (5 females) from a local university with an average age of 27 years. 7 of them wear optical glasses and 8 of them had tested a head-worn device. None of them participated in our previous studies. They were rewarded with a drink.

#### Design

We used 2×8×6 within-participant design with Device (Face or HWD), Gesture (Pan-D and Pan-F panning, L and C zooming) and Social context (Alone, Family or friend, Stranger, Home, Public space and Workspace) as factors.

#### Procedure

The study was performed in the presence of the interviewer. Participants watched a video of an actor performing panning and zooming gestures in front of a wall and then performed themselves the same gestures 3 times. The order of the videos was counterbalanced between participants. After completion of all the gestures, participants completed a questionnaire containing one question for each gesture and social context. This type of exploration based on video watching has already been used in previous studies [2]. Although it misses the ecological validity of an experiment in a real environment, it allows for a first exploration of a novel technique. For every participant we collected 48

responses to the Likert scale questions (2 × 8 gestures × 6 social contexts), written comments and oral feedback.

### Results

We collected 576 answers (48 answers × 12 subjects) for the 5-points Likert scale questionnaire (5= full agree).

#### Comparison of face and HWD acceptability

We used a Wilcoxon signed-ranks test to compare our main factors. We found no difference in social acceptability between Face and HWD ( $Z=-1.817$ ,  $p=.069$ ), but with a trend showing better acceptance for interaction on the HWD. The acceptance rate for both face and HWD gestures in any social context is above 50% (Figure 12). Results were rather homogeneous on the HWD, with a constant 10%-12% of disagreement for all social contexts except in front of strangers, where this value is 18%. We found more differences on the Face, with no disagreement when at Home or Alone, but with 31% disagreement in Public places and 25% in front of strangers. Comments from participants also show that most of them don't mind using the face: "I don't think it would disturb me to do the gesture either on the skin or on the temple." [P6]. One female participant indicated the problem of dirty hands on the face: "the face can be affected when perspiring" [P7].

#### Hand-to-face gestures acceptability

A Friedman test reveals a significant difference in social acceptability among techniques ( $p<.001$ ). We used a Wilcoxon signed-rank test for our analyses. We found a significant difference between the panning techniques ( $Z=-2.06$ ,  $p=.039$ ), Pan-D being better perceived than Flick. We also found a difference between the zooming techniques ( $Z=-6.2$ ,  $p<.001$ ), L being better perceived than C. Participants commented that Cyclo might be perceived as insulting, as it could signal that "you are crazy" [P4] in many cultures. This gesture seemed also more visible: "I feel all the gestures are quite subtle except Cyclo which might attract attention" [P6].

### DISCUSSION

#### Extending HWD through Hand-To-Face Interaction

Interacting with HWDs via the device temple presents a number of limitations, such as the small surface, movement due to contact and user fatigue. While Face interaction is not meant to replace HWD interaction, particularly for rapid selection, it is a promising alternative for extended tasks such as navigation. In our experiments we discovered that face interaction offers the benefit of faster panning + zooming as well as lower fatigue than the HWD.

#### Hand-To-Face gestures acceptability

While our social acceptability study is a laboratory exploration that should be expanded to the real environment, it indicates that hand-to-face interaction could be acceptable in different social contexts. Users find calm gestures such as displacement or pinching to be more socially acceptable than brisk ones such as Flick. Unexpectedly, the Cyclo gesture which ranked among the

most efficient from study 2 was seen as not acceptable. Users also identified some potential resistance to using Hand-to-Face input, due to practicalities such as facial cosmetics. One alternative to explore would be mid-air gestures near, but not touching the face. However, mid-air gestures need to be carefully designed to include proper delimiters, to not appear socially unacceptable, and to be evaluated against on-face input as a touch-less solution for addressing the issue of facial cosmetics.

#### **Hand-To-Face detection**

We explored hand-to-face gestures for HWDs without emphasizing the technology that would ultimately support this style of input. Several options exist, such as a camera mounted on the HWD [15], body-implanted sensors [13] or instruments worn on the finger [34]. Ideally such instruments should consider contactless hand-to-face input, as a means to avoid the limitations discussed above.

#### **Casual and explicit hand-to-face interaction**

Hand-to-face gesture recognition will need to consider the Midas-touch problem [12], i.e. how to differentiate casual and explicit hand-to-face gestures. Two obvious solutions include touching the HWD to initiate the face detection or using a voice command. Another solution is to use gestures that are very different from casual ones. While casual gestures are rather static and use the whole hand [18], gestures such as pinching could be easier to differentiate. Further investigation is needed to explore the possibilities.

#### **Other applications for hand-to-face interaction**

While hand-to-face interaction appears to be particularly well suited for HWDs, other contexts may benefit from hand-to-face interaction. One of the main benefits of hand-to-face interaction is to allow interaction with one hand when no input device is available. It could thus be used in mobile situations, such as on a bike to control a music player or when the other hand is busy holding a handle bar on a bus. Further work is needed to assess the value of hand-to-face interaction in eyes-free situations.

#### **Lessons learned**

This first exploration allows us to sum up a number of prescriptive design recommendations. Gestures should be performed on the cheek or on the lower areas of the face (chin, jaw), which are less tiresome than the Forehead. While the cheek has demonstrated value for navigating a document, other areas of the face can be used to enhance the symbolism and memorization of certain tasks. For instance, touching the chin can be used to activate a command related to speech, such as making a call. Hand-to-face gestures should be calm, since obvious or vigorous gestures can be less socially acceptable.

#### **Gesture alternatives**

Prior work on user-generated gestures, indicated that designers came up with only 60% of the gestures produced by end-users [33]. This gives some evidence that users may, in some instances, generate a larger set of alternatives than a few designers. Some gestures may have been missed

though, such as using pressure for zooming. Our final gestures resemble known multi-touch gestures as we had to select a set that was compatible with our baseline, the HWD temple: given its limited surface area, our choice of gestures was limited. In future, we intend on exploring other interactions without the baseline restrictions.

#### **Limitations and future work**

Our exploration was limited by the apparatus used in our experiments, by the participants profile and by the context of our social acceptability study. The overall apparatus, including the large HWD and the glove, was bulky and heavy, although it allowed us to demonstrate the benefits of hand-to-face interaction. While these benefits should be further noticed with lighter form factors (due to limited temple area), further evaluation is needed. Our comparison focused mainly on the impact of the temple's size. However other aspects such as the HWD's grip on the user's head need to also be considered when characterizing interaction with a HWD. In particular, the results of our guessability study apply best to HWDs with a similar form factor as our mock-up prototype. In future, we will replace the heavy HWD with an improved, lighter version. We also plan to integrate the sensing technology on the HWD itself, for instance using a camera fixed on its frame (similar to [15]).

Our participants were mainly local students. The findings of our work should be verified with a more diverse population, particularly for better understanding guessability and social acceptability issues. Our social acceptability study was based on video browsing gestures. Future work should test the acceptability of such interactions in real settings, such as in [22]. Factors such as gender and cultural ethnicity also need to be considered for improved Hand-to-Face input. For instance we can compare users' perceived acceptability between Asian and Western populations as in [22].

#### **CONCLUSION**

We presented an exploration of hand-to-face gestures, a novel type of on-body interaction especially well-suited for HWDs. We first described the results of a guessability study that shows that for navigation tasks such as panning and zooming participants prefer using hand-to-face gestures rather than gestures on the HWD. We calculated a similarity score among these gestures that demonstrates that participants converged to similar hand-to-face gestures for panning and zooming. In a first study we explore different areas and techniques for panning. We found that the cheek is the most promising area on the face, being larger, more efficient, less tiring and preferred to a regular temple on the HWD. In a second study we investigate different techniques for zooming and show that using the cheek is more efficient than using the temple of the HWD. Finally, we enquire into the social acceptability of these gestures and show that gestures on the face could be as acceptable as on the HWD.

#### **REFERENCES**

1. Alexander, J., Han, T., Judd, W., Irani, P. Subramanian, S. 2012. Putting your best foot forward: investigating

- real-world mappings for foot-based gestures. In *Proc. of CHI '12*. ACM, 1229-1238.
2. Bailly, G., Müller, J., Rohs, M., Wigdor, D., Kratz, S. 2012. ShoeSense: a new perspective on gestural interaction and wearable applications. In *Proc. of CHI '12*. ACM, 1239-1248.
  3. Borg, G. Borg's Perceived Exertion and Pain Scales. *Human Kinetics* (1998), viii 104pp.
  4. Epson Moverio, <http://www.epson.com/moverio>
  5. GlobalWebIndex, <https://www.globalwebindex.net/Top-global-smartphone-apps>
  6. Google Glass, Project Glass One Day video, <http://www.google.com/glass>
  7. Gustafson, S., Rabe, B., and Baudisch, P.. 2013. Understanding palm-based imaginary interfaces. In *Proc. of CHI '13*. ACM, 889-898.
  8. Harrison, C., Tan, D., and Morris, D. 2010. Skinput: appropriating the body as an input surface. In *Proc. of CHI '10*. ACM, 453-462.
  9. Harrison, C., Benko, H. and Wilson, A. 2011. OmniTouch: wearable multitouch interaction everywhere. In *Proc of UIST '11*. ACM, 441-450.
  10. Hidayat, A. Flick list with its momentum scrolling and deceleration. <http://ariya.ofilabs.com/>
  11. Hincapie-Ramos, J-D., Guo, X., Moghadasian, P. and Irani, P. 2014. Consumed Endurance: a metric to quantify arm fatigue of mid-air interactions. In *Proc. of CHI '14*. ACM, to appear.
  12. Hinckley, K., Wigdor, D., Input Technologies and Techniques. Chap. 9 in *The HCI Handbook*, 3rd Edition, Taylor & Francis.
  13. Holz, C. Grossman, T., Frizmaurice, G and Agur, A. 2012. Implanted user interfaces. In *Proc. of CHI '12*. ACM, 503-512.
  14. Kim, D. Hilliges, O., Izadi, S., Butler, A., Chen, J., Oikonomidis, I. and Olivier, P.. 2012. Digits: freehand 3D interactions anywhere using a wrist-worn gloveless sensor. In *Proc. of UIST '12*. ACM, 167-176.
  15. Kimura, S., Fukuomoto, M., and Horikoshi, T. 2013. Eyeglass-based hands-free videophone. In *Proc. of ISWC '13*. ACM, 117-124.
  16. Lissermann, R. Huber, J., Hadjakos, A. and Mühlhäuser, M. 2013. EarPut: augmenting behind-the-ear devices for ear-based interaction. In *CHI '13 EA*. ACM, 1323-28.
  17. Lucero, A. Lyons, K., Vetek, A., Järvenpää, T., White, S. and Salmimaa, M. 2013. Exploring the interaction design space for interactive glasses. In *CHI '13 EA*. ACM, 1341-1346.
  18. Mahmoud, M. and Robinson, P. 2011. Interpreting Hand-Over-Face Gestures. In *Proc. of Affective Computing*, Springer, 248-255.
  19. Malacria, S., Lecolinet, E. and Guiard, Y. 2010. Clutch-free panning and integrated pan-zoom control on touch-sensitive surfaces: the cyclostar approach. In *Proc. of CHI '10*. ACM, 2615-2624.
  20. Nicas, M. and Best, D. 2008. A Study Quantifying the Hand-to-Face Contact Rate and Its Potential Application to Predicting Respiratory Tract Infection. In *J. Occup. Env. Hyg.*, 347-352.
  21. Perrault, S., Lecolinet, E., Eagan, J., Guiard, Y. 2013. Watchit: simple gestures and eyes-free interaction for wristwatches and bracelets. In *Proc. of CHI '13*. ACM, 1451-1460.
  22. Profita, H., Clawson, J., Gilliland, S., Zeagler, C., Starner, T., Budd, J. and Do, E. 2013. Don't mind me touching my wrist: a case study of interacting with on-body technology in public. In *Proc. of ISWC '13*. ACM, 89-96.
  23. Rico, J. and Brewster, S. 2010. Usable gestures for mobile interfaces: evaluating social acceptability. In *Proc. of CHI '10*. ACM, 887-896.
  24. Ruiz, J, Li, Y., Lank, E. 2011. User-defined motion gestures for mobile interaction. In *Proc. of CHI '11*. ACM, 197-206.
  25. Saponas, T., Kelly, D., Parviz, B. and Tan, D. 2009. Optically sensing tongue gestures for computer input. In *Proc. of UIST '09*. ACM, 177-180.
  26. Serrano, M., Lecolinet, E., and Guiard, Y. 2013. Bezel-Tap gestures: quick activation of commands from sleep mode on tablets. In *Proc. of CHI '13*. ACM, 3027-3036.
  27. Seyed, T., Burns, C., Costa, M, Maurer, F. and Tang, A. 2012. Eliciting usable gestures for multi-display environments. In *Proc. of ITS '12*. ACM, 41-50.
  28. Špakov, O. and Majaranta, P. 2012. Enhanced gaze interaction using simple head gestures. In *Proc. of UbiComp '12*. ACM, 705-710.
  29. Tamaki, E. Miyak, T. and Rekimoto, J. 2010. BrainyHand: a wearable computing device without HMD and its interaction techniques. In *Proc. of AVI '10*. ACM, 387-388.
  30. Telepathy, <http://tele-pathy.org>
  31. Vuzix M100, <http://www.vuzix.com>
  32. Wagner, J. Nancel, M., Gustafson, S., Huot, S. and Mackay, W. 2013. Body-centric design space for multi-surface interaction. In *Proc. of CHI '13*. ACM, 1299-1308.
  33. Wobbrock, J., Ringel Morris, M. and Wilson, A. 2009. User-defined gestures for surface computing. In *Proc. of CHI '09*. ACM, 1083-1092.
  34. Yang, X-D. Grossman, T., Wigdor, D. and Fitzmaurice, G. 2012. Magic finger: always-available input through finger instrumentation. In *Proc. of UIST '12*. ACM, 147-156.



# Identifying Suitable Projection Parameters and Display Configurations for Mobile True-3D Displays

Marcos Serrano<sup>1,2</sup>, Dale Hildebrandt<sup>1</sup>, Sriram Subramanian<sup>3</sup>, Pourang Irani<sup>1</sup>

<sup>1</sup>University of Manitoba  
Winnipeg, MB, R3T 2N2, Canada  
{umhilde6,irani}@cs.umanitoba.ca

<sup>2</sup>IRIT - University of Toulouse  
Toulouse, 31062, France  
marcos.serrano@irit.fr

<sup>3</sup>University of Bristol  
Woodland Road, BS81UB UK  
sriram@cs.bris.ac.uk

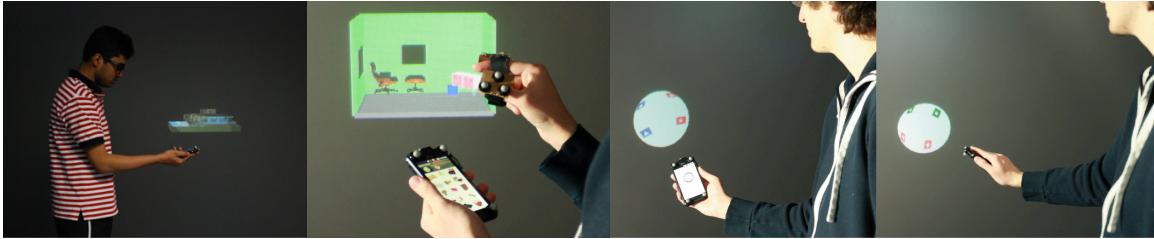


Figure 1. a) Illustration of a true-3D visualization on a mobile device. b) We implemented a 3D interior design application in our true-3D mobile prototype. c-d) Subject doing visual search during the first experiment.

## ABSTRACT

We present a two-part exploration on mobile true-3D displays, i.e. displaying volumetric 3D content in mid-air. We first identify and study the parameters of a mobile true-3D projection, in terms of the projection's distance to the phone, angle to the phone, display volume and position within the display. We identify suitable parameters and constraints, which we propose as requirements for developing mobile true-3D systems. We build on the first outcomes to explore methods for coordinating the display configurations of the mobile true-3D setup. We explore the resulting design space through two applications: 3D map navigation and 3D interior design. We discuss the implications of our results for the future design of mobile true-3D displays.

## Author Keywords

Mobile True-3D, intangible displays, mid-air pointing, displays.

## ACM Classification Keywords

H.5.2. Information interfaces and presentation: Interaction.

## INTRODUCTION

Benefiting from 3D on mobile devices is pertinent, beyond video games, for mobile scenarios such as 3D interior design or 3D map exploration. Mobile devices (smartphones and gaming consoles) incorporate at present

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MobileHCI '14, September 23 - 26 2014, Toronto, ON, Canada

Copyright 2014 ACM 978-1-4503-3004-6/14/09 \$15.00.

<http://dx.doi.org/10.1145/2628363.2628375>

auto-stereoscopic displays as a first step toward achieving true-3D content [21]. True-3D, i.e. displaying volumetric 3D content in mid-air, can be accomplished through different means: stereoscopic displays with head-tracking, optical illusions, moving parts or augmented reality glasses [14,15,17,20,27]. Integrating true-3D on mobile devices, apart from facing hardware challenges, presents a number of unresolved human factors questions concerning its use.

We possess limited knowledge on both the numerous constraints imposed on viewing and interacting with mobile true-3D interfaces and the usage scenarios suitable for such displays. These include knowing about the ideal angular position, size and distance of the volumetric projection, relative to the mobile device, the projection limits on visual search and direct interaction and how to coordinate the mobile and true-3D views. Answers to such questions will equip manufacturers and designers with tools to begin exploring a range of technologies that can be tailored for true-3D input on mobile devices.

We study the various factors that can potentially influence the effective deployment of true-3D on mobile devices in an emulated environment (Figure 1). We focus on mobile-mounted 3D projection, which means that the true-3D projection moves with the mobile device as if both were attached. We systematically address the above listed fundamental questions for this form of Mobile Multi-Display Environment (MMDE).

Our contributions include: 1) an identification of suitable display parameters and constraints for true-3D mobile display; 2) an exploration of the coordinated display configurations for the 2D mobile and true-3D displays; and 3) the application of these configurations to the design of two proof-of-concept applications, a 3D map and a 3D interior design application (Figure 1).

## STATE OF THE ART

### True-3D

True-3D refers to any 3D digital display capable of producing mid-air, full-depth-cue (or volumetric), multi-angle and/or multi-user images without the need for user instrumentation [15,17].

Recent work studying the value of ‘true’ 3D displays has relied on emulation for recreating this type of display [1,15,20]. To produce intangible and mid-air images, most systems use optical illusions. Vermeer is an interactive 360° viewable 3D display based on the optical illusion of using two parabolic mirrors [2]. HoloDesk is a system that exploits a see through display with kinetic motion for creating a ‘true’ 3D effect [15]. Virtual Reality (VR) has mostly emulated the ‘true’ 3D experience using stereoscopic projections, an approach we adopt to evaluate the display projection parameters for true-3D mobile displays. Head-mounted augmented reality systems have largely evolved and can facilitate 3D stereo viewing capabilities, with lighter glasses and see-through displays, such as the Vuzix Star 1200 [28]. Depth cues can be produced using auto-stereoscopic displays, which have appeared on mobile and handheld gaming devices. Such screens can be viewed by only one user and have a narrow viewing angle. Coupling such displays with head-tracking allows various angles of view, as in Fuwa-Vision [20], an auto-stereoscopic display presenting images in mid-air.

Aside from good ‘emulations’, the last few years witnessed ‘true’ 3D displays developing at an accelerated pace. The HelioDisplay generates a mid-air display using floating particles as a projection surface [14], but is limited to projection on a two-dimensional plane. Researchers more recently have demonstrated the use of laser plasma technology to produce 3D-objects with light dot arrays in space [17]. This method does not allow for direct interaction with the true-3D display.

While such advances bring us closer to experiencing true-3D, we possess limited knowledge on how to design for such displays to co-exist with other forms of displays, such as mobile devices. To our knowledge, this work is the first exploration of the human factors considerations for a true-3D mobile multi-display environment.

### Mobile Multi-Display Environments

Previous works have classified Mobile Multi-Display Environments (MMDE) on a continuum from being partially mobile (one mobile phone used with fixed displays [1]) to fully mobile systems [5]. Fully mobile can be further divided into multi-device-single-display and single-device-multi-display [5] MMDEs. We focus our explorations on the category of fully mobile, single-device and multi-display systems.

Foldable mobile devices such as the Nintendo 3DS [21] integrate a second screen to extend its display capabilities. Using an additional screen has the disadvantage of

increasing the overall size and weight of the device. Pico-projectors alleviate the above problem by allowing use of a secondary display as needed [25]. Their main limitations are the need for a flat projection surface and the large distance between the projected display and the main display. While the latter drawback can be resolved by finding a closer projection surface [5] or by using steerable projectors [6], the former still limits where such systems can be used. The use of a true-3D display could overcome this limitation.

### Interaction with mobile 3D and true-3D interfaces

Recent work demonstrates methods for interacting with 3D content using different smartphone interaction capabilities such as the touchscreen, the accelerometer or the camera [29,18]. Song et al. presented a set of techniques for visual exploration of volumetric data [29]. PalmSpace proposes the use of continuous around-device gestures for 3D rotation tasks [18]. Sitcky tools are a set of multi-touch techniques for 6DOF manipulation on flat tabletops [13]. However it is unobvious whether such solutions transfer well to handheld mobile true-3D, where the 3D content moves with the device.

In the case of true-3D displays, we can borrow knowledge from the literature on 3D user interfaces [11,19]. However, with actual systems very few results exist. Chan et al. investigated direct-touch interaction for 2D intangible displays [7]. The apparatus was based on the use of a Fresnel lens to create the optical illusion of true-3D. Results showed users performed poorly in distinguishing the z-coordinate of the targets. The use of visual pseudo-shadow feedback improved user performance. Vermeer [1] and HoloDesk [15] present interactions with holographic images, such as direct-touch, scooping and grasping. The growing body of literature on 3D user interfaces [12, 13] inspires the design of the basic direct input methods we developed in our prototype applications.

### DISPLAY PROPERTIES

The following factors, which we derive in part from the MMDE literature, can influence the display properties of a true-3D mobile multi-display environment.

#### Projection area

While pico-projectors need a projection surface, true-3D projectors may display an image in mid-air around the mobile device. Prior work has generally kept the mobile projection pointing downward [5], straight [30] or steerable (using a motor to direct the projector) [6]. These solutions provide significant flexibility in finding a suitable projection surface. A true-3D mobile display needs not be constrained by the position of the projection throw. Therefore after considering the potential projection areas around the smartphone, we decided to focus on the top area of the phone (Figure 2-left). This area always remains visible when the user rotates the phone to inspect the 360° true-3D image.

### Projection distance to the 3D object

The distance between the mid-air 3D projection area and the smartphone (Figure 2-center) may have an impact on users' visual perception and direct input. If the projection is far from the device, it may affect the continuity of the visual search [5] but even further limit direct-touch interaction with the true-3D and require indirect forms of interaction. Previous research has investigated the effect of visual separation (angle and distance between displays) on MMDE [5]. Results show there was no effect on task performance but a higher number of eye context switches occurred if both displays were not in the same field of view (approximately 30% higher). We further investigate this factor in our first experiment.

### Projection angle

We define the projection angle as the angle between the phone's y-axis and the 3D object. Traditional depictions of mobile true-3D envision the 3D content at a 90° angle relative to the phone's plane (Figure 2-right) or displayed directly over the touchscreen (see Samsung's concept vision in [26]). These depictions assume the best projection 'extends' the mobile display into a 3D volume. However, this projection style presents several drawbacks. First, it is unclear how both displays would properly overlay as there could be issues with color mixing and light intensity since the touchscreen would likely be brighter than the true-3D. Second, this vision is limited as it considers the true-3D simply as an extension of the touchscreen instead of viewing it as a secondary display that can extend the mobile phone's capabilities. Our first experiment examines this factor, studying angles from 90°, perpendicular to the phone's plane, to 0°, collinear to the phone's plane (Figure 2-right).

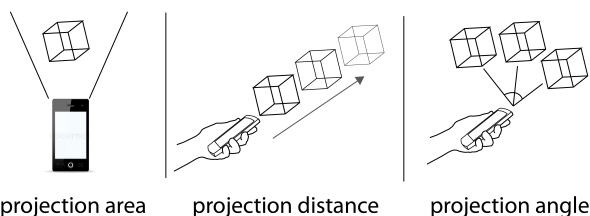


Figure 2. Projection area, distance and angle properties.

### Projection volume

On traditional MMDE, display sizes are heterogeneous (i.e. pico-projection and mobile touchscreen will vary in size) and further can change while being used if the device is hand held. Some studies with fixed displays show better performance for identical size compared to different size displays [5,24]. In the case of mobile true-3D, its volume may affect visual search as well as direct-touch interaction, a factor we investigate in our experiment.

### Point-of-view

The visual exploration of a mobile true-3D display can rely on wrist rotation dexterity to avoid complex interactions for rotating it. There are three main rotation axes around the

wrist: flexion/extension, pronation/supination and ulnar/radial deviation. Previous work on wrist dexterity has identified that the maximum angles are 60°-45° for flexion/extension, 65°-60° for pronation/supination, and 15°-30° for ulnar/radial deviation [23]. We thus expect limited accessibility to occluded areas on the true-3D projection, which we explore across the entire 360° true-3D display using wrist dexterity alone.

### MOBILE MULTI-DISPLAY TRUE-3D PROTOTYPE

To gain an understanding of the various properties that could influence the human factors aspects with mobile multi-display true-3D we implemented a running emulation.

### Prototype

Our implementation was based on a stereoscopic display coupled with head tracking, on the VisCube [27] platform, an immersive environment composed of a projection wall and floor (Figure 3). In this system the user has to wear polarized glasses with IR markers to allow visual head-tracking. 3D content was developed using GLUT. We used an LG-P925 smartphone (4.3'' screen, weight 60z, dimensions 5"(H)×2.7"(W)×0.4"(D)) running Android 2.3. The position of the mobile device in the environment was tracked using a Vicon IR motion tracking system and IR markers.



Figure 3. We built our prototype in a VisCube 3D [27] immersive system, tracking the mobile device using optical tracking.

### Emulating True-3D

Emulating true-3D in a CAVE environment has been achieved by others. The Personal Cockpit project investigated the use of virtual windows around the user through emulation in a VisCube immersive system [9].

### EXPERIMENT 1: DISPLAY PROPERTIES

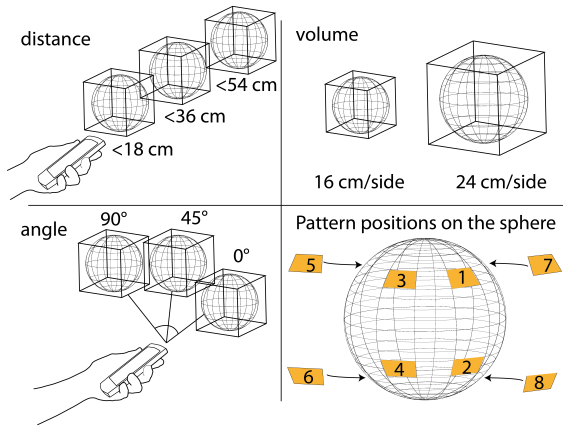
The goal of this experiment was to identify the best spatial configuration for the projection of a mobile true-3D display to ensure effective visual search. We explore the case of natural user interaction, i.e. wrist rotation for search without interface support, but do not explore any type of user input. We focus on the properties described in the previous section (Figure 2): projection's **angle** to the phone plane, **distance** to the phone, **volume** and **pattern position** on the true-3D (point-of-view).

### Apparatus, task and instructions

We used the VisCube prototype described previously. Participants were required to identify the location of a



graphical pattern on a 3D opaque sphere on the true-3D display. The sphere was separated into eight parts, each one containing a unique pattern. All eight patterns had the same volume and were symmetrically positioned on the sphere (Figure 4).



**Figure 4. Properties evaluated in the first experiment: true-3D display's distance to the phone, volume size, angle to the phone and item position on the sphere (4 positions on the front and 4 symmetrical positions on the back).**

We used a basic geometrical shape as the stimulus pattern: circle, triangle, square or star. The stimulus was displayed on the mobile screen to ensure that all users would start the task with a focus at the same location. Once the stimulus was displayed, in black and white, users had to find its coloured version on the sphere (three possible colors) and click a button of the same color on a wand (held with the other hand) to select the answer (Figure 1). Color allowed differentiating correct and false answers with minimum cognitive load. We used the wand merely as an experimental input to collect answers from users. Buttons on the wand are disposed overall at the same distance of the thumb. In case of an error, we recorded the number of attempts it took the user to find the correct answer. We asked participants to minimize selection time (primary demand) without neglecting the error rate (secondary demand).

The position of the sphere was determined by the angle and distance factors (Figure 4). The angle represents the position of the sphere in relation to the mobile device plane: 0°, 45° and 90°. We defined three approximate distances to the center of the sphere: *close* when the sphere center is 18cm away from the top of the device, *middle* when it is 36cm away and *far* when it is 54cm away. The volume factor represents the volume of the sphere's bounding cube (Figure 4): the small cube was approximately 16cm/side and the big cube 24cm/side.

#### Participants

11 men and 1 woman, aged 25.25 years on average, volunteered for the experiment. All of them were right handed; 9 held the mobile device in the right hand while 3

preferred to hold it with the left hand. We mirrored the results from those 3 participants for the position factor (others factors not being dependent on the hand used).

#### Procedure

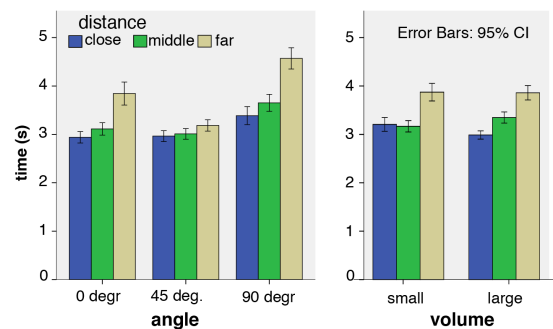
We used a 3×3×2×8 within-subject design with Angle (0°, 45° and 90°), Distance (close<18cm, medium<36cm and far<54cm), Volume (small=16cm/side and big=24cm/side) and Item Position (8 areas of the sphere) as factors. We ran three blocks of trials for each condition. Angle was counter-balanced using a 3×3 Latin square while other factors were presented randomly. We ran one training block before the experiment. The experiment consisted of 144 conditions×3 blocks×12 users = 5184 trials.

#### Data Collection

We recorded the smartphone's position and angle with the Vicon. Besides success rate, we measured trial completion time, from stimulus onset to button pressed. Participants filled a 5-point Likert scale with nine questions to indicate preference for the four factors. The ninth question asked whether users liked the concept of mobile true-3D.

#### Results and Discussion

We used the univariate ANOVA with post-hoc comparisons using the Bonferroni adjustment for all our analyses. Results reveal a main effect of angle ( $F_{2,22}=97.6$ ,  $p<.001$ ), distance ( $F_{2,22}=91.6$ ,  $p<.001$ ) and position ( $F_{7,77}=139.7$ ,  $p<.001$ ) on completion time. There was no consistent effect of volume ( $F_{1,11}=.116$ ,  $p=.734$ ). We found an interaction between angle and distance ( $F_{4,44}=12.6$ ,  $p<.001$ ) and volume and distance ( $F_{2,22}=5.5$ ,  $p=.004$ ). We also found interaction effects for position and angle ( $F_{14,154}=11.7$ ,  $p<.001$ ) and position and distance ( $F_{14,154}=11.6$ ,  $p<.001$ ). We did not find any main effects for accuracy rate, with an average success rate of 97.8% (std. dev. 0.2%).

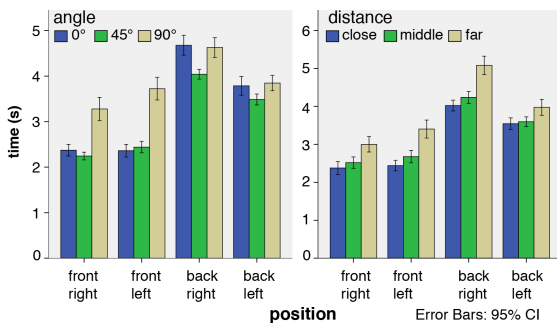


**Figure 5. Completion time (s) for Distance at each Angle (left) and Distance at each Volume (right).**

**Angle and distance interaction:** Overall, completion time increased with distance and angle (Figure 5). Post-hoc analysis revealed a significant difference between all three angle positions, 45° (3.0s on average) being slightly faster than 0° (3.3s) and considerably faster than 90° (3.8s). Surprisingly, the worst projection angle for a visual search task on a mobile true-3D display is 90°, which contradicts popular conceptions of such displays. Subject 4 reported

“when the sphere is farther horizontally it is easier to search than if it is vertically distant”. Concerning distance, we also found a significant effect between all three distances. Performance with the close (3.0s) being slightly better than with middle (3.2s) and considerably better than far (3.8s). The results do not reveal a linear degradation of distance on visual search (while distances values are linear) and can thus expect a stronger negative impact with further positions. Concerning the interaction between both factors, we notice in the case of 45° that the difference between all three distances is less important than for the other two angles, which are particularly bad when the 3D image is far.

**Position on the sphere:** As expected, the front hemisphere is significantly faster than the back (2.7s vs. 4s). Our experiment required searching for objects in the back part of the projection to capture conditions such as 3D object occlusion. We find a significant difference between the back right (3.7s) and left elements (4.4s) (Figure 6). Several users reported severe difficulty reaching the back left elements and particularly the top left object (4.6s) due to a physiological limitation in wrist rotation. This result is consistent with previous works on wrist rotation [23].



**Figure 6. Completion time (s) for Position at each Angle (left) and Position at each Distance (right).**

#### User Preference

From the 5-point Likert scale most participants found it difficult to search for content on the large sphere (Q1), did not show a specific preference for projection at 0° or 45° (Q7), and generally preferred the closer distance (Q6). All participants liked the concept of having mobile true-3D.

#### Summary of Experiment 1 Results

This first study demonstrates that a 3D object positioned at an angle of 0° or 45° and at a distance of less than 36 cm (middle or closer) performs best for visual search. We also found that the region in the back and opposite to the hand holding the device is the weakest for object search. This is primarily due to the wrist dexterity as observed during the experiment and from participant feedback. Wrist dexterity also affects objects located further away, i.e. these become hard to inspect under all angles. Thus, our results recommend shorter distances if the device is to solely rely on wrist rotation for viewing the display. We consider these constraints in experiment 2.

## EXPERIMENT 2: DIRECT-TOUCH INTERACTION

The goal of this experiment is to investigate the effect of volume size on visual search and direct input.

#### Task and instructions

Participants were required to identify the location of a graphical pattern on the true-3D display and to select it with direct mid-air touch. To complete the task, they used their finger, equipped with IR markers, to touch the pattern on the sphere. We added one larger volume size than in experiment 1 to further investigate this factor (note we did not see effects of the volume sizes selected in the previous study). The small volume was approximately 16cm/side, the medium 24cm/side and the large 36cm/side (the volume is constant and does not scale to the distance to the user).

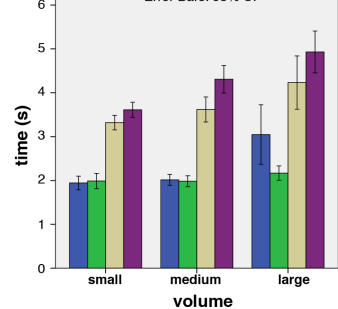
#### Participants

Ten men and two women, of an average age of 26.3 years volunteered for the experiment. Five participated in the first experiment. Eleven were right handed and one left-handed. They all held the mobile device with their left hand and used their right index finger for interaction.

#### Procedure and data collection

This experiment followed a 3×2×8 within-subject design with Volume (small, medium and large), Angle (0° and 45°) and Item Position (8 areas of the sphere) as factors. We set the Distance factor to *middle* (<36cm) to allow enough space for the large Volume to be displayed. Three blocks of trials were run for each condition, the angle factor being counter-balanced while others were selected randomly. We also ran one training block before the experiment to reduce the learning effect between new participants and those who participated in the first study. The experiment consisted of 48 conditions×3 blocks×12 users = 1728 trials. We

collected the same data as in the previous experiment.



**Figure 7. Completion time (s) for Volume at each Position.**

We found an interaction between volume and position ( $F_{14,154}=2.5$ ,  $p=.002$ ). There was no consistent effect of angle. Concerning error rate, there was a significant effect of volume ( $F_{2,22}=9.6$ ,  $p<.001$ ) but no significant effect of other factors (position and angle).

**Volume:** Overall, completion time increased with the volume of the display (Figure 7). A post-hoc analysis revealed there is no significant difference between the mean

time for the small (2715 ms) and the medium volumes (2980 ms). There was however a significant difference between those two times and the mean time for the large volume (3594 ms). Results showed that success rate increased with volume: 91.3% for the small volume, 95.8% for the medium volume and 97.2% for the larger volume. Most of the errors were due to the intangible nature of the display. In some cases users had a bad perception of 3D depth and touched a back or adjacent item while pointing: this happened more often on the small volume, since items were closer, thus explaining the difference in success rate. In other cases users inadvertently touched another item on the way to the target.

**Position:** Overall, as in the previous experiment, post-hoc analysis revealed a significant difference between the front (2.6s) and the back items (4s). On the back, the right side (4.3s) was significantly slower than the left side (3.7ms). This can be explained as users held the phone with the left hand and most of the selection time corresponds to the visual search (back right items are more difficult to see).

#### Summary of Exp. 2

This second experiment allowed us to explore suitable values for the projection volume for direct interaction. Projections smaller than 24cm/side improved efficiency. The slight cost in accuracy at smaller volumes suggests that target sizes need to be considered carefully for such displays, and may be limited by the 3D input tracking capabilities. Other forms of feedback for mid-air input, as proposed in [7], could also help improve targeting accuracy. We leave this exploration for future work.

#### IMPLICATIONS OF EXPERIMENTS 1 AND 2

From this first exploration of mobile-mounted true-3D displays we identified suitable display parameters along with some display constraints.

##### Suitable display parameters

Our two experiments enabled us to identify the suitable values for the explored display factors (angle, distance, volume and position of true-3D content). To improve user search and direct input interaction, a true-3D display should be positioned at an angle between 0 and 45 degrees. The center of the projection should be less than 36cm away from the mobile device. The volume of the display should be smaller than a cube of 24cm/side. The back area of the display opposite to that of the hand holding the device is difficult to access. This suggests target reaching techniques could be developed for such displays as with large displays or tabletops [30]. We apply these parameters in the design of our applications, as described later.

##### Implications

*Display Volume:* Based on our experiments, mobile true-3D displays may consider a limited volume if users are to rely on direct input and wrist rotation as their primary means of interacting with the display. This limitation suggests that if 3D content is larger than this volume (for example if displaying a very large map) or if the user wishes to

translate 3D objects, the content should be clipped to the boundaries of the volume space. Our applications are restricted to displaying content within such limits. However, alternative methods for space allocation and effective space management need to be considered.

*Occlusion:* Both viewing and touching the back of the mobile true-3D are particularly difficult. We deduce two main implications from this constraint. First, when creating 3D content for this display, designers should avoid including important content (such as controls or interactive objects) on the back region. The second implication is that to allow easier access to the back region, applications should include rotation techniques for rotating the content of the true-3D and not solely rely on wrist control. The user could then access the back region by rotating the content instead of rotating his wrist.

We use our derived parameters and constraints in the rest of our work: first to frame and define usage scenarios, and second as requirements for our applications.

#### ILLUSTRATIVE EXAMPLES

Our implementation of the two representative examples was informed by two participatory design sessions. These sessions helped us define the display configurations for mobile true-3D displays: how the 2D and the 3D displays content can be coordinated.

##### Participatory design sessions

In both sessions participants filled Post-it notes with their ideas on how users should interact with mobile true-3D. All Post-it notes were read and participants could comment on one another's ideas. We took notes throughout the sessions.

**Design session 1:** The goal of this first session was to collect general mobile true-3D design ideas. We first asked participants to describe how they envision using mobile true-3D and then to think specifically about the map application and give ideas on display content and interaction. Eight HCI graduate students took part in this session. Participants' ideas for mobile true-3D included: maps, instruction delivery (Chinese calligraphy, sign language), online shopping and 3D sculpting. Concerning the 3D map participants suggested displaying "volumetric buildings" and "3D signposts and landmarks". Several general ideas emerged on the relation between mobile and true-3D displays, such as "2D cross-sections" of the 3D image, "screen replication" to avoid the fat finger problem and "real pop-ups" in 3D. Concerning the input, some participants would like to touch the true-3D image ("nice to touch 3D") while others imagined ways of indirect interaction ("3D cursor at a distance from the finger").

**Design session 2:** The goal of the second design session was to collect ideas on how to interact in our 3D environment. Ten students with HCI background participated. On the mobile phone, participants cited "moving/rotating the phone for rotating the true-3D content" and "using the device angle to initiate scale". On

the true-3D space, some participants indicated the use of 3D gestures such as “pinching” for selecting or translating and using the “distance from the hand in mid-air to the phone” for zooming. A participant also cited using “eye tracking” for rotation, and another one using “face tracking” for translation.

### Coordinated display configurations

From these design sessions and the results from the first exploration, we identified a set of display coordination configurations for mobile true-3D. Our configurations are compliant with the **volume constraint** identified in the first exploration: we only consider configurations with a limited true-3D volume. We define four dimensions to characterize the coordinated display configurations (CDC) of 3D and 2D displays:

**CDC1:** Independent (Figure 8-a): both displays show separate content. For example, a GUI on the mobile display used to control the 3D content on the true-3D display.

**CDC2:** Extruded replication (Figure 8-b): the true-3D display extrudes the content of the mobile display in 3D. For example, the 3D view of a map or spatial workspace.

**CDC3:** Cross-section (Figure 8-c): the mobile display shows a cross-section view of the true-3D display. An example is views in 3D Computer Aided-Design (CAD) or in architecture visualization.

**CDC4:** Extruded detail (Figure 8-d): the true-3D display extrudes a detailed 3D view of a section of the content presented on the mobile display. This corresponds to the idea of a “3D pop-up” in design session 1.

In terms of input, we used three interaction modes: on the mobile, on the true-3D area and mixed true-3D + mobile. Interaction can be single handed (on the mobile) or bi-manual (one hand on the mobile, one on the true-3D).

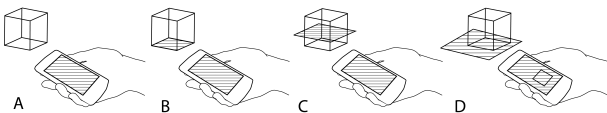


Figure 8. Coordinated Display Configurations of 2D mobile and true-3D displays.

We apply our coordinated display configurations in two applications: the 3D map and the interior design application. Our applications were designed based on the projection parameters we obtained from the first two experiments.

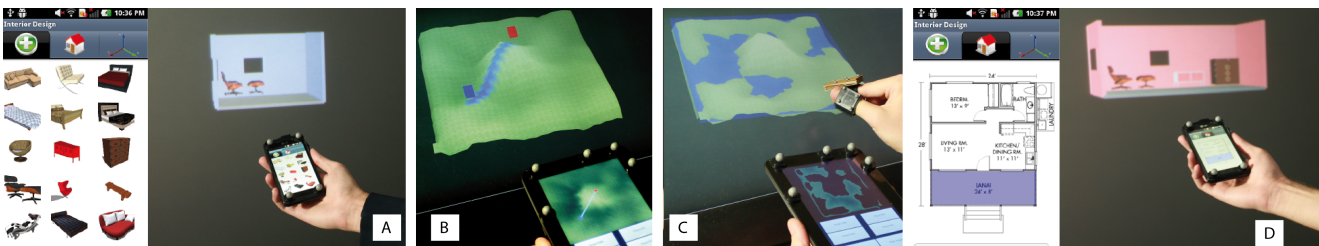


Figure 9. Illustration of the display configurations: A. Independent; B. Extruded replication; C. Cross-section; D. Extruded detail

### Illustrative application 1: 3D map

3D maps can provide information on terrain height. This is useful for military personnel and hikers, who often traverse areas with many rapid changes in elevation. 3D maps provide advantages over conventional 2D maps as landmarks on the former closely resemble their physical counterparts. They also decrease mental load in comparison to 2D maps [22]. True-3D display maps enable users to align their viewpoint with objects in the real-world. Such 3D maps are being constructed in projects such as the Urban Photonic Sandtable Display [8].

Our application consists of a top-down view of a 2D map on the mobile device’s display, while a 3D version of the map is shown on the true-3D display (CDC2, Figure 9-b). The 3D map is clipped to the volume of the true-3D display (**volume constraint** from our first exploration). Interactions can be performed through either the touchscreen or by using pinch gestures in the true-3D display volume. Various overlay information, such as paths and points/regions of interest can be placed by interacting with the 3D map. These are represented by visual markers on the 3D map, which are replicated on the 2D map as well. This display relationship allows for modification of markers on one display, with all changes appearing on both displays. This design helps overcome the difficulty of interacting with the back region of the display (**occlusion constraint** from our first exploration); the 2D view can be used to verify the position of an out-of-sight marker as it is being placed in 3D. Map content can be filtered based on a number of different properties. In our application, we focus on height filtering. A user can move their hand up or down to control a horizontal selection plane, which intersects the 3D map (CDC3, Figure 9-c). The 2D view displays a horizontal slice of the 3D map at the intersection height. This takes advantage of the 3D and multi-display setup by allowing a visualization of the selected height on the original 3D map, while the 2D version displays the height-filtered map.

### Illustrative application 2: Interior design

Interior design is a field with a long history in computer-aided design (CAD). Recently, with the advent of mobile devices, interior design applications such as AutoCAD provide mobile support with great success (more than 5 million downloads on the Android Market). They facilitate the design process in the field and as a result can integrate features into this process, such as taking pictures of the real

space being designed. An obvious limitation of such applications is the difficulty of manipulating 3D content on a mobile device. Mobile true-3D displays could be used to perform and brainstorm 3D interior design in the field.

Our interior design application consists of a 3D representation of the room to design and a mobile interface with three views. The room occupies the true-3D volume and cannot be translated outside of it (a solution to accommodate the **volume constraint**). The front wall is transparent to allow the user to easily inspect and position the 3D furniture (a solution to the **occlusion constraint**). The three views of the mobile phone illustrate different aspects of our display configurations: the first view displays a set of images of furniture (**CDC1**, Figure 9-a). When the user selects furniture, it is added to the 3D room (it appears in mid-air so that the user can select it and position it). The second view represents the floor plan of the house. On the floor plan, the room currently selected is displayed in a different color (**CDC4**, Figure 9-d). The user can select another room to change the 3D room on the true-3D display. The third view allows users to rotate and scale the selected furniture by using a slider or by setting the direct-touch mode to “rotate” or “scale” so that the user can rotate or scale the 3D furniture using his/her fingers by pinching. The user can rotate and position the object at the same time by performing a bi-manual interaction: with the right hand he/she directly rotates the object on the 3D environment, while translating the phone can be delegated to the left hand, thus moving the room.

## DISCUSSION

### Lessons learned

*Mobile true-3D - opening new usages.* Most previous works on MMDE were motivated by the need for expanding mobile displays. One lesson learned from our explorations is that mobile true-3D displays will represent a shift in how users interact with mobile platforms. Accordingly designers will need guidelines to create content for this novel interaction environment.

*Design guidelines:* Our Coordinated Display Configurations (CDC) can help designers explore design alternatives for usage applications. For example, an application to view 3D photos can be implemented using the Extruded detail configuration (CDC4). A 3D image is shown on the true-3D projection while the mobile phone displays the collection of images. Another view can be based on the Independent configuration (CDC1), for example to provide an interface to also allow editing the image. In reality, various CDCs will co-exist in a single application, as we demonstrated in our two proof-of-concept applications. By integrating several CDCs in their applications designers will enrich mobile true-3D interfaces and user interaction capabilities.

*Display properties:* In the context of natural user interactions with mobile true-3D, we can take away the

resulting properties that emerged from our first two experiments. These were designed specifically to allow full content viewing through wrist rotation, and interaction using direct input. The use of other interaction techniques with mobile true-3D may allow overcoming some of these constraints: for example, a technique for content rotation can replace natural wrist-based rotation and allow access to occluded elements on the true-3D.

### Limitations and future work

Our work is built upon some assumptions on the capabilities and limitations of future mobile true-3D displays. We assumed mobile true-3D projectors will be able to display a volumetric image of any volume, at any distance and any angle of the phone. We made this assumption to evaluate the user limitations independently from technology capabilities.

The results of our experiments are influenced by the technology we used to emulate the true-3D displays. The obvious differences between this technology and the final true-3D display, in terms of color, brightness or 3D perception, may alter the results from our experiments. However, most of those results are strongly influenced by human physiological limitations on wrist-based rotation and arm reach. Thus we believe these technical differences do not have a fundamental impact on our findings. Moreover, researchers have used such platforms for developing and testing novel technologies.

There is still much to learn on how mobile true-3D displays will be used in a real mobile situation. A next step in our work will be to use existing mobile augmented reality glasses [28] to evaluate the display properties in mobile situation. A second perspective to our work is to explore the input space of mobile true-3D. In our work we have identified some input configurations and implemented several interaction techniques, including finger gestures such as pinching. We will explore the input configurations space to propose a full suite of interaction techniques. Finally, a third perspective derives from user collaboration on the true-3D display. This will allow us to elaborate on our coordinated display and input configurations.

### CONCLUSION

In this paper we present a two-part exploration of the concept of mobile true-3D. In the first exploration we do the first ever study on mobile true-3D display properties. Results reveal the best distance to phone, angle to phone, display volume and positions on the display for good visual search and direct touch input interaction on mobile true-3D displays. We use these results to guide the design of our applications. Finally, in the last exploration we identify coordinated display configurations between the 2D mobile display and the true-3D display. We define four display configurations named Independent, Extruded replication, Extruded detail and Cross-section. We illustrate those configurations using our two proof-of-concept applications.



In HCI we find numerous examples of novel technologies whose adoption, from discovery to commercial use, take decades [4]. Buxton refers to this process as the Long Nose of Innovation [3]. Our paper is motivated by our will to reduce the long nose for mobile true-3D. While extensive research is taking place to engineer mobile true-3D [8,10,16,20] we provide a contribution in this vein to identify application guidelines, limitations and scenarios to help future adoption of this technology.

#### REFERENCES

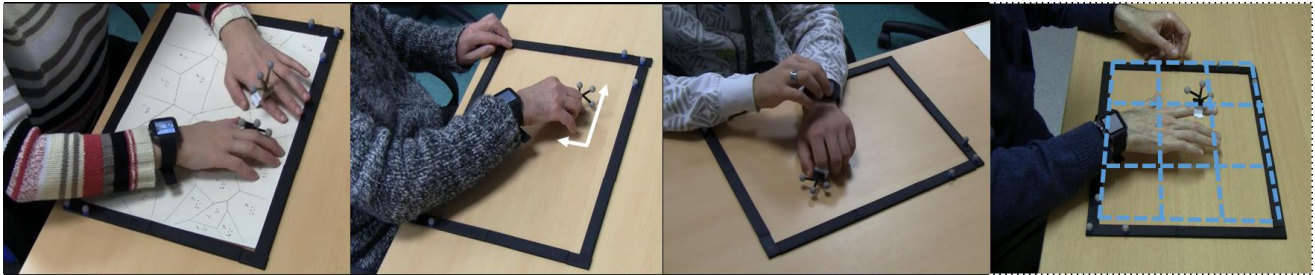
1. Bergé, L.-P., Serrano, M., Perelman, G., and Dubois, E. 2014. Exploring Smartphone-Based Interaction with Overview+Detail Interfaces on 3D Public Displays. In *Proc. of MobileHCI '14*, ACM, 10 pages.
2. Butler, A. et al. 2011. Vermeer: direct interaction with a 360° viewable 3D display. In *Proc. of UIST '11*, ACM, 569-576.
3. Buxton, B. The Long Nose of Innovation. CHI 2011 Keynote.
4. Buxton, B. Buxton Collection (last accessed February 2014). <http://research.microsoft.com/buxtoncollection>
5. Cauchard, J., et al. 2011. Visual separation in mobile multi-display environments. In *Proc. of UIST '11*, ACM, 451-460.
6. Cauchard, J., et al. 2012. Steerable projection: exploring alignment in interactive mobile displays. *Personal Ubiquitous Comput.* 16, 1 (January 2012), 27-37.
7. Chan, L. et al. 2010. Touching the void: direct-touch interaction for intangible displays. In *Proc. of CHI '10*, ACM, 2625-2634.
8. DARPA (last accessed February 2014), <http://www.darpa.mil/newsevents/releases/2011/11/23.aspx>
9. Ens, B., Finnegan, R. and Irani, P. 2014. The Personal Cockpit: A Spatial Interface for Effective Task Switching on Head-Worn Displays. In *Proc. of CHI '14*. ACM, 10 pages (to appear).
10. Fattal, D. et al. 2013. A Multi-Directional Backlight for a Wide-Angle Glasses-Free Three-Dimensional Display. In *Nature* 495, 348-351 (21 March 2013).
11. Froehlich, B. et al. 2006. The GlobeFish and the GlobeMouse: two new six degree of freedom input devices for graphics applications. In *Proc. of CHI '06*, ACM, 191-199.
12. Hancock, M. et al. 2007. Shallow-depth 3d interaction: design and evaluation of one-, two- and three-touch techniques. In *Proc. of CHI '07*, ACM, 1147-1156.
13. Hancock, M., Ten Cate, T., and Carpendale, S. 2009. Sticky tools: full 6DOF force-based interaction for multi-touch tables. In *Proc. of ITS '09*, ACM, 133-140.
14. HelioDisplay, <http://www.io2technology.com>
15. Hilliges, O., Kim, D., Izadi, S., Weiss, M. and Wilson, A. 2012. HoloDesk: direct 3d interactions with a situated see-through display. In *Proc. of CHI '12*, ACM, 2421-2430.
16. Karlsson, N et al. 2012. iAR: an exploratory augmented reality system for mobile devices. In *Proc. of VRST '12*, ACM, 33-40.
17. Kimura, H., et al. 2011. True 3D display. In *ACM SIGGRAPH 2011 Emerging Technologies*. ACM, Article 20, 1 pages.
18. Kratz, S. et al. 2012. PalmSpace: continuous around-device gestures vs. multitouch for 3D rotation tasks on mobile devices. In *Proc. of AVI '12*, ACM, 181-188.
19. Martinet, A., Casiez, G. and Grisoni, L. 2010. The effect of DOF separation in 3D manipulation tasks with multi-touch displays. In *Proc of VRST '10*. ACM, 111-118.
20. Nii, H. et al. 2012. Fuwa-Vision: an auto-stereoscopic floating-image display. In *SIGGRAPH Asia 2012 Emerging Technologies*. ACM, Article 13, 4 pages.
21. Nintendo 3ds, <http://www.nintendo.com/3ds>
22. Nurminen, A. and Oulasvirta, A. 2008. Designing interactions for navigation in 3D mobile maps. In *Lecture Notes in Geoinformation and Cartography*, Springer, pp. 198-224
23. Rahman, M., Gustafson, S., Irani, P. and Subramanian, S. 2009. Tilt techniques: investigating the dexterity of wrist-based input. In *Proc. of CHI '09*, ACM, 1943-52.
24. Rashid, U. Nacenta, M. and Quigley, A. 2012. The cost of display switching: a comparison of mobile, large display and hybrid UI configurations. In *Proc. of AVI '12*, ACM, 99-106.
25. Rukzio, E., Holleis, P. and Gellersen, H. 2012. Personal Projectors for Pervasive Computing. *IEEE Pervasive Computing* 11, 2 (April 2012), 30-37.
26. Samsung Flexible AMOLED vision (last accessed Feb. 2014), <http://youtube.com/watch?v=f4AhTiQkWwk>
27. Visbox, <http://www.visbox.com/>
28. Vuzix, <http://www.vuzix.com/home/>
29. Song, P. et al. 2011. WYSIWYF: exploring and annotating volume data with a tangible handheld device. In *Proc. of CHI '11*, ACM, 1333-1342.
30. Winkler, C., Pfeuffer, K. and Rukzio, E. 2012. Investigating mid-air pointing interaction for projector phones. In *Proc. of ITS '12*, ACM, 85-94.

# From Tactile to Virtual: Using a Smartwatch to Improve Spatial Map Exploration for Visually Impaired Users

Sandra Bardot<sup>1</sup>, Marcos Serrano<sup>1</sup>, Christophe Jouffrais<sup>2</sup>

<sup>1</sup> University of Toulouse – IRIT  
Toulouse, France  
{first\_name.last\_name}@irit.fr

<sup>2</sup> CNRS - IRIT  
Toulouse, France  
{first\_name.last\_name}@irit.fr



**Figure 1.** From left to right: exploration of a raised-line map with an infrared tracker on the finger; exploration of a virtual map based on hand tracking and smartwatch feedback; left/right swipes on the smartwatch to filter the data; and raising hands to gather the data within a virtual grid layout.

## ABSTRACT

Tactile raised-line maps are paper maps widely used by visually impaired people. We designed a mobile technique, based on hand tracking and a smartwatch, in order to leverage pervasive access to virtual maps. We use the smartwatch to render localized text-to-speech and vibratory feedback during hand exploration, but also to provide filtering functions activated by swipe gestures. We conducted a first study to compare the usability of a raised-line map with three virtual maps (plain, with filter, with filter and grid). The results show that virtual maps are usable, and that adding a filter, or a filter and a grid, significantly speeds up data exploration and selection. The results of a following case study showed that visually impaired users were able to achieve a complex task with the device, i.e. finding spatial correlations between two sets of data.

## Author Keywords

Visually impaired users; wearable devices; map exploration; geospatial data.

## ACM Classification Keywords

H.5.2. Information interfaces and presentation: User interfaces.

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

*MobileHCI'16*, September 06-09, 2016, Florence, Italy  
© 2016 ACM. ISBN 978-1-4503-4408-1/16/09...\$15.00  
DOI: <http://dx.doi.org/10.1145/2935334.2935342>

## INTRODUCTION

Visually impaired (VI) people need regular access to geospatial information during education but also during everyday life. Tactile raised-lines maps are the most commonly used tools for that. However, they are difficult to make, they are expensive, and they depend on the intervention of a tactile graphics specialist. Because raised lines elements and legends are cumbersome, tactile maps cannot contain many details. In addition, they must be edited and printed again when an update is necessary. They generate cognitive issues because Braille legend is usually situated outside of the map, and the reader must interrupt the exploration to read it. Finally, according to the National Federation of the Blind, the Braille literacy rate among blind people in 2009 was reported to be near ten percent. To pass along these limitations, recent prototypes of interactive tactile maps combine a tablet with a tactile overlay [3]. However, the tactile overlay must still be printed and cannot easily be modified, thus limiting the benefits of interactive audio feedback.

Delogu and colleagues explored non-visual access to digital geospatial data relying on the sonification of multiple views, including tables and maps [6]. This work provided an interesting framework for data sonification and manipulation, such as brushing or filtering, but was mainly based on a discrete (keyboard-based) exploration of the data. Especially, they did not study a direct spatial exploration of the map, which may provide the user with accurate hand position awareness on that map.

In this paper, we propose a technique to support direct spatial exploration of data on virtual maps, without any overlay, as a mean to replace raised-line maps (see Figure 1). Our technique relies on the combined use of hand

tracking and a smartwatch. The content of the virtual map is rendered using the smartwatch text-to-speech (TTS) engine, for example, a region's name and corresponding unemployment rate, and its vibratory functions to render region borders. The smartwatch is also used as an input device to filter the data by the mean of simple swipe gestures on its touchscreen.

We designed three versions of this exploration technique called Plain, Filter and Grid-Filter. With the Plain technique the watch directly renders the underlying region. With the Filter, the user can select a target data on the watch, which will be the only data to be rendered. With the Grid-Filter technique, the Filter is preserved but, in addition, a spatial layout with nine cells gathers the data. While the Plain and Filter techniques are performed on a surface, the Grid is rendered above the surface and relies on mid-air gestures. When the hand is on the surface, the grid is deactivated to allow precise spatial exploration. When the hand is raised over the surface, the grid is activated to allow a rapid cell-by-cell exploration. Using this technique, we investigated the interest of a rapid but coarse-grained access to the map based on mid-air gestures.

We conducted two studies on these techniques: a comparative experiment and a case study. The comparative experiment involved 12 VI participants, and aimed to compare the usability of the four techniques (regular raised-line maps vs. virtual maps using the three techniques) to access simple geospatial information (name of fictive regions with the name of their main cereal production). Participants explored maps of increasing complexity (increasing number of regions) randomly generated. The second study involved four VI users and aimed at specifying the usage of the Filter and Grid-Filter techniques in scenario including two datasets to be compared (i.e. the population and unemployment rates of the 48 states of the USA). This type of comparison, where the user looks for spatial correlations among multiple types of data, is usual in geo-data visualization (see e.g. [28]).

Our contributions are: 1) a novel solution for VI users to explore virtual maps with data filtering, using a hand tracking technique and a smartwatch, 2) the experimental comparison of the usability of raised-line maps and virtual maps, and 3) a case study relying on a real map exploration with comparison of two sets of data.

## RELATED WORK

Our work relates to research on non-visual exploration of maps. We specifically focus on free map exploration as opposed to guided exploration where the user wants to reach a specific point, which involves other interaction techniques (see for instance [13]).

### Raised-line interactive maps

As mentioned in the introduction, regular raised-line maps present several major limitations. In order to overcome these limitations, one solution consists in adding a tactile

overlay over a touchscreen [3]. Users must double-click identified zones of the tactile map in order to receive audio feedback such as the name and description of points of interest (buildings, parks, streets, etc.) Such a device improves map accessibility because it speeds up learning and it enhances user experience [3]. Furthermore, it provides the user with many layers of information in addition to the name and description of points of interest (e.g. time schedule of public transportation, current movies in a cinema, etc.) However, the tactile overlay must be printed out each time an update is required.

### Tangible maps

Using tangible maps instead of raised-line paper maps presents the advantage of adding dynamic haptic feedback, which can be combined with interactivity. McGookin *et al.* [16] designed a device for helping VI people to access graphs with physical objects (phicons) that represent points of the graph. Touching phicons in predefined cells within a tangible grid, the user was able to explore scatter plots or bar charts. However, the system was not designed to explore more complicated spatial representation with numerous points of interest that do not fit in pre-defined cells. They also observed that the objects were regularly knocked over during the exploration.

More recently, Ducasse *et al.* [7] proposed a novel type of more steady and versatile phicons to construct maps, called Tangible Reels. These phicons are sucker pads that represent points of interest with reels that represent links between pads. Thus, they are specifically suited to represent connections between metro stations or sides in geometric shapes for instance. This device allows VI people to dynamically create maps but also explore existing maps and retrieve specific information related to points and links. Each phicon can be placed anywhere on the screen, and be linked to any other phicon. Hence, this approach overcame the limitation of both raised-line maps (i.e. being static) and tangible maps based on a limited number of positions (see e.g. [16]).

However, tangible maps present major drawbacks. First, the number of phicons that can be used simultaneously is limited because they are cumbersome. In addition, the presence of many phicons (12 being yet large) tends to slow down the hand exploration process [7].

Our main objective was to improve the exploration of geospatial data by visually impaired users. We aimed to design accessible virtual maps that do not rely on any physical artifact (e.g. raised-line map or phicons), but on a mobile device that visually impaired users may already own (e.g. smartphone or smartwatch). Such mobile virtual maps may then be used in many different places (at school or at home for instance).

### Virtual maps

Previous work has already explored how VI people can access virtual maps. There are three main approaches for



exploring digital maps by visually impaired users: using the keyboard, using haptic devices (e.g. mouse or phantom), and using a touchpad or a tablet.

Using the keyboard, VI people can navigate through maps regions sequentially [18]. The keys are used to move from one region to another one. Another possibility is to divide the map according to a grid layout [28], usually made of 3x3 cells. Users can target a cell with the numeric keypad to quickly get the information related to that cell. These keyboard-based approaches are efficient to navigate a spreadsheet but are less efficient to provide the user with a mental spatial representation of the relative locations of regions on a map. In fact, keyboard-based exploration, being strictly symbolic and discrete, required more cognitive effort for reconstructing the explored layout [6].

Map exploration based on haptic devices generates cognitive issues too. Haptic devices do not provide any stable reference frame, thus differences exist between perceived distances and real distances [12]. In addition, users can lift up the mouse and move it [14], which generates disorientation when operating the mouse without vision [19,24].

Finally, some studies focused on map exploration on a touchscreen or tablet. [25] investigated the sonification of simple navigation maps on handheld touch screens. The evaluation showed that users of the Timbre prototype recognized a few shapes in a matching to sample protocol, but can also develop a rough understanding of an indoor floor plan. [20] asked users to explore a smartphone-based map with vibrotactile and audio output. In that study too, participants correctly perceived basic spatial information. More recently, [9] used a tablet instead of a smartphone, and observed that they were usable but still rely on demanding cognitive processes. In fact, in all these studies relying on touch sensitive screens, it appears that the task of following a line with a finger is difficult, even though the guidance is improved with vibratory and audio feedback. On the contrary, the detection of adjacent zones based on auditory and audio feedback is quite easy. In addition, the memorization of the relative positions of the different zones is effortless because hand exploration is direct (in register with the map). Finally, an obvious drawback of handheld devices is that they only provide a limited surface for exploration, and hence require recurrent panning and zooming operations that are very difficult to perform by visually impaired users.

In the current study, we used a hand tracking technique instead of touch input on a tablet. Hand tracking techniques are now frequent and cheap [15]. They allow multiple hands and fingers tracking, as well as much larger surfaces for exploration than tablets. Furthermore, we used a smartwatch to provide audio and vibrotactile feedback. There are many advantages of the smartwatch over a smartphone or tablet as a personal device. First, with the smartwatch, the two hands are completely free, which is

very important for tactile exploration [1]. Then, the smartwatch provides audio and vibrotactile feedback that is collocated in space with the point or the area of interest, which may reinforce the understanding of the map. Finally, the smartwatch provides a second interactive surface that may be used to provide the user with input functions clearly separated from the exploration surface. This segregation of input surfaces may also enhance the general understanding of the map.

### **NON-VISUAL GEOSPATIAL DATA EXPLORATION**

The current work focuses on the spatial exploration of geographic maps with associated data (e.g. demographic or weather maps). Recent work has focused on non-visual techniques to “visualize” complex data (the visualization being tactile or auditory). [28] designed a device relying on multiple views (e.g. spreadsheets or map views) to access demographic data. In that system, users can filter the data within the table, and then navigate the map view with the keyboard. Using that filtering function, they hear the selected data only.

Our work was inspired by this previous study on non-visual data exploration. However, we focused on the complementary objective of improving the spatial component of exploration, providing the user with large surfaces and collocated feedback. We also aimed at providing the user with map access in any situation (home, school, work, etc.) relying on a personal device such as a smartwatch.

Transforming visual maps into virtual maps that make sense for visually impaired users is not straightforward. In the following section, we first analyze the layout and content of regular raised-line maps. We identified the elements that should be preserved to design accessible virtual maps. We then detail the techniques that we designed to explore virtual maps based on hand tracking and a smartwatch.

### **Tactile raised-line maps for VI people**

Tactile raised-line maps have two main advantages: information is tactile, and spatial exploration is direct. They are made according to guidelines (simplification of contours, reduction of the number of elements, legends, etc., see e.g. [26]). These maps contain important elements:

- Contour of areas, rendered through raised lines;
- Points of Interest (POI) and labels represented with specific tactile symbols ;
- a Braille legend describing each area and POI;
- Data associated to each area or POI, for instance the population of a region. In raised-line maps this information is usually written outside the map with Braille.

In addition, tactile exploration of novel spaces relies on behavioral strategies. For instance, [11] has observed specific strategies for the elaboration of object to object relationships. We observed similar behaviors when visually impaired users explore raised-line maps: they frequently

come back to previously explored objects or areas; but they also adopt more global strategies, such as scanning the image from left to right and top to bottom. Finally, it is important to note that raised-lines maps support bimanual exploration.

We aimed to preserve these strategies during spatial exploration of virtual maps. We used hand tracking to locate the hands, and we used the audio and vibratory features of the smartwatch to render the information that is under the hand.

### Hand tracking for VI people

Spatial exploration by using hand tracking instead of touch input offers several advantages for VI people. First, VI people tend to put multiple fingers and hand palm down on the surface, which, in absence of visual feedback, generates unexpected events [1]. Instead, hand tracking can simply associate one point with each hand or finger. Second, hand tracking allows performing mid-air gestures, for instance to change the information level of the map when raising the hand. Although mid-air exploration may seem difficult for VI people, in this paper, we explore its use to perform coarse-grained spatial exploration (i.e. by using a 3x3 grid). Finally, mobile and low-cost hand tracking solutions have recently been proposed [15], which could leverage making map exploration possible in different contexts and on different surfaces, such as a regular desk, without the need for an underlying touch sensitive surface. Coupling hand tracking with a wearable device for input and feedback (e.g. a smartwatch) makes it possible for VI people to explore maps in many places such as school or at home.

### Using a smartwatch for localized feedback

Recent work has shown the interest for VI people to use wearables [27], and particularly smartwatches: they are small, easily accessible and unobtrusive to wear, and can improve information access and social interactions. Wrist is the preferred part of the body for a wearable [21]. Current smartwatches have the advantage of including speakers and vibratory feedback.

We decided to use a smartwatch to leverage hands-free map exploration. We used the smartwatch both as input and output. As input, the device's touchscreen is used to filter or brush data by performing simple horizontal or vertical swipe gestures (Figure 2). As output, the device is used to render localized Text to Speech (TTS), for instance the name of regions. The vibratory feedback is also used to render information, such as the name of POIs or the border between regions. We identified different mappings between these input/output modalities and the map exploration task. We designed three different exploration techniques (Plain, Filter and Grid-Filter) based on the smartwatch.

### Plain exploration

Plain exploration is the exploration of a virtual map equivalent to the exploration performed on a raised-line map: each element on the map is rendered.

*Input interaction:* The smartwatch is only used as an output for this technique.

*Feedback:* We combined auditory and vibratory feedback. TTS reads out information underneath the hand, such as the name of the region and its population. A 100 ms vibration notifies the transition from one region to another one. A continuous vibration means that the hand is outside the map.

### Filter exploration

Filtering data before exploration allows reducing the amount of information to render through TTS, and thus reduces the user's cognitive load. The filtering allows selecting a sub-range of values, for instance regions with more than 100 thousand residents. To perform the filtering, users make swipe gestures on the smartwatch.

*Input interaction:* A succession of horizontal finger swipes on the smartwatch reads out the filter values (depending on the scenario). A double-tap selects the current filter.

*Feedback:* After selection, only the data that corresponds to the selected filter is read out. According to the filter state, TTS reads out information underneath the hand, such as the name of the region and its population. As in the Plain mode, a 100 ms vibration notifies the transition from one region to another one. A continuous vibration means that the hand is outside the map.

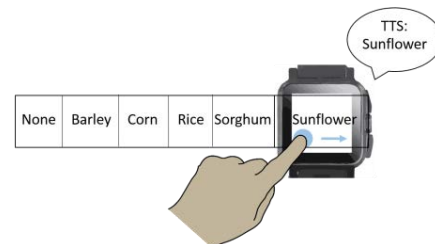


Figure 2. Swipe gesture to select a cereal among others

### Grid-Filter exploration

As previously mentioned, filtering reduces exploration time and user's cognitive load. However, it can be difficult to find certain regions in a map especially if they are small. To get a full glance of a map without missing any region, one solution consists in using a 3x3 grid [28], i.e. reading out the information concerning all the regions contained in each cell of the grid.

However, when gathered within a grid, the spatial relationships between regions are masked. To overcome this limitation, we combined the Plain exploration mode with a Grid-based exploration mode. The user can use one or the other interaction level according to hand height above the map.

*Input interaction:* When the hand is lying onto the table, the user explores the map in Plain mode. When the hand is moving over the table, the user explores the map in Grid mode.

*Feedback:* At the surface level, the interaction is identical to the aforementioned Filter technique. At the Grid level, a 100 ms vibration notifies the border between two cells of the grid. A double vibration pattern is used to notify the user when he changes the interaction level, i.e. when he is raising or lowering the hand.

### COMPARATIVE STUDY: MAP EXPLORATION

The goal of our study was to compare the effectiveness of our virtual map versions against a raised line printed version during an exploration task. We evaluated each technique described above on fictive maps.

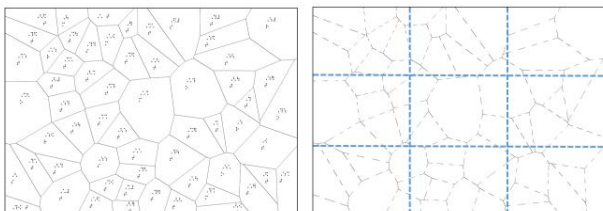
#### Task and instructions

The task was to explore a map, and answer a question as fast as possible. The question was the same for all trials: *Give the name of the four regions that contain <name of a cereal>*. If the participant had not found the regions in less than 4 minutes, we considered the trial as a failure. The participant was allowed to provide each response sequentially (during the exploration) in order to avoid memorization.

#### Maps

We designed four sets of fictive maps having the same size (A3), but different number of regions: 30, 45 and 60 regions (Figure 3). To create the maps, we used the Voronoi algorithm [8] configured to randomly generate regions of different areas fitting in a 29.7 \* 42 cm surface (A3 format). We chose that size because it is the format preferred by visually impaired people [2].

Twenty-seven raised-line maps were printed on A3 sheets of papers. In order to reduce the number of printed maps, we reused raised-line maps with different participants. However, for a given map, we changed the name of the cereal to be found to ensure that we never repeated any condition.



**Figure 3. Raised-line (left) and virtual (right) map with 60 regions. Each region includes a number and a letter (braille in the raised-line map). The grid is shown on the virtual map.**

Each region within the raised-line maps included two Braille labels corresponding to the region's name and the most common cultivated cereal. To reduce the size of the labels, the name of the region was represented by a number and the cereal by its initial. For instance, the region 11, which contains corn, was described as "11 C".

We generated 324 virtual maps so that we never reused any virtual map across participants and trials: each participant used 27 maps for each virtual technique. Virtual maps

present the same general layout and the same labels (number and initial), which were rendered with a TTS engine. We used a 100 ms vibration of the smartwatch to render borders between regions.

#### Participants

Twelve visually impaired participants (5 females), aged between 20 and 65 years (M=47, SD=13), volunteered for this experiment. All of them were proficient in Braille reading: they declared an average subjective reading proficiency of 2.9 on a scale of 1 to 5 (SD=1.3, min=2, max=5). The level of visual impairments varied: 7 of them were legally blind, and 5 had residual light perceptions (they were blindfolded during the study). Six of them had a bachelor degree, five had a master degree and one was a PhD student. Concerning their occupation, 2 were students, 3 were pensioners, 3 had a work and 4 were unemployed. Eleven of them have a smartphone that they daily use; one has a tablet and none of them had ever used a smartwatch.

#### Design and procedure

The experiment followed a 4x3 within-participants design with Exploration Technique (*Raised-line, Plain, Filter and Grid-Filter*) and Number of Regions (30, 45, and 60) as factors. The conditions were grouped in blocks including only one Exploration Technique. Within each block, participants repeated three trials for each Number of Regions. We counterbalanced the Number of Regions within each block. The whole order of blocks and maps was counter-balanced across participants. We informed users that they could take a break between blocks. Before using a technique, participants completed a training session. During the training session, we asked users to definitely choose one hand to perform the exploration, and the other one to interact with the smartwatch. They were told how to use the technique, and they were asked to find some regions with associated data. Once they felt comfortable with the technique, they could start the block. All participants chose to use their non-dominant hand to explore and wear the watch, and their dominant hand to perform swipe gestures on the watch.

#### Apparatus

For hand tracking, we used infrared optical markers tracked by 8 OptiTrack cameras (1mm precision). The system senses the 3D position of markers (x, y and z) at 100HZ. Markers were positioned on the index finger of each hand and on the corners of the interaction surface (Figure 4). We used an Android smartwatch SimValley AW-414 (91 grams, 45.3\*44.3\*14.1 mm, 28\*28 mm touchscreen) with Google TTS. We used TCP sockets over a local Wi-Fi network to connect the watch and the cameras to the main computer.

To set the different parameters of the smartwatch (TTS speed and volume), we carried several user testing. The speed of the TTS engine was twice faster than the default Google TTS speed. Vibratory feedback was set to 100 ms long when crossing a border between two regions or two

cells within the grid. It was set to two pulses of 100 ms when raising the hand up or down, and thus changing of exploration level (Plain to Grid level and vice-versa). A continuous vibration indicated that the hand was out of the map.



**Figure 4. Left: Experimental setup with the infrared cameras. Right: marker on the finger.**

Concerning interaction with the watch, we used the default Android *onFling* callback to detect swipe gestures, and the *onDoubleTap* callback to detect double taps. We defined distance and velocity parameters based on user testing to ensure that swipes could be easily performed.

**Collected data**

We logged all tracking data (hand movements). For each trial, we measured the completion time as follows: we started the timer when the user had understood the question and was ready to explore the map; we ended the timer when the user had answered the question.

After each block, participants had to fill a NASA-TLX questionnaire [10] about the technique that was just used. At the end of the session, we collected users’ preference, and the aspects they liked and disliked about each technique. We also asked whether they used exploration patterns or strategies.

We collected 4 Techniques x 3 Number of Regions x 3 repetitions x 12 participants = 432 trials.

**Results**

We used a Shapiro-Wilk test to determine the normality of the distributions of collected data. Because the distributions were not normal and could not be normalized, we used non-parametric Wilcoxon and Friedman tests for two or multiple comparisons respectively. We used the Bonferroni correction when needed ( $p < .008$  for multiple comparisons between conditions).

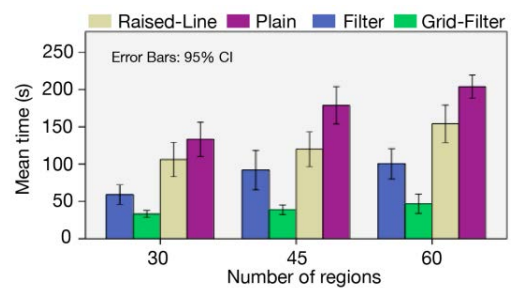
*Time performance*

A Friedman test revealed a significant effect of the Technique factor on the mean time needed to answer questions ( $\chi^2(3)=30, p < .01$ ). More precisely, a series of Wilcoxon tests with correction showed a significant difference between the Grid technique and all other techniques: Raised-Line ( $Z = -5.15, p < .001$ ); Plain ( $Z = -5.23, p < .001$ ); and Filter ( $Z = -4.57, p < .001$ ). We also found a difference between Filter and Plain ( $Z = -5.18, p < .001$ ). Overall, Grid-Filter was faster than the other techniques: on average, answering a question with the Grid-

Filter technique took 40 s, with Filter 83 s, with Raised-line 127 s and with Plain 172 s (Figure 5).

When we analyzed the results according to the Number of Regions, we found no difference between Grid-Filter and Filter. However Grid-Filter was always faster than Raised-Line (30 regions:  $Z = -2.98$ ; 45 regions:  $Z = -2.82$ ; 60 regions:  $Z = -3.05$ ; with  $p < .01$ ). Grid-Filter was also always faster than Plain (30 regions:  $Z = -3.05$ ; 45 regions:  $Z = -3.05$ ; 60 regions:  $Z = -3.05$ ; with  $p < .01$ ). The Filter technique was faster than Plain for 30 regions ( $Z = -2.90, p < .01$ ) and for 60 regions ( $Z = -3.05, p < .01$ ) (Figure 5).

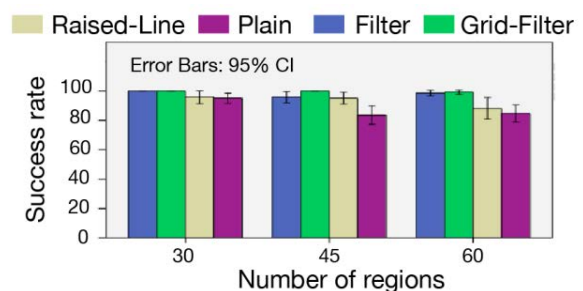
Response times were longer with Raised-Line and Plain. This is due to the fact that users had to thoroughly explore the map in order to find the targeted region and the associated data to answer a question. On contrary, Filter and Grid-Filter renderings quickly provide access to the answer.



**Figure 5. Average time to answer a question for each technique and different numbers of regions.**

*Precision*

Concerning the success rate, i.e. the percentage of regions found, our tests reveal a significant effect of the Technique factor ( $\chi^2(3)=25, p < .01$ ). A Wilcoxon test confirmed a difference between Grid-Filter and Raised-Line ( $Z = -3.29, p = 0.02$ ), Grid-Filter and Plain ( $Z = -4.62, p < .01$ ), and Filter and Plain ( $Z = -4.37, p < .01$ ). On average, success rate was 93.1% with Raised-Line, 87.8% with Plain, 98.1% with Filter, and 99.7% with Grid-Filter (Figure 6).



**Figure 6. Percentage of correct responses for each technique and different numbers of regions.**

For 30 Regions, we only found a difference between Filter and Plain ( $Z = -2.20, p = 0.02$ ). For 45 Regions, Plain was more efficient than all other techniques: Grid-Filter ( $Z = -$



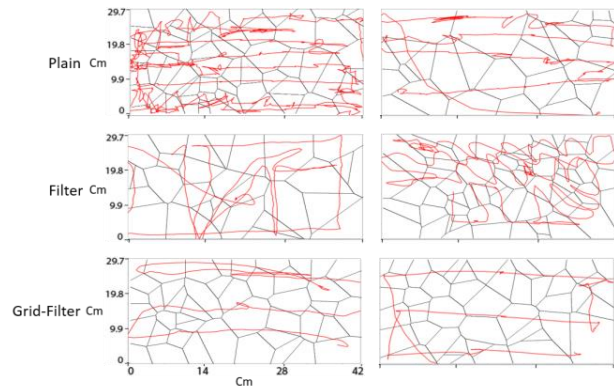
3.05,  $p < .01$ ), Raised-line ( $Z = -2.71$ ,  $p = 0.03$ ), and Filter ( $Z = -2.66$ ,  $p = 0.01$ ). For 60 Regions, we found a difference between Filter and Plain ( $Z = -2.80$ ,  $p < .01$ ) and Plain and Grid-Filter ( $Z = -2.93$ ,  $p < .01$ ).

In general, Plain technique was the most difficult technique to perform the task. During the exploration, it was easy to miss several regions. Sometimes users had to browse the map a second time in order to find all regions.

#### Exploration strategies and hand movements

The observation of hand movements revealed interesting exploration patterns. When exploring the raised-line map, 11 participants used their two hands. With the leading hand, they first read the letter corresponding to the cereal. If needed, they read the number of the region. The other hand was used to find the contours of the neighboring regions to anticipate the following movement (finding the next data).

With the three virtual techniques, most users made either a horizontal or a vertical scanning (Figure 7). Some participants had different behaviors (Figure 7).



**Figure 7. Examples from different participants of hand scanning movement with the three virtual techniques.**

#### NASA-TLX

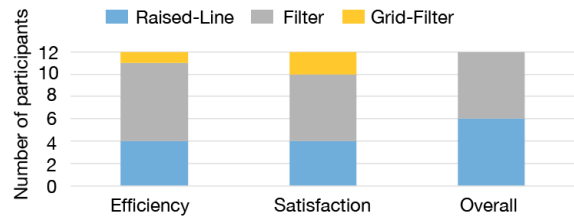
We found no significant differences among techniques on any of the six NASA-TLX properties. Overall, participants were satisfied with the usability of the techniques: on average, on a 0 to 100 scale (0 being low), participants rated Mental Demand 24, Physical Demand 26, Temporal Demand 36, Effort 26, Performance 70, and Frustration 29.

#### User preferences

Participants rated the four techniques in order of preference along three criteria: subjective efficiency (which technique is the more efficient?), subjective satisfaction (which one is more pleasant?), and overall preference (if only one, which technique would you use?) Most participants ranked the Filter technique first on the three criteria: 7 out of 12 on efficiency, 6 out of 12 on satisfaction, and 6 out of 12 in general (Figure 8).

Interestingly, the Grid-Filter technique, which was more efficient according to completion time, was only ranked first on efficiency by 1 participant, and satisfaction by 2

participants. It was never ranked first according to the overall preference. One participant reported (P2) that, with this technique, “it was difficult, in mid-air, to estimate his/her own hand location in relation to above map”. P6 reported “it was tiring to keep the hand in mid-air”. However, other participants reported that the “technique is nice because it allows fast exploration” (P5), and because “it gathers information” (P6).



**Figure 8. Subjective Efficiency, Satisfaction and Overall ranking.**

Concerning the use of a smartwatch, P3 found the swipe technique easy to perform, but, on the contrary, P5 thought it was difficult to use. In fact, P5 is not used to swipe gestures because he has a special phone case that prevents swipe gestures. P7 appreciated “performing hands-free exploration with the watch”. Overall, 11 participants liked using the watch.

#### Summary

Interestingly, the Grid-Filter technique was the more efficient but not the preferred one. We further explored this technique, compared to the Filter alone in our second experiment.

#### CASE STUDY: EXPLORING MULTIPLE DATASETS

The goal of this second study was to validate the use of the best techniques (Filter and Grid-Filter) in a more realistic and complex scenario: exploring the map of the USA with two types of data (unemployment rate and density population). Finding data correlation trends is a usual task in spatial data visualization [28].

#### Map and data

We used a map of the USA with 48 states (we removed Hawai and Alaska, see Figure 9). This map was not familiar to our participants (average of 2.3 on a scale of 1 to 5). For each state, the user could explore two types of data, unemployment rate and density population. We used two different datasets, one for each technique, from two different years: 1980 and 2010. We used the USA unemployment rate and density population reported by the US Bureau of Labor Statistics and by the United States Census, respectively. For the training, we used a different dataset generated randomly.

#### Techniques and task

We asked five different questions for each technique, inspired by a previous study on data exploration [28].

1. (*Find Max/Min*) Give all states with a low unemployment rate

2. (*Find a specific state*) Give the population of a certain state
3. (*Compare data*) Among the states with a high population, which ones have a low unemployment rate?
4. (*Find neighbors*) Among the neighbors of the state X, which ones have a high unemployment rate?
5. (*Value in geographical context*) Does population density grow from East to West?

Theoretically, using the Grid-Filter should provide a benefit for the first three questions, as it allows a rapid and synthetic exploration. The question 4 relies on a more precise spatial exploration. Then, the Grid-Filter can help to locate the state but will prevent a correct identification of the neighbors. For the question 5, the Grid-Filter could be used to compare the states in western cells (1, 4, 7) with those in eastern cells (3, 6, 9).



Figure 9. Map of the USA with a 3x3 grid.

#### Data filtering and user feedback

Data was divided into three types of values: low, medium or high. The TTS engine read out the type of value, for instance: “low”. We used different terms for unemployment rate (low, medium, high) and population (small, average, large). For each dataset, the user could select one range of values or all values.

*Data selection:* Horizontal swipes on the smartwatch selected the data (unemployment rate or population). There were four possible horizontal selections: population, unemployment rate, only state names or all data (Figure 10).

*Filtering:* vertical swipes selected the range of values for the current data.

*Feedback:* the TTS engine rendered the name of the state underneath the hand, and then unemployment rate and population density, respectively. Vibrations were as described in the first study (100 ms for changing states or areas and 100 ms twice for changing between exploration levels).

#### Participants

Four visually impaired women, 47 years old on average, took part in this study. None of them participated in the previous study. Participant 1 (P1), 58 years old, is a Braille teacher and is legally blind. Participant 2 (P2), 56 years old, is a teacher for VI persons and had residual light perception (she was blindfolded for this study). Participant 3 (P3), 19 years old, is a university student and is legally blind. Participant 4 (P4), 55 years old, is a Braille teacher and is

legally blind. All of them have a smartphone, which they use daily, and none of them owns a tablet or a smartwatch.

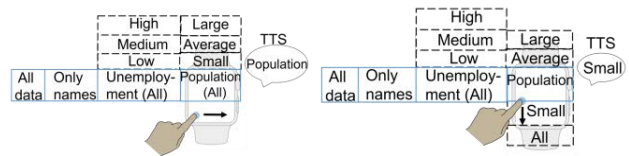


Figure 10. Users can perform left/right swipes to select data (left) and up/down swipes to select values (right).

#### Procedure and collected data

We counterbalanced the two techniques and the two datasets from different years, so that the (data; technique) couple was specific to each user.

During a preliminary familiarization phase, the participants practiced until they felt comfortable with the technique. Participants had a few minutes to browse the map with only state names rendered to get comfortable with it. During that session, they were shown how to use the filter and/or the grid to retrieve data values concerning specific states. This phase lasted 7 minutes on average.

After the familiarization, they had to answer the first question, without any comment or suggestion from the experimenter. When the strategy used during this first trial was not the optimal one, the experimenter described the optimal strategy. He then asked the user to use it to answer that question once more, but about a different state and data. The same procedure was used for the five types of question mentioned earlier. Hence, each user completed 5 to 10 trials with each technique. We logged the hand movements and we measured the completion time for each trial.

Finally, the participants ranked the two techniques according to subjective efficiency (which one is the more efficient?), subjective satisfaction (which one is more pleasant?), and overall preference (if only one, which technique would you use?) We asked them about the strategy that they eventually used to get the responses. They finally had to mention the aspects they had liked and disliked about the techniques. We also asked them what they would change about the two techniques.

#### Results

##### Answers to the questions

All the participants were able to answer the five questions. They performed respectively 13, 11, 10 and 11 trials in total, which means that they were able to find a good strategy on the first trial. On average, they performed 5.7 trials with Filter and 5.5 trials with Grid-Filter to answer the five questions. The only questions where participants were asked to repeat the trial were questions 1 and 3.

On average, they needed 208 s (SD= 68) per question with the Filter, and 148 s (SD=43) per question with the Grid.

Included in that time, the swipe gestures took 39 seconds on average ( $SD=20$ ).

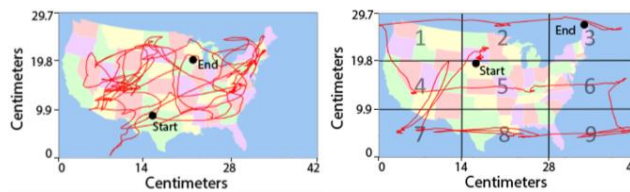
Sixty-six percent of the answers were correct with the Filter and 76.2% with the Grid-Filter. Errors only concerned Question 4 (find a neighbor) with the Grid-Filter, because users provided more states than correct.

#### Data correlation

One user (P3) said that it was convenient to get the two values at the same time, on the same map. She spontaneously compared that condition to raised-line bar charts that she used to explore, and she said that it was more suitable. P1 reported that it was interesting to relate two different types of data on the same map. One of the Braille teachers (P1) thinks filtering on the smartwatch is faster than using a braille map with several data.

#### Overall strategies with Filter and Grid

Logged hand movements reveal interesting differences between the Filter and Grid-Filter conditions. With the Filter condition, participants browsed the entire map and had to slow down in areas where states were small. However, they still miss some states (see Figure 11). With the grid, users did not miss any state but sometimes had to repeat the TTS feedback. In that case they quickly moved the hand out and back in the cell (see cells 4 and 7 on Figure 11).



**Figure 11. Two examples of logged hand movements for the Question 1 (find max/min) with the Filter (left) and with the Grid-Filter (right).**

Two participants (P1, P3) applied the optimal strategies (properly using the two levels of filtering, i.e. data type and value) for all questions except for question 4 (find neighbor) with Grid-Filter: as they used the grid level only, they gave more states than expected. In general, these two participants used systematically the grid level with the Grid-Filter technique. P2 did not like the Filter and used the grid level of the Grid-Filter technique to answer the questions 1, 3 and 4. For the others questions, she browsed the map using only the filter as she already knew some USA states. P4 chose to use only the filter with both techniques: she found it difficult to use the grid level as she felt she could not get a mental representation of the map.

#### Swipe gestures for selecting and filtering

Most users were able perform horizontal and vertical swipes to filter data. However, we observed that participants had more difficulties performing horizontal than vertical swipes, probably due to the arm orientation that led to diagonal swipes. Three participants (P1, P3, and P4) systematically used the filtering function. P2 never

filtered data: she had some difficulties performing swipe gestures, and hence did not like it. She reported that she never uses swipe gestures on her own smartphone.

#### Using the Grid layout

Although three participants used the grid level, only P3 preferred the Grid-Filter to Filter alone. Users were confident with the Grid-Filter because they found it easy and fast to browse the whole map. However one participant (P4) found that the “grid rendered too much information”. In addition, she did not like it because “it is difficult to get a mental representation of the size of each state”.

#### Hand-raising and mid-air gestures

Overall, although three of them used mid-air gestures frequently, participants did not appreciate mid-air interaction, which confirms the observations made in the previous study. Three of them suggested replacing the hand raising by another type of gesture or input. Two participants (P2 and P4) said that it was difficult to estimate precisely hand elevation and that they felt uncomfortable. Moreover, most of them reported having difficulties in knowing the precise location of their hand over the map.

#### Preference

Participants rated the two techniques along three criteria: efficiency, satisfaction and overall. Results were equivalent on efficiency and satisfaction. Two users preferred each technique for each criterion. Overall, most participants (3 out of 4) ranked the Filter technique first.

## DISCUSSION AND FUTURE WORK

### Raised-lines vs. virtual maps

Our work focused on the usability of virtual maps, including different exploration techniques (Plain, Filter and Grid-Filter), as opposed to the regular raised-line maps. Overall, the results suggest that VI people are able to explore geo-spatial data in virtual maps. More precisely, they show that when filtering functions are added to virtual maps, which is impossible to achieve with tactile maps, they provide the user with an efficient mean to retrieve specific answers about the elements of the map.

Concerning the filtering techniques added to virtual maps, the results showed that Grid-Filter is more efficient than all other techniques, but less preferred. The Filter technique alone appears as a satisfying and efficient compromise. The following case study highlighted that both Filter and Grid-Filter techniques are usable to perform more complex tasks, such as comparing two sets of geo-spatial data over a map. Moreover VI participants appreciated both techniques.

### Smartwatches for pervasive access to maps

One of our goals when considering virtual techniques is to leverage map exploration for visually impaired users everywhere (office, school, home, etc.) Our solution is based on hand tracking combined with a smartwatch used to provide the user with both feedback (TTS, vibration) and input to filter data (by type and value). Although it is out of the scope of our paper, embedded and wearable solutions

for hand tracking already exist, for instance using a head-mounted camera [5]. Future work will investigate how these solutions adapt to the exploration of geospatial data by VI people.

### **Two hands for virtual exploration**

During our experiment, most participants used two hands to explore raised-line maps. However, our virtual techniques only involved one hand in the exploration task (the other hand was used to interact with the watch before the exploration). Virtual exploration techniques could probably be improved by involving the second hand in the exploration. Most hand tracking systems allow tracking multiple hands. Concerning the feedback, a second smartwatch or just a Bluetooth bracelet with vibration could also be used on the second hand. However, this solution would probably be cumbersome. Another option would be to use vibratory rings, which have the additional advantage of allowing precise feedback on a single finger. In the future, we plan to investigate how to combine multiple vibratory feedbacks on both hands and/or several fingers and, most importantly, how VI people would perceive them. In addition, the vibratory feedback that we provide during exploration could be much richer and rely on a wide range of vibrotactile patterns [17].

### **Zooming and panning**

We have shown that exploration techniques based on virtual maps allow comparing two sets of geospatial data over a map. This task is considered as a complex and useful task. The design space provided by our device is large and allows performing many more tasks that are impossible with raised-line maps, such as zooming and panning. Panning could simply be performed using gestures on the watch or voice commands. Zooming techniques for VI people have already been proposed [23]. They rely on zoom levels with significantly different content, and which preserve the cognitive grouping of information. A simple solution could be to associate a hand gesture with a specific level to zoom in or out.

### **Mid-air gestures**

Our work explored a novel and even provocative approach by proposing VI people to use mid-air gestures. Mid-air gestures are not frequently used in interfaces for VI users. In fact, during design sessions with VI users, it appeared that they are not, at first glance, in favor of these gestures. Touch, including tactile perception on the skin, is the main sensory modality to perceive objects for VI people. Raising the hand away from the object is then not natural. However, our two studies revealed that they are effective to dynamically change the brushing of the geo-spatial data [6], but also the exploration mode that is provided. We observed that all but one participants actively used mid-air exploration when using the Grid-Filter technique, even though they were not required to. In addition, the Grid-Filter technique was the most efficient technique for retrieving specific information from the map. However, many participants reported that it is tiring if it is too long,

and that it is difficult to build a mental representation of the map when their hand is moving above the map.

Our studies point out a perceptual issue described by [22]. Indeed, they showed that tactile cues contribute to accurate hand positioning. In the Filter condition of our studies, the VI users explore a virtual map by moving the hand on a surface. They can use both tactile (fingers sliding along the surface) and kinesthetic (arm position and movement) feedback to estimate their own hand position. When they move in mid-air, users must rely on their kinesthetic feedback only (the tactile feedback is missing), resulting in a less precise encoding of hand location in space. However, since mid-air gestures appeared as efficient and useful in our studies, we believe that their use should be further explored. One solution could be to use mid-air feedback, such as Ultrahaptics [4], a system that creates mid-air multitouch haptic feedback.

### **CONCLUSION**

In this paper we proposed virtual spatial map exploration techniques as an alternative to regular raised-line maps. Our techniques are based on the combined use of hand tracking and a smartwatch for feedback and input. We defined three types of map exploration: Plain, Filter and Grid-Filter. In a first study, twelve visually impaired users explored a set of randomly generated maps by using these three techniques as well as the classical raised-line approach. Results show that using the Grid-Filter approach is the fastest, but generates discomfort. In a second study, we observed four VI people who explored two types of data (unemployment and population) on the USA map. Results show that virtual techniques are usable to perform complex tasks such as finding correlations between the two sets of data.

### **ACKNOWLEDGMENTS**

We thank all the visually impaired users who participate to the studies. We also thank the LACII lab as well as IJA and Lestrade institutes, Toulouse, FR. This work was part of the AccessiMap project (research grant AccessiMap ANR-14-CE17-0018).

### **REFERENCES**

1. Brock, A. M., Lebaz, S., Oriola, B., Picard, D., Jouffrais, C., and Truillet, P. 2012. Kin' touch: Understanding How Visually Impaired People Explore Tactile Maps. *Conference on Human Factors in Computing Systems - CHI*, ACM, 2471–2476.
2. Brock, A. M., Truillet, P., Oriola, B., and Jouffrais, C. 2010. Usage of multimodal maps for blind people: why and how. *ITS'10: International Conference on Interactive Tabletops and Surfaces*, ACM Press, 247 – 248.
3. Brock, A. M., Truillet, P., Oriola, B., Picard, D., and Jouffrais, C. 2015. Interactivity Improves Usability of Geographic Maps for Visually Impaired People. *Human-Computer Interaction* 30:



- 156–194.
4. Carter, T., Seah, S.A., Long, B., Drinkwater, B., and Subramanian S. 2013. UltraHaptics. *Proceedings of the 26th annual ACM symposium on User interface software and technology - UIST '13*, ACM Press, 505–514.
  5. Colaço, A, Kirmani, A., Yang, H.S., Gong, N., Schmandt, C., and Goyal, V.K. 2013. Mime. *Proceedings of the 26th annual ACM symposium on User interface software and technology - UIST '13*, ACM Press, 227–236.
  6. Delogu, F., Palmiero, M., Federici, S., Plaisant, C., Zhao, H., and Belardinelli, O. 2010. Non-visual exploration of geographic maps: Does sonification help? *Disability & Rehabilitation: Assistive Technology* 5, 3: 164–174.
  7. Ducasse, J., Macé, M., Serrano, M., and Jouffrais, C. 2016. Tangible Reels: Construction and Exploration of Tangible Maps by Visually Impaired Users. *CHI'16 - To appear*, 2186–2197.
  8. Fortune, S. 1987. A sweepline algorithm for Voronoi diagrams. *Algorithmica* 2, 1-4: 153–174.
  9. Goncu, C., Madugalla, A., Marinai, S., and Marriott, K. 2015. Accessible On-Line Floor Plans. *World Wide Web Conference Committee (IW3C2)*, 388–398.
  10. Hart, S.G., and Staveland, L.E. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (eds.). Elsevier, 139–183.
  11. Hill, E.W., and Rieser, J.J. 1993. How persons with visual impairments explore novel spaces : Strategies of good and poor performers. *Journal of Visual Impairment and Blindness* 87: 8–15.
  12. Jetter, H., Leifert, S., Gerken, J., Schubert, S., and Reiterer, H. 2012. Does (multi-)touch aid users' spatial memory and navigation in “panning” and in “zooming & panning” UIs? *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '12*, ACM Press, 83–90.
  13. Kane, S.K, Morris, M. R., Perkins, A. Z., Wigdor, D., Ladner, R. E., and Wobbrock, J. O. 2011. Access overlays: Improving non-visual access to large touch screens for blind users. *Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST '11*, ACM Press, 273–282.
  14. Lawrence, M. M., Martinelli, N., and Nehmer, R. 2009. A Haptic Soundscape Map of the University of Oregon. *Journal of Maps* 5, 1: 19–29.
  15. Marin, G., Dominio, F., and Zanuttigh, P. 2014. Hand gesture recognition with leap motion and kinect devices. *2014 IEEE International Conference on Image Processing (ICIP)*, IEEE, 1565–1569.
  16. McGookin, D., Robertson, E., and Brewster, S. 2010. Clutching at Straws: Using Tangible Interaction to Provide Non-Visual Access to Graphs. *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, ACM Press, 1715–1724.
  17. Paneels, S., Anastassova, M., Strachan, S., Van, S. P., Sivacoumarane, S., and Bolzmacher, C. 2013. What's around me? Multi-actuator haptic feedback on the wrist. *2013 World Haptics Conference (WHC)*, IEEE, 407–412.
  18. Parente, P., and Bishop, G. 2003. BATS : The Blind Audio Tactile Mapping System. *Proceedings of ACM South Eastern Conference*, ACM Press.
  19. Pietrzak, T., Crossan, A., Brewster, S. A., Martin, B., and Pecci, I. 2009. Creating usable pin array tactons for non-visual information. *IEEE Transactions on Haptics* 2, 2: 61–72.
  20. Poppinga, B., Magnusson C., Pielot, M., and Rasmus-Gröhn, K. 2011. TouchOver map: Audio-Tactile Exploration of Interactive Maps. *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI '11*, ACM Press, 545–550.
  21. Profita, H. P., Clawson, J., Gilliland, S., et al. 2013. Don't mind me touching my wrist. *Proceedings of the 17th International symposium on wearable computers - ISWC '13*, ACM Press, 89–96.
  22. Rao, A., and Gordon, A. 2001. Contribution of tactile information to accuracy in pointing movements. *Experimental Brain Research* 138, 4: 438–445.
  23. Rastogi, R., and Pawluk, D. T.V. 2010. Automatic, intuitive zooming for people who are blind or visually impaired. *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility - ASSETS '10*, ACM Press, 239–240.
  24. Rice, M. T., Jacobson, R. D., Gollledge, R. G., and Jones, D. 2005. Design considerations for haptic and auditory map interfaces. *Cartography and Geographic Information Science* 32, 4: 381–391.
  25. Su, J., Rosenzweig, A., Goel, A. , Lara, E. de , and Truong, K. N. 2010. Timbemap: Enabling the Visually-Impaired to Use Maps on Touch-Enabled Devices. *Proceedings of the 12th international*

*conference on Human computer interaction with mobile devices and services - MobileHCI '10*, ACM Press, 17.

26. The Braille Authority of North America. 2010. Guidelines and Standards for Tactile Graphics. Retrieved from <http://brailleauthority.org/tg/web-manual/>
27. Ye, H., Malu, M., Oh, U., and Findlater, L. 2014. Current and future mobile and wearable device use by people with visual impairments. *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, ACM Press, 3123–3132.
28. Zhao, H., Plaisant, C., Shneiderman, B., and Lazar, J. 2008. Data Sonification for Users with Visual Impairment. *ACM Transactions on Computer-Human Interaction* 15, 1: 1–28.

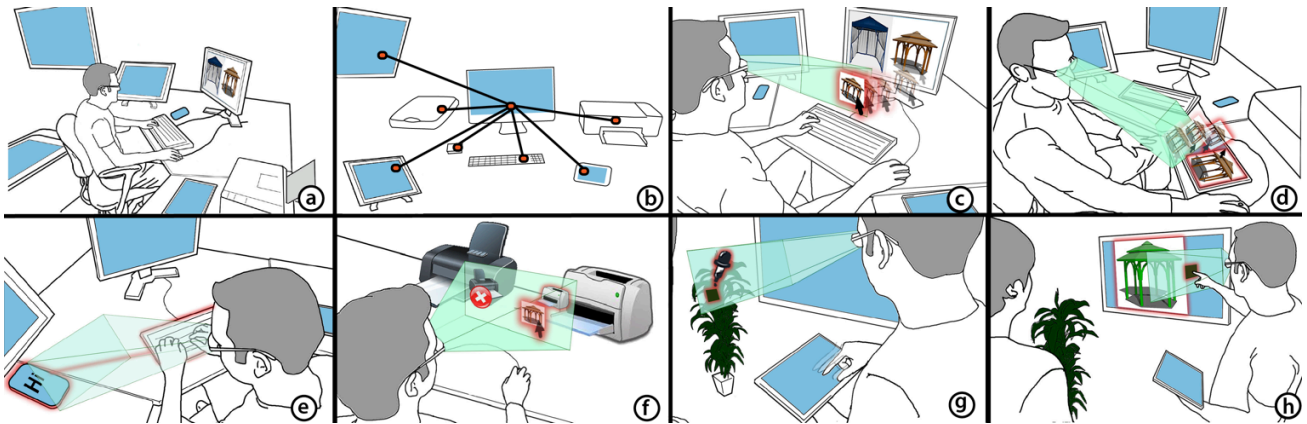
# Gluey: Developing a Head-Worn Display Interface to Unify the Interaction Experience in Distributed Display Environments

Marcos Serrano<sup>1</sup>, Barrett Ens<sup>2</sup>, Xing-Dong Yang<sup>3</sup>, Pourang Irani<sup>2</sup>

<sup>1</sup>IRIT-University of Toulouse  
Toulouse, France, 31062  
marcos.serrano@irit.fr

<sup>2</sup>University of Manitoba  
Winnipeg, MB, Canada, R3T 2N2  
{bens, irani}@cs.umanitoba.ca

<sup>3</sup>Dartmouth College  
Hanover, NH, USA  
xing-dong.yang@dartmouth.edu



**Figure 1.** In this scenario, an architect uses Gluey in his daily work routine (a). He first registers devices and creates a spatial map using Gluey (b). He copies a sketch from his desktop screen to Gluey’s always-available display (c), and pastes it to a tablet (d). With Gluey’s field-of-view tracking feature, he can redirect keyboard input to his phone (e) or select a printer (f). Gluey’s unified input capability allows him to pick a real-world color with a touchscreen (g) and apply it to his sketch using a touchscreen (e).

## ABSTRACT

Distributed display environments (DDEs) allow use of various specialized devices but challenge designers to provide a clean flow of data across multiple displays. Upcoming consumer-ready head-worn displays (HWDs) can play a central role in unifying the interaction experience in such ecosystems. In this paper, we report on the design and development of Gluey, a user interface that acts as a ‘glue’ to facilitate seamless input transitions and data movement across displays. Based on requirements we refine for such an interface, Gluey leverages inherent head-worn display attributes such as field-of-view tracking and an always-available canvas to redirect input and migrate content across multiple displays, while minimizing device switching costs. We implemented a functional prototype integrating Gluey’s numerous interaction possibilities. From our experience in this integration and from user evaluation results, we identify the open challenges in using

HWDs to unify the interaction experience in DDEs.

## Author Keywords

Head-Worn Display; Distributed displays; Multi-display environments; Input redirection; Content migration.

## ACM Classification Keywords

H.5.2. Information interfaces and presentation: Interaction

## INTRODUCTION

A new generation of light-weight, see-through head-worn displays (hereafter referred to as ‘HWDs’) is emerging for general purpose use (e.g. Google Glass, Meta, Microsoft HoloLens). These wearable devices will soon co-exist within a larger ecosystem of displays (i.e. desktop monitors, tablets, smartphones) we depend on for daily tasks. Researchers have sought ways to unite multiple displays and input devices into distributed display environments (DDEs) [13,24,30,37]. However, DDEs are commonly afflicted by device switching [11,47] and data transfer costs [37], which can be mitigated by inventive techniques for content migration (e.g. [32,38,43]) and input redirection (e.g. [25,34]). In such display-saturated ecosystems, what unique role can HWDs play beyond simply being an additional display medium?

To this end we implement Gluey, a user interface that acts as a ‘glue’ to facilitate seamless information flow and input redirection across multiple devices (Figure 1). Gluey derives its bonding force from several features unique to

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).  
MobileHCI '15, August 24 - 27, 2015, Copenhagen, Denmark  
© 2015 ACM. ISBN 978-1-4503-3652-9/15/08...\$15.00  
DOI: <http://dx.doi.org/10.1145/2785830.2785838>

HWDs. For example, the embedded cameras and spatial sensors of next generation HWDs allow Gluey to maintain a spatial model of the relative positions of various devices, which can be used to facilitate cross-device interactions [48]. Gluey's spatial model can track which devices the user is looking at, and use this information to move content across devices. The handsless operability of HWDs, allows users to freely interact with other displays using any available input device. Finally, HWDs' transparent displays can 'disappear' and reappear only when needed with implicit sensing of the user's actions.

Our work builds on previous research that envisioned HWDs taking on a pivotal role in DDEs [10,21]. While each of these early systems made valuable contributions, we contribute a holistic set of design requirements to drive forward the use of HWDs in DDEs with a complete system implementation. In this paper we report on the lessons learned during our design and development of Gluey on an Epson Moverio HWD. Using our proof-of-concept prototype, we implement and evaluate a set of interaction techniques for content migration and input redirection.

This work has 3 main contributions: 1) we propose a set of design requirements to leverage inherent HWD capabilities to facilitate seamless content migration and input redirection in DDEs; 2) we introduce and evaluate Gluey, an implementation of this vision; and 3) we articulate a set of open challenges in using HWDs to unify DDEs.

#### LEVERAGING HWD CAPABILITIES IN DDES

Gluey is inspired by Weiser's vision of 'invisible' devices [45]. In his seminal paper, he proposed to that in a highly usable, distributed display environment, the details regarding device usage should become unnoticeable and 'fade' into the background, allowing users to fully attend to their tasks. However, several workplace studies [13,30,37] reveal that, although tasks often span multiple devices [13,30], these devices are far from being 'invisible'; the work required to facilitate inter-operability occupies a significant portion of users' awareness. For instance, from taking a photo and presenting it in a finished document, data is often passed along a long chain of devices using ad-hoc methods [37]. In contrast to these reflections on the current reality, we present a scenario depicting our vision of Gluey, which addresses these issues by using HWDs' unique affordances for 'transparency'.

#### Gluey Scenario

John, an architect, relies on numerous digital devices while juggling between ad-hoc tasks in his daily work (Figure 1a). Most of his drawing takes place on a desktop computer attached to two monitors and a printer. John uses a tablet for sketching and for showing his drawings to clients. He uses a smartphone for his communication needs.

John completes a plan for a gazebo on his desktop computer and prepares to use his tablet for presenting it to a client. After saving his presentation, John moves it with his mouse

to a clipboard on his HWD (1a), so that he can later copy it onto any other device. He then glances at his tablet and uses his mouse to grab the presentation from the HWD clipboard and transfers it to the device that is in the HWD's direct view, i.e. the tablet screen (1b). Meanwhile, his business partner James sends a text message to his smartphone, noting the urgency of their soon to begin meeting. He is able to quickly reply to James, without opening his phone's soft keyboard. He instead uses the desktop keyboard, already near his hands, to type his reply. He does this by turning to look at his phone, which is detected by Gluey's head-mounted camera. As he types, the text gets redirected to the smartphone (Figure 1c). Before leaving to meet his client, John prints a copy of the same presentation he had on the HWD clipboard by dragging its document icon to a printer icon, which comes in view when he looks in the direction of the printer (1d). At his client's office, he is led to a boardroom equipped with a large wall display for presentations. Since John has previously registered this display within Gluey's spatial configuration, he can immediately use his tablet as a trackpad to drag the presentation of his gazebo drawing from his HWD onto the large display. The client discusses different options, and suggests a different color for the gazebo, which matches the leaves of a nearby plant. John glances over to the plant, and selects its deep green color with Gluey (1e), then fills in the gazebo using a finger gesture (1f).

This scenario captures how John seamlessly interleaves many tasks in an ad hoc manner. He minimizes his need to switch input devices and can seemingly migrate content to where it is needed by glancing at devices. While the actual data transfer is happening in the 'cloud', John need only concern himself with simple, intuitive gestures for manipulating abstract representations of objects that suit his immediate needs. Furthermore, he can extract content from his physical surroundings (such as color or captured images) which he copies onto his digital display.

#### Design Criteria

To realize the above scenario a DDE interface would need to fulfill the following design requirements (or DRs), which we ground in both existing literature and our experience with HWDs. A fundamental feature in any DDE system is **Input Redirection** (DR1) [34,44], which allows available input devices to be used across multiple displays. Such input redirection features should facilitate **Content Migration** (DR2) across devices [8,12]. Moreover, these features should be designed to support cross-compatibility between **All Devices** (DR3) in the environment [2,37], including touch gestures and other available input devices, as well as different display types. However, implausible pairings (e.g. a mouse cursor on a printer) should be omitted to avoid confusion. Initiating input redirection and content migration quickly and painlessly requires a robust **Device Registration** (DR4) mechanism [7,36] (i.e. identifying the presence and location of a device). Implicit device registration requires maintenance of a **Spatial**

**Model** (DR5) of the environment [34,36] (i.e. updating the spatial coordinates of registered devices captured via head-tracking, depth sensor and RGB data). However, such implicit methods are prone to false activations and solutions are needed to overcome these while minimizing the need for explicit input. All of the above features will benefit from **Always Available** (DR6) feedback [10,13] as the user's attention often shifts among the available displays. Furthermore, to enable such operations beyond a single instrumented environment, **Mobility** (DR7) is crucial [2,30]. Our Gluey design is largely dedicated toward fulfilling these design requirements for a 'transparent' and mobile, next-generation computing platform.

**STATE OF THE ART**

A long history of research on HWD interfaces has led to the development of advanced platforms for superimposing digital information on the real world [16,17,33]. Whereas such Augmented Reality applications are aimed toward providing the user with context-dependent information, Gluey is designed to enhance the user's interaction experience. Gluey is related to work on Device Composition [2] and draws inspiration from several areas of research on multi-display systems [5,12,20].

**Input Redirection and Content Migration in DDEs**

With input redirection (DR1), one input device is used to control data on multiple displays [44]. An early ubicomp concept, Mouse Anywhere [9], uses spatial tracking to direct mouse input to the display panel nearest to the user. In later techniques, multiple displays are 'stitched' together to form a large distributed workspace [25,34]. In this space, users can migrate content between separate displays using variations of the traditional dragging technique [25,34]. A later comparison of stitching techniques [29] notes that techniques traversing interstitial space face a disadvantage due to lack of visual feedback in this space. Gluey's always available display can be used to mitigate this disadvantage.

In contrast to input redirection, clipboard techniques (DR2) store a temporary instance of an object while transferring it to a new location. This metaphor has been extended for use in DDEs [27]. To make such interaction intuitive, designers can introduce a physical or virtual object to represent the intermediary storage [32]. Other techniques provide visual representation of the copied object [8], support storage of multiple objects on a tabletop [38] or allow migration of data captured from real-world objects [7].

More recently, WatchConnect [23] uses a smartwatch as a central device in DDEs. However a watch is not always available, such as when typing on a keyboard, and lacks the unique benefits of HWDs. For example, HWDs can overlay virtual information on the real world and allow for drag & drop across displays as demonstrated in an early prototype called EMMIE [10]. HWDs are by definition always available (DR6) and have been combined with head, hand and object trackers to manipulate virtual information in DDEs [21]. Gluey's always available, hands-free display is

ideal for visual storage of multiple migratory objects. Using Gluey, users can easily copy and paste items in multiple interleaved operations using head motion.

**Head-orientation and Gaze Based Techniques in DDEs**

Head or gaze orientation can facilitate input redirection or content migration in DDEs. Zhai et al.'s MAGIC technique [48] was one of the earliest systems to exploit users' gaze to coarsely reposition a cursor. Gaze direction has also proven useful for redirecting keyboard input to a specific device at which a user is looking [14]. Initially, head-orientation based content migration techniques were designed to move objects in 3D graphical workspaces [31,34]. Turner et al. [43] also explored several gaze-based techniques for transferring content between displays. In user studies, they found a trade-off between performance and ease of use, with users preferring techniques that demand lower hand-eye coordination. Unlike gaze-tracking systems, which pinpoint the precise location of the user's view [42], Gluey uses head orientation as a proxy for gaze as gaze-modelling studies show that target switching motions can be primarily accounted for by head motion [18,26]. With head orientation Gluey can determine the display currently in use and detect switches between displays.

**Device Registration and Spatial Models in DDEs**

Most DDEs employ external techniques to register and locate devices, such as GPS [21], external cameras [35], or ultrasonic and magnetic sensors [10]. Smartphone cameras have been used to identify displays without help from external devices by recognizing the screen content [8,12]. For example, TouchProjector [8] allows users to drag and drop objects from remote screens through live video image and DeepShot [12] captures the work state on a device and allows that state to be resumed on a different device. Unlike Gluey however, these techniques require the user to hold the device throughout the entire process.

	DR1	DR2	DR3	DR4	DR5	DR6	DR7
	Input Redirection	Content Migration	All Devices	Device Registration	Spatial Model	Always Available	Mobility
Fix & Float [35]	×	✓	×	×	×	×	✓
Sync. Clip. [27]	×	✓	×	×	×	×	✓
Toolspace [31]	×	✓	×	×	×	×	✓
EMMIE [10]	×	✓	×	×	×	✓	✓
MARS [21]	×	×	×	×	×	✓	✓
Aug. Surface[34]	✓	✓	×	✓	✓	×	×
Hybrid UI [36]	✓	×	✓	✓	✓	✓	✓
Touch Proj. [8]	✓	✓	×	×	✓	×	✓
DeepShot [12]	×	✓	×	✓	✓	×	✓
WatchConnect [23]	✓	✓	×	×	×	×	✓
MultiFi [19]	✓	×	×	×	✓	✓	✓
<b>Gluey</b>	✓	✓	✓	✓	✓	✓	✓

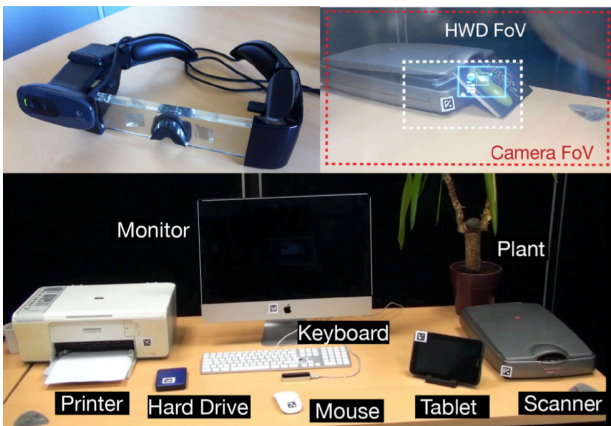
Table 1. Prior work summary in relation to the Design Requirements (DRs) from the literature.

Head direction has been previously used to help determine the best topology for facilitating cursor movement between screen edges [28] and to automatically redirect input across displays based on gaze estimation [1,4]. HWDs have also been used to display the configuration of input devices and objects in DDEs [36] or to extend wearable device interaction capabilities [19]. Gluey builds on these concepts by building a complete spatial model of displays in the environment, but with only the use of the HWD-mounted camera. Gluey also differentiates from previous work by leveraging field-of-view (FoV) tracking from the same camera to implicitly support input redirection and content migration. Table 1 situates this prior work in relation to our holistic set of design requirements.

**GLUEY HARDWARE**

We implemented our prototype on an Epson Moverio BT-100 head-worn display. The display has a resolution of 960×540 pixels, a diagonal FoV of 23°, 24-bit color depth and a perceived image size of 80 inches at 5 meters. The display’s tethered controller runs Android v2.2. In this initial implementation, we did not explore design issues related to binocular depth and used it in monocular mode.

We equipped the HWD with a Logitech C270 HD webcam (Figure 2 top left), chosen for its high resolution and frame rate (1280×720 at 30 fps) with a small footprint (17/45 g without/with cable, 70×18×30mm). The camera’s FoV (52° diagonal) is wider than the limited field of the Moverio, allowing Gluey to track objects beyond the visualization range of the Glueboard (Figure 2 top right).



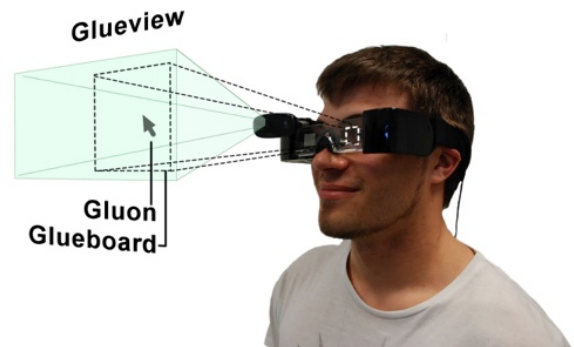
**Figure 2.** HWD equipped with a webcam (top left). First person view of the HWD content and comparison of HWD and

The camera is connected to the desktop computer running Mac OSX v10.7, thus all computer vision tasks are conducted on a single computer, ensuring a high frame rate. The communication between the desktop computer and the HWD was performed using TCP sockets over USB. The HWD runs the TCP server, while the computer and tablet connect as clients at an estimated frequency of 60Hz. Our implementation tested Gluey’s capability in a DDE that includes several devices (Figure 2 bottom).

**GLUEY COMPONENTS**

Gluey incorporates several key components that support simple and intuitive techniques for unifying the interaction experience of distributed display environments. To provide cohesion between otherwise disjoint displays, we exploit the unique features of HWDs such as *view-fixed displays*, *cameras* and inclusion of *spatial sensors* (Figure 3).

We describe the key components of our system using the terms Glueboard, Glueview and Gluon as these define concepts in our implemented system but have slightly different meanings than what might traditionally be considered a ‘clipboard’, ‘head-gaze input’ and ‘input redirection’. The Glueboard allows the user to have an always visible clipboard, which allows content to move across devices, even those that might not be in the user’s vicinity. The term Gluon denotes input from any source, but also goes beyond its primary purpose of facilitating input direction. For example, the Gluon, when aligned with the Glueview, can allow picking colors from physical objects in the environment. This function deviates from the traditional definition of input redirection. We thus felt new terms would define our functionality more accurately.



**Figure 3.** Gluey exploits the unique features of HWDs with three components: Glueview, Glueboard and Gluon. Together, these components enable several novel interaction techniques.

**Glueview**

Many HWDs now include built-in *cameras* and *spatial sensors* (e.g. Meta), from which we can determine a user’s head position in relation to the environment and parse details about what they are viewing. Because head motion dominates over eye movement when switching targets (accounting for 55% of vertical and 85% of lateral motion) [26], we can gain information based on a user’s general *head orientation*, for instance which device they are looking at. With head-direction, we can approximate the functionality of gaze-based systems while avoiding many of their drawbacks, such as the need for additional hardware and difficulties aligning HWD visual content with the real world [16]. We use the term *Glueview* to denote the use of FoV tracking (i.e. head orientation as a proxy for gaze tracking) in combination with Gluey’s spatial model (**DR5**) for enabling implicit device registration (**DR4**): for example by simply having the smartphone in the FoV, the user can link his desktop keyboard and the smartphone.



### Implementation

We register and track the position of surrounding devices with the marker-based ARToolkitPlus. Each tracked object has a unique 2×2 cm marker, allowing identification and tracking in 3D space. During an initiation stage, the Gluey user views each marker and adds it to a 3D spatial model. The tracking application runs at 60 fps; in each frame, the tracker detects all visible markers, maps them to known devices and sends both device id and position to the HWD. Finally webcam coordinates are transformed into HWD coordinates for aligning visual output. Despite the drawbacks of adding a marker to each tracked object and dependency on light conditions, we found this approach sufficient for building our proof-of-concept application in a controlled environment.

### Glueboard

Rather than treat the HWD as just another information display, we can use the device's *view-fixed display* as an always-available 'canvas' for displaying *visual feedback* about interactions in the DDE. The HWD is advantageous for this purpose as the feedback is always-available (**DR6**) even when the user is mobile (**DR7**). For example, this display can show text next to a keyboard, augment devices with useful status information or provide a *visible clipboard* space to store *multiple data objects* in transit between copy/paste operations (**DR2**) to multiple interleaved destinations (**DR3**). We call this combined, always-available feedback and multi-object-clipboard space the *Glueboard*.

### Implementation

During use, the Glueboard is dynamically updated with information on surrounding objects, including status of registered devices (e.g. power on, ready to print) and incoming input data. To display the Glueboard, we set the background color to black to maintain display transparency (black is not visible against the real-world background).

### Gluon

A main limitation of multi-device use is the need to change input modes when switching between devices. Since HWDs do not need to be held, the Gluey user can use any input device at hand, whether it be a keyboard, mouse, a mobile touchscreen or mid-air finger gestures to control multiple displays (**DR1**, **DR2**). The *Gluon* pairs input devices with displays in the DDE to provide a unified interaction experience, independent of input mode or display type. For instance a user can use a desktop keyboard to enter text on a smartphone SMS. Other input modes can also be used to interact with any display or even the real world; for instance a user can select a physical object's color using a tablet as a trackpad. The *Gluon* can be thought of as the visible pointer for various cross-device input modes (e.g. mouse-Gluon, finger-Gluon – see Figure 7), but more so, the Gluon represents the concept of input-agnostic pairing between all available input and display types (**DR3**). This concept is primarily possible as the display on the HWD can be

operated in a hands-free manner, a characteristic that is not easily possible with other types of mobile devices.

### Implementation

The Gluon is implemented according to the following principles to promote usability. First, each input device is associated with only one display at any given time. Second, the Gluon is a singleton, meaning that it can only appear on one display at a time, be it the desktop monitor, the HWD or any another device, such as a tablet or smartphone. Third, the graphical representation of the Gluon pointer is different on the HWD than on other devices to give its current location and delineate transitions between overlapped displays. In our implementation, the computer sends mouse positions, keyboard input and Leap motion gestures at a frequency of 60 Hz.

### GLUEY INTERACTION CAPABILITIES

We illustrate Gluey's three main components through an array of interaction techniques. All features presented hereafter were implemented in our prototype.

### Assorted Content Migration Techniques across Multiple Devices

Gluey facilitates at least three forms of content migration.

1) *Migrating content across multiple displays*: As with prior systems [10,27,32], Gluey facilitates fluid *digital content migration*, for example moving a text document or an image from one display to another. To copy multiple items in Gluey from a variety of devices, we implemented a link between the desktop computer's clipboard and our Glueboard. When the user issues a copy command on the desktop system, the object is added to a list of objects maintained by the HWD and a thumbnail showing its presence appears on the Glueboard (Figure 4). This item can be any object of the file system, such as an image, a video, a pdf, a web link or simply text. For efficiency, we also implemented a feature to allow quick text-entry onto the Glueboard, for replying to interleaving SMS bursts during a task, for example. As the user looks at the keyboard and types, Gluey stores the text in a buffer. When the user looks away, the text appears on the Glueboard for subsequent placement on any device via the Glueview.

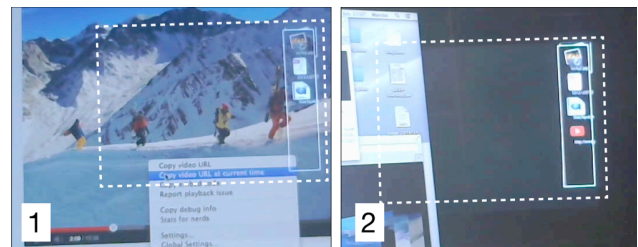
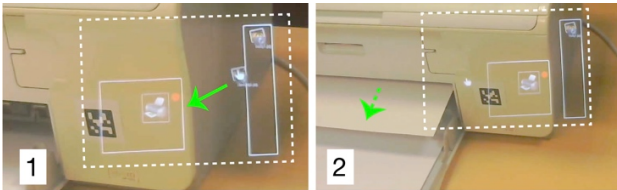


Figure 4. Copying a Youtube link (left) onto the multi-object Glueboard alongside an image, a video and a pdf file (right).

We implemented two methods for moving items from the Glueboard onto other devices. With the mouse-Gluon, the user can simply select and drag the Glueboard item onto a device. The drop region is defined by a rectangle, centered



at the marker (Figure 5). The item is removed from the Glueboard once it is moved onto another device. An alternative solution takes advantage of the mid-air Gluon whereby the user can select an item on the Glueboard via finger movement (detected by our Leap device) and flick it in the direction of the device as per Gluey's spatial model. For instance, the user can look at the desktop display and make a swiping gesture to paste the item to the tablet on the right. In this implementation we cannot differentiate among multiple devices in a given direction, but this could be mapped to flick speed or rely on contextual or recency data.



**Figure 5. Printing an image via drag-and-drop from the Glueboard to an idle printer can mitigate wait queues.**

2) *Migrating content across remote devices*: Through its mobility feature, Gluey enables *remote content migration*, for example by grabbing text from a display and ‘dropping’ it on an available printer or scanner in an adjacent room. This can mitigate delays incurred during interaction with remote devices, such as printers. For example, a user can walk with their HWD to a printer room and determine which devices are operational and jobless before pasting the item onto a printer of choice (Figure 5). Such remote migrations are as fluid as those used for migrating content across devices.

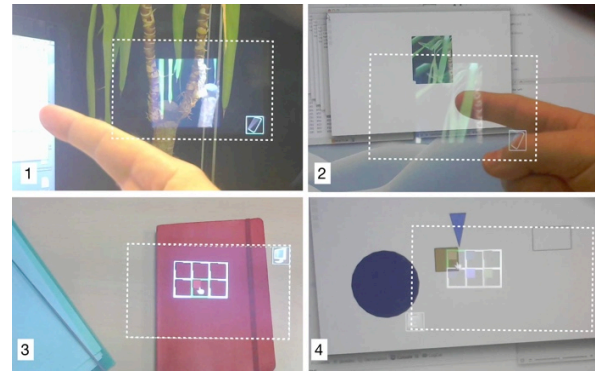
3) *Migrating physical world features onto a display*: Gluey allows capturing features of the user's immediate environment for use on any display, such as in our implemented *image capture* and *color picker* techniques. As the HWD is equipped with a camera, the user can capture an image of content visible through the Glueview. We implemented this feature using the Leap device. A clockwise circular mid-air gesture with the finger grabs the image in the Glueview (Figure 6). The image then appears on the Glueboard and can be transferred to any display as with other digital content. Similar operations are also possible with other input methods; for example, with a mouse-Gluon, the user can lasso a region under the Glueview to select an object in the environment.

Inspired by the Toolglass technique [6], we also implemented a color picker widget that allows the user to choose colors from the user's surroundings. For example, the user can look at a book or a cup and pick its color to use it in a drawing. The color is copied into an available swatch on the Glueboard where it can be later selected with the Gluon. While this approach depends on correct calibration of the HWD camera, it shows Gluey's capability to migrate content from outside the boundaries of the digital

ecosystem. Other forms of capture are also possible, such as selecting text from a book, or capturing faces at a meeting.

#### Input Redirection across Multiple Contexts

Gluey's framework also facilitates diverse input redirection methods. To demonstrate this, we implemented four instantiations of the Gluon, each for very different contexts and applications.



**Figure 6. Top: 1) Taking a snapshot with a finger gesture; 2) pasting the snapshot on a display with a swipe forward gesture; 3) Picking colors from real-world objects. 4) Pasting the colors into graphical items on the computer display.**

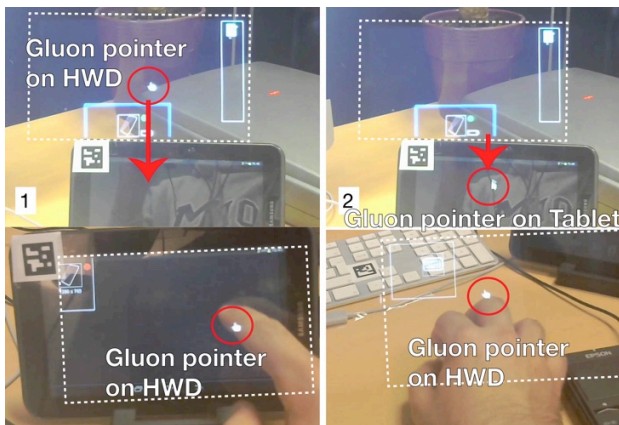
The *mouse-Gluon* allows the user to control any other device with the desktop mouse cursor. This is particularly effective for interleaving tasks between the desktop and other devices or for precise control or selection of objects among all devices. To transfer the mouse-gluon onto the Glueboard, the user aligns the Glueboard to be adjacent to the display on which the mouse cursor exists, such that the two appear ‘stitched’ together. When the mouse cursor is moved beyond the display's boundary it lands on the Glueboard (Figure 7-top). The mouse-Gluon changes appearance to indicate its current location. The *touch-Gluon* replaces the mouse input when the latter is unavailable as in mobile contexts. Using an active touch device, such as a smartphone, the user can invoke the touch-Gluon, in a similar manner as with the mouse-Gluon, to interact with content on other devices, including with content on the Glueboard (Figure 7-bottom). It is also possible to integrate an always-available shortcut, such as a bezel-tap gesture [40] to transfer control from a touchscreen to the HWD.

The *keyboard-Gluon* redirects input from an active keyboard, to any other device. This redirection is triggered by the user's glance at the target display. For instance, while editing a text document on a desktop, the user can turn her head to a tablet to automatically redirect text there, without needing to open its native text-entry mechanism.

The *midair-Gluon* is useful for creating quick ad-hoc links between devices. HWDs equipped with depth cameras can track finger motions in mid-air. We implemented two types of finger gestures, pointing and swiping, using the Leap Motion affixed to a desk. A pointing gesture selects an object that is intersected by an imaginary ray projected

from the pointing finger. For instance, to print a photo, the user points to select it on the desktop and then points at the printer. Swiping is used for moving objects between devices in a given direction, corresponding to relative device locations in Gluey's spatial model.

We implemented two complementary methods to handle false positives, i.e. accidental input redirections. The first method verifies whether the Gluon is compatible with the selected device. For instance the user cannot redirect a mouse to a printer. The second method adds a brief dwell time before starting input redirection. Visual feedback on the HWD informs the user of the time left and allows her to cancel the redirection by shaking her head.



**Figure 7. Mouse-Gluon transfer between stitched Glueboard and tablet display (top); touch-Gluon (bottom left); and mouse-Gluon (bottom right).**

#### Always-Available Visual Feedback

An always-available display, such as the Glueboard can provide at least two forms of visual feedback. The Glueboard can show multiple items that have been copied for transfer to other devices. It can also show feedback about the user's Gluon usage and text-entry. In addition, Gluey can reveal hidden properties and states of devices in the DDE. For example, when the user looks at a device, she can request feedback about the desktop mouse battery life, display resolution of a device or the list of jobs on a printer.

#### Environment's spatial model

To assist with useful operations such as finger pointing or display stitching, Gluey uses the spatial model created during an initial device registration phase. This spatial model captures the position of each device around the user. When the user needs to recall the locations of registered devices, she can view a 2D representation of the spatial layout. The spatial model also allows interaction with out-of-view devices: the user can point at the printer on the left side of the desk without turning her head to send a document. Small icons on the border of the Glueboard indicate the position of nearby devices.

#### PRELIMINARY USER FEEDBACK

We gathered preliminary user feedback with Gluey to examine the potential of its interaction capabilities, and to seek ideas for improving its features and implementations.

#### Participants

We recruited 12 students (1 female, between 21 and 36 years of age) to give feedback on Gluey. Our participants were "naïve" to our implementation and are not actively doing research on HWDs. Only 3 participants had previous experience with the commercial Oculus Rift.

#### Procedure

Participants viewed a video presenting the Gluey prototype, its main components, and interaction capabilities. After showing all interactions participants used the Gluey prototype for roughly 5 minutes for each of the following techniques: device registration; input redirection; content migration (across displays, printing and scanning, color picking and image picking); and out-of-view interactions with the spatial map. For example, for input redirection they were asked to redirect the mouse-Gluon to the HWD and then to a table. After each technique, the experimenter interviewed the participant for roughly 3-5 minutes and recorded his rankings for the technique on a 7-point Likert scale (1: strongly dislike and 7: strongly like). We recorded additional information for each technique: 1) what participants liked and disliked most about a feature; 2) how they would do the same action without Gluey; 3) suggestions on how the current feature could be improved; and 4) the level of distraction resulting from Gluey's information overlay. We collected the answers to the questionnaire and videotaped the entire experiment.

#### Results

On average the experiment took 55 minutes (15 minutes for presenting Gluey and showing the videos, and 40 minutes for testing the prototype). We report the results in terms of general usability and then discuss the impact of our implemented design requirements in Gluey.

#### *General acceptability of a HWD interface for unifying the interaction experience*

Overall, participants' rankings were highly positive. They gave mostly favorable comments on content redirection features (copy and paste on the Glueboard) and input redirection. Seamlessly taking a snapshot and pasting it later on a device was another highly appreciated interaction.

#### *Input Redirection (DR1 & DR4)*

**Forms of input redirection:** All forms of input redirection were ranked high (83% rated redirection favorably, with a score of 5 or higher). Keyboard and the touch redirection were the most highly favored. The keyboard was perceived as "useful to interact with a tablet or smartphone" (p4) and the touch "for using Gluey in a mobile context" (p3). The mouse was liked for its "performance" (p5) and "familiarity" (p11). Some pointed that "transferring the mouse from the PC to Gluey can be tedious" (p4) and that "it is not adapted to all devices" (e.g. tablet) (p2). Some

users preferred using gestures over the mouse as they are “more natural” (p2) and there is “no need to handle an input device” (p6). A suggested improvement for gestures is to “visualize the area where gestures can be performed” and “provide continuous feedback on the gesture itself” (p8).

**Handling false positives:** Users appreciated our two mechanisms to handle false positives. Participants liked the visual feedback, which is “informative” (p6) and “easy to understand” (p9). A possible improvement would be to “set the timer value according to the device or context” (p8).

#### *Content Migration (DR2 & DR4)*

**Copy and paste:** was the most preferred interaction with Gluey. Participants rated it on average a 6.4. The benefits cited by participants include “not needing USB cables or keys” (p9), being able to “replace PCs dedicated towards data transfer” (p5) and its “immediacy” (p4). A possible perceived limitation of this approach is its scalability, i.e. how many items a user can copy (p2). Suggestions for improvement include “using gestures or voice” (p1), “being able to copy several items” and grouping them at once (p8) and “having an undo command” (p6).

**Print:** Most participants (75%) appreciated being able to print using Gluey. Some users pointed out that Wi-Fi printers are gaining popularity and “allow printing from any device” (p5). However, with Gluey, the user can walk up to a row of printers and identify the one that is not busy before issuing the print command. Participants who did not like this feature cited problems such as “turning the head towards the printer” (p8) or that “this task is already easy on PCs” (p6). Interestingly, one user greatly appreciated the feature, explaining that she owns a Wi-Fi printer that does not allow printing from her mobile device. This requires her to manually transfer data to a PC before printing, a problem Gluey was specifically designed to alleviate.

**Physical content migration:** Taking colors from the real world was liked by 83% of the participants. While some of them said color picking was “more interesting for graphic designers” (p10), others saw it as a “tool for replacing the PC color picker” (p6) or as being useful “to find a similarly colored cloth on the Internet” (p5). Furthermore, 83% of participants liked taking snapshots and scanning.

#### *Always available visual feedback (DR6)*

Visual feedback, i.e. seeing the state, position and input interaction of devices was perceived to be useful by 83% of the participants. Participants liked “seeing the state of devices, such as battery status” (p3), “memory left” (p4), “connection” (p4) or “shortcuts while typing” (p5). Most participants also pointed out that such feedback should be disabled at will. One participant suggested showing only a small icon that can be expanded when needed (p2).

**Gluey visual distraction:** from Glueview is considered minimal. Most users (75%) considered the visual distraction acceptable. Some indicated that they ignored distractions after getting used to Gluey (p5). Participants that rated

Gluey as distracting cited mobile contexts (p3), driving (p7) or looking at another display (p6) as situations where they would not want to have a visual overlay. As a solution, one participant suggested providing a simple input to turn off the GlueView at will, such as a tap on the HWD temple.

#### *Mobility (DR7)*

Gluey’s inherent mobility was perceived to be a useful feature. Interestingly, while some participants cited mobility to be useful for work, others favoured its benefits at home. At work Gluey allows “sharing code with a colleague” (p2) or “taking a PowerPoint presentation to a meeting” (p5). At home you can “share content among the different devices in different rooms” (p9), “copy a post-it from my fridge and paste it on my computer” (p10) or “replace my Google Chromecast” (p6).

#### *Spatial Model (DR5)*

Being able to use the spatial model to select an out-of-view device via pointing was identified as a very useful feature. It allows users to “keep the current visual context” and is “more effective” than repeatedly turning the head (p10). A possible improvement is “to add a selection sound to each device” (p8) for improved feedback.

#### *Summary*

Overall, participants were enthused at the concept of Gluey. These results are encouraging taking into account the current limitations of the prototype, such as its weight or the limited display FoV. This exploratory study allowed participants to envision Gluey’s utility and to generate significant feedback for future improvements.

#### **CHALLENGES IN UNIFYING DDES WITH HWDS**

We further discuss our observations on the challenges that need to be addressed for future improvements on Gluey.

*Transition from Prior Multi-Display Use to Gluey.* We probed our participants to describe their routine methods for redirecting information across devices and for minimizing device switching costs. Most users indicated they have adopted a mix of USB memory sticks, email or a cloud service to move data across devices, even if the migration is for a temporary file, i.e. moving a document to the system that is connected to the printer. These ad-hoc solutions are common among multi-device users [37]. Users viewed Gluey as addressing many of their data migration challenges and potentially making it more efficient. An effective HWD-based system should take into account these prior uses to facilitate the transition.

*Unwanted Activations.* Users commented favorably on having the head direction trigger the active display. To avoid unwanted activations, we implemented a false positive mechanism that only redirects to a device after it has been fixed in for a certain period of time, giving time for cancelation. However this solution adds a latency, which can affect the overall performance of the system.

*Visual Overload.* Users highly favored the Glueboard for storing multiple items during migration. However, the

Glueboard's visual feedback mechanism needs to also take into account the possibility for visual overload, although users did not find the Glueboard overly distracting. Gluey should provide users with explicit control over what gets placed on the always-available display, such as preview devices states, or the list of jobs on a printer.

*Data Fragmentation.* Gluey does not fully solve the issue of data fragmentation, although it does provide an intuitive experience for migrating content. Gluey allows transferring files to others' devices and for asynchronous, opportunistic and unplanned copy and paste. Study participants welcomed the idea of seeing a specific document at work get copied on the HWD and taken home for further use. We plan to link Gluey with existing cloud services and study how to integrate it with such approaches.

*Connectivity and Deployment.* For our proof-of-concept system we relied on a simple yet effective TCP/IP stack over Wi-Fi to interconnect devices with the HWD. One value of this star-formation is that any device can communicate via another through the HWD. This can provide a highly modular method of adding new devices to a user's personal display ecosystem. While such an approach sufficed in our exploration, we intend on exploring the effectiveness of new communications and interconnectivity protocols designed for DDEs [22].

Future Gluey versions can be deployed in at least two ways. First, Gluey can behave as a distributed application, similar to current cloud applications such as Dropbox. In this manner it will allow for content migration and simple input redirection (keyboard, gestures-based commands). Alternatively, it can exist as a module of the OS. In this manner it may facilitate a tighter integration of the various Gluon inputs. For instance, Gluey as a module of Android OS could convert input events from non-tactile devices into touchscreen events at the system level.

*Other Technical Challenges.* Other technical considerations that need to be addressed for a final implementation of Gluey are adapting data to the device capabilities (or informing the user if data cannot be used on a specific device) and handling network hiccups (which can affect Gluey's fluid experience).

#### LIMITATIONS AND FUTURE WORK

Our current prototype is limited by the marker-based tracking mechanism, the weight of the device and the resolution of the head-worn display. Our prototype is not fully mobile as the head-mounted camera is powered by our computer. We plan to replace the current webcam with a depth camera to enable tracking gestures [39,46]. System latency also needs to be substantially reduced to provide a smooth and pleasant interaction experience. However, these limitations did not impede our ability to implement and receive user feedback of our proof-of-concept prototype. Also, the limitations of HWD hardware, such as FoV, will improve with newer commercial products. We plan to

further improve the mobility of our system and identify novel interaction scenarios for mobile DDEs [11] as we integrate Gluey to operate with other devices, such as smart watches and tabletops. Another concern for transparent displays is the color blending effects for which we plan to implement in Gluey recently developed color correction methods [41].

#### CONCLUSION

We introduced Gluey, a head-worn software interface that acts as a 'glue' for bonding information across the various displays in a distributed environment. Gluey exploits three key components: Glueview (i.e. using users' *head-direction as a proxy* to detect which device they are looking at), GlueBoard (i.e. using the device's *view-fixed display* as an always-available 'canvas' for displaying visual feedback about interactions) and Gluon (i.e. using the input device at hand seamlessly between displays). We created a proof-of-concept prototype using a HWD equipped with a webcam, guided by a number of design requirements. This allowed us to implement a large set of dedicated interaction techniques that illustrate Gluey's potential, which users commented on favorably through preliminary usage feedback. As the appearance of improved HWDs continues to accelerate, Gluey can be fully realized to resolve many of the device switching and data transfer costs in DDEs.

#### ACKNOWLEDGEMENTS

We thank all our participants for the many valuable suggestions and feedback on this project as well as our lab members for early thoughts and many discussions on iterations of this work. We also thank our reviewers who provided many insightful comments that helped improve the quality of the paper. Finally, we acknowledge an NSERC Discovery Grant awarded to the last author.

#### REFERENCES

1. Ashdown, M., Oka, K., and Sato, Y. Combining head tracking and mouse input for a GUI on multiple monitors. *CHI '05 EA*, 1188-1191.
2. Bardram, J.E. Activity-based computing: Support for mobility and collaboration in ubiquitous computing. *Journal of Personal and Ubiquitous Computing* 9(5), 312-322, 2005.
3. Baur, D., Boring, S., and Feiner, S. Virtual projection: exploring optical projection as a metaphor for multi-device interaction. *CHI '12*, 1693-1702.
4. Benko, H. & Feiner, S. Multi-Monitor Mouse. *CHI '05*, 1208-1211.
5. Bi, X. and Balakrishnan, R. Comparing usage of a large high-resolution display to single or dual desktop displays for daily work. *CHI '09*, 1005-1014.
6. Bier, E., Stone, Maureen., Pier, K., Buxton, W., and DeRose, T. Toolglass and magic lenses: the see-through interface. *SIGGRAPH '93*, 73-80.



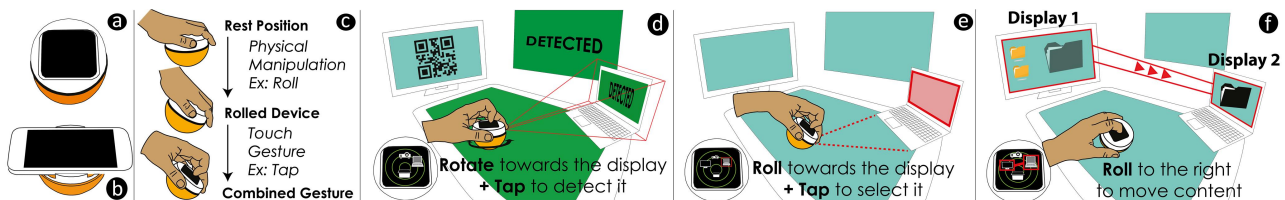
7. Boring, S., Altendorfer, M., Broll, G., Hilliges, O. and Butz, A. Shoot & Copy: Phonedcam-based information transfer from public displays onto mobile phones. *Mobility '07*, 24-31.
8. Boring, S., Baur, D., Butz, A., Gustafson, S., and Baudisch, P. 2010. Touch projector: mobile interaction through video. *CHI '10*, 2287-2296.
9. Brumitt, B., Meyers, M., Krumm, J., Kern, A. and Shafer, S. EasyLiving: Technologies for intelligent environments. *HUC '00*, 12-29.
10. Butz, A., Höllerer, T., Feiner, S., MacIntyre, B., and Beshers, C. 1999. Enveloping Users and Computers in a Collaborative 3D Augmented Reality. *IWAR*. IEEE, 35.
11. Cauchard, J., Löchtefeld, M., Irani, P., Schoening, J. Krüger, A., Fraser, M. and Subramanian, S. Visual separation in mobile multi-display environments. *UIST '11*, 451-460.
12. Chang, T.-H. and Li, Y. Deep shot: a framework for migrating tasks across devices using mobile phone cameras. *CHI '11*, 2163-2172.
13. Dearman, D. and Pierce, J. It's on my other computer!: computing with multiple devices. *CHI '08*, 767-776.
14. Dickie, C., Hart, J., Vertegaal, R., and Eiser, A. LookPoint: an evaluation of eye input for hands-free switching of input devices between multiple computers. *OZCHI '06*, 119-126.
15. Ens, B., Finnegan, R. and Irani, P. The Personal Cockpit: A spatial interface for effective task switching on head-worn displays. *CHI '14*, 3171-3180.
16. Feiner, S., Macintyre, B. and Seligmann, D. Knowledge-based augmented reality. *Communications of the ACM*, 53-62, 1993.
17. Fitzmaurice, G.W. Situated information spaces and spatially aware palmtop computers. *Communications of the ACM*, 39-49, 1993.
18. Freedman, E. G., and Sparks, D. L. Coordination of the eyes and head: movement kinematics. *Exp. Brain Res.* 131(1), 22–32, 2000.
19. Grubert, J., Heinisch, M., Quigley, A. and Schmalstieg, D. MultiFi: Multi-Fidelity Interaction with Displays On and Around the Body. *CHI'15*. ACM, to appear.
20. Grudin, J. Partitioning digital worlds: focal and peripheral awareness in multiple monitor use. *CHI '01*, 458-465.
21. Höllerer, T., Feiner, S., Terauchi, T., Rashid, G. Hallaway, D. 1999. Exploring MARS: Developing Indoor and Outdoor User Interfaces to a Mobile Augmented Reality System. In *Computers and Graphics*. Vol 23, 779-785.
22. Houben, S., Nielsen, S., Esbensen, M, and Bardram, J.E. Noosphere: An activity-centric infrastructure for distributed interaction. *MUM '13*.
23. Houben, S. and Marquardt, N. WatchConnect: A Toolkit for Prototyping Smartwatch-Centric Cross-Device Applications. *CHI'15*. ACM, to appear.
24. Hutchings, D.R., Stasko, J. and Czerwinski, M. Distributed display environments. *Interactions*, 12(6), 50-53, 2005.
25. Johanson, B., Hutchins, G., Winograd, T., and Stone, M. PointRight: experience with flexible input redirection in interactive workspaces. *UIST '02*, 227-234.
26. Kim, K.-H., Reed, M.P. and Martin, B.J. A model of head movement contribution for gaze transitions, *Ergonomics*, 53(4), 447-457.
27. Miller, R.C. and Myers, B.A. Synchronizing clipboards of multiple computers. *UIST '99*, 65-66.
28. Nacenta, M., Sallam, S., Champoux, B., Subramanian, S., and Gutwin, C. Perspective cursor: Perspective-based interaction for multi-display environments. *CHI '06*, 289-298.
29. Nacenta, M, Mandryk, R., and Gutwin, C. Targeting across displayless space. *CHI '08*, 777-786.
30. Oulasvirta, A. and Sumari, L. Mobile kits and laptop trays: Managing multiple devices in mobile information work. *CHI '07*, 1127-1136.
31. Pierce, J., Conway, M., Dantzich, M., and Robertson, G. 1999. Toolspaces and glances: storing, accessing, and retrieving objects in 3D desktop applications. *I3D '99*. ACM, 163-168.
32. Rekimoto, J. Pick-and-drop: a direct manipulation technique for multiple computer environments. *UIST '97*, 31-39.
33. Rekimoto, J. and Nagao, K. The world through the computer: Computer augmented interaction with real-world environments. *UIST '95*, 29-36.
34. Rekimoto, J. and Saitoh, M. Augmented surfaces: A spatially continuous workspace for hybrid computing environments. *CHI '99*, 378-385.
35. Robertson, G. and Card, S. Fix and float: object movement by egocentric navigation. In *Proc. of UIST '97*, 149-150.
36. Sandor, C., Olwal, A., Bell, B., and Feiner, S. 2005. Immersive Mixed-Reality Configuration of Hybrid User Interfaces. In *Proc. of ISMAR '05*. IEEE, 110-113.
37. Santosa, S. and Wigdor, D. A field study of multi-device workflows in distributed workspaces. *UbiComp'13*, 63-72.
38. Schmidt, D., Sas, C. and Gellersen, H. Personal Clipboards for individual copy-and-paste on shared multi-user surfaces. *CHI '13*, 3335-3344.
39. Serrano, M., Ens, B. and Irani, P. Exploring the use of hand-to-face input for interacting with head-worn displays. *CHI '14*, 3181-3190.

40. Serrano, M., Lecolinet, E. and Guiard, Y. Bezel-Tap gestures: Quick activation of commands from sleep mode on tablets. *CHI '13*, 3027-3036.
41. Sridharan, S.K., Hincapié-Ramos, J.D., Flatla, D.R. and Irani, P. Color correction for optical see-through displays using display color profiles. *VRST'13*, 231-240.
42. Stellmach, S. and Dachsel, R. Still looking: Investigating seamless gaze-supported selection, positioning, and manipulation of distant targets. *CHI '13*, 285-294.
43. Turner, J., Alexander, J., Bulling, A., Schmidt, D. and Gellersen, H. Eye pull, eye push: Moving objects between large screens and personal devices with gaze and touch. *INTERACT '13*, 170-186.
44. Wallace, J.R., Mandryk, R.L., and Inkpen, K. Comparing content and input redirection in MDEs. *CSCW '08*, 157-166.
45. Weiser, M. The computer for the 21<sup>st</sup> century. *Scientific American*, 94-104, 1991.
46. Wilson, A. and Benko, H. 2010. Combining multiple depth cameras and projectors for interactions on, above and between surfaces. *UIST '10*. ACM, 273-282.
47. Yang, X-D., Mak, E., McCallum, D., Irani, P., Cao, X. and Izadi, S. 2010. LensMouse: augmenting the mouse with an interactive touch display. *CHI '10*, 2431-2440.
48. Zhai, S., Morimoto, C., and Ihde, S. Manual and gaze input cascaded (MAGIC) pointing. *CHI '99*, 246-253.

|

# TDome: A Touch-Enabled 6DOF Interactive Device for Multi-Display Environments

Houssein Saidi<sup>1</sup>, Marcos Serrano<sup>1</sup>, Pourang Irani<sup>2</sup>, Emmanuel Dubois<sup>1</sup>  
<sup>1</sup> University of Toulouse – IRIT  
 Toulouse, France  
 {first\_name.last\_name}@irit.fr  
<sup>2</sup> University of Manitoba  
 Winnipeg, Canada  
 Irani@cs.umanitoba.ca



**Figure 1.** TDome combines a small (a) or large (b) touchscreen with a dome-like mouse [23]. TDome supports performing combined gestures (c), i.e. a 6 DOF physical manipulation followed by a touch input. TDome facilitates several common tasks in MDEs, such as d) display registration using its embedded camera, e) device selection and f) cross-display data transfer.

## ABSTRACT

The rapid evolution of multi-display environments (MDEs) has created a vacuum in need of novel input devices to optimize interaction in MDEs. In this paper, we propose TDome, a novel touch-enabled 6DOF input and output device to facilitate interactions in MDEs. TDome offers a private display as output, and multiple degrees of freedom as input by combining touch gestures on the display with physical rotation, roll and translation manipulations of the device. TDome allows versatile interactions that address major MDE tasks, which we illustrate through various proof-of-concept implementations: detect surrounding displays, select one display, transfer data across displays, reach distant displays and perform private interactions. We explore TDome's usability and suitability for MDEs through three user studies. First we explore combined *physical+touch* gestures from which we discard uncomfortable combinations. We experimentally validate their feasibility and come up with a set of 71 combined gestures that can be easily performed and efficiently detected. Finally, we collect user feedback to identify natural mappings between gestures and MDE interactions.

## Author Keywords

Multi-display environments; input device; touch input; rolling device.

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).  
 CHI 2017, May 06-11, 2017, Denver, CO, USA  
 © 2017 ACM. ISBN 978-1-4503-4655-9/17/05...\$15.00  
 DOI: <http://dx.doi.org/10.1145/3025453.3025661>

## ACM Classification Keywords

H.5.2. Information interfaces and presentation (e.g., HCI): Interaction.

## INTRODUCTION

Multi-displays environments (MDEs), combining vertical and horizontal displays of varying shapes and forms, have shown significant value for interacting with heterogeneous data sources and in multiple contexts such as 3D exploration [7], collaborative scenarios [12], crisis management [14] and scientific data visualisation [33]. MDEs offer numerous advantages for organizing information across displays, for enhancing individual and group work, for providing support to peripheral information and for extending the interaction space. The incidental emergence of MDEs has resulted in a device vacuum: to our knowledge no device has been specifically implemented for optimizing interactions in such spaces.

Researchers have mainly proposed adapting existing devices to tackle individual MDE tasks, such as the mouse for multi-monitor pointing [5], or smartphones for cross-display data-transfer or distant pointing [11,22]. However such adaptations can result in undesirable side effects: mice are not appropriate when the user is standing [22] and smartphones held in mid-air can be tiring and cumbersome for long interactions [16]. Recent research has demonstrated the use of wearable devices to perform cross-device interactions [18,32]. However, current wearables lack proper input mechanisms and mainly serve private purposes. If MDEs are to become the office of the future, as envisioned by many [26,28], can we design a device specifically tuned for such an environment? Adopting a unique device would avoid the homing effect when switching from one device to another, enhance privacy in such environments through personal data control and visualization, lead to a coherent set of interactions with the



varied MDE applications, and ultimately contribute to a more fluid task flow, a key element in MDEs [4].

To this end we present a novel touch-enabled device, TDome, designed to facilitate interactions and address a range of tasks in MDEs [9,32]. TDome is the combination of a touchscreen, a dome-like Mouse [23] providing 6 DOF, and a camera that can sense the environment. TDome thus inherits properties of other existing mouse-like devices but includes many novel features to tackle the needs of common MDE tasks [9,32]: TDome identifies the spatial layout of displays (Figure 1-d); facilitates distant interaction and data transfer across displays (Figure 1-ef); and enables personal interactions by using the touchscreen as a private output medium. In our work we designed and implemented different techniques employing two versions of TDome (small and large touchscreen) to address these MDE tasks.

In this paper we address two major challenges for applying TDome in MDEs: first, the device's usability, which demands the user to coordinate a physical manipulation with a touch gesture (we refer to as combined gestures - see Figure 1-c); second, the mapping between TDome gestures and MDE tasks. To validate TDome's usability and suitability for MDEs, we conducted three user studies. We first carry out a formative study to discard gestures deemed too uncomfortable. We followed this with a controlled system validation in which we identified the success rate and performance of combined gestures. Results show that users can comfortably and precisely perform 54 combined gestures with TDome when it is equipped with a large touchscreen and 17 when it has a small touchscreen. Finally, using this set of gestures, we collected user feedback on the best mappings from TDome gestures to common MDE tasks. Results reveal that users take benefit of the semantics of TDome gestures to easily map them to certain tasks, such as rolling the device to view private information, or performing a Drag on the touchscreen to send content across displays.

We offer two contributions. (i) We introduce TDome, a novel device aimed directly at facilitating MDE interactions and discuss its properties with regards to MDE common tasks. (ii) We introduce a diverse range of MDE interaction possibilities with our prototypes. To validate these contributions, we identify the most usable TDome interactions. We then obtain qualitative feedback on mappings from TDome features to MDE tasks.

## RELATED WORK

We present a range of techniques designed to support interaction in MDEs. In most cases, a new technique was implemented using an already existing device.

### Early multi-display interaction techniques

While mouse input is suited for interactions with multiple desktop monitors [5], such a device does not adapt well to multi-display environments (MDEs), where displays may be scattered within the physical space [34]. One of the

earliest solutions, Pick'n Drop [27], proposes the use of a stylus to transfer information from one device to another. The stylus has also been used for pointing [20], a particularly difficult task when displays are large or far from the user. However, due to the limited DOF of such devices, these solutions focus on specific interaction techniques but do not cover the broad set of MDE tasks.

### Adapting mobile and wearable devices for MDEs

Performing physical gestures with mobile devices leverages significantly more DOF than those available with existing devices, such as mice. The main reason is that such devices combine a number of sensors that expand the input/output space (e.g. touch, tilt). Examples include the use of mobiles for pointing [22], for continuous map navigation [10,21], for copy-and-paste operations [30] or for interacting with distant 3D content [8]. The camera on the mobile device has been used to transfer data between MDE displays [11,13]. Another common approach is to use mobile devices for multi-display overview+detail tasks [3,7]. However, holding a mobile device in mid-air can be tiring [16], especially as mid-air manipulations lack precision. To overcome this problem, mobile devices can be actuated [31,19], which poses new limitations (robustness, speed).

Researchers have recently proposed the use of wearable devices for MDE operations, such as data transfer [18, 32]. For instance, Gluey [32] is a user interface based on the combination of a head-worn-display with a camera, which facilitates seamless input transitions and data movement across displays. Such devices offer mostly personal capabilities in MDEs.

### Roly-Poly Mouse

A number of multi-DOF input devices have been proposed in the literature. Although they were not specifically designed in the context of MDEs, their capabilities are related to those of TDome. TDome's rounded shape allows tilting the device, similar to the Rockin' Mouse [2] or the VideoMouse [17], and is directly inspired by a more recent device, the Roly-Poly Mouse (RPM) [23]. Using a completely rounded bottom as RPM has been shown to provide larger amplitude of movement than previous tilting devices, and it also enables compound gestures (see Table 1 in [23] for a summary on the differences between RPM and previous tilting and multi-DOF mice).

While the rounded dome-like shape of RPM offers multiple degrees of freedom, it hinders the device's stability. Unintended physical manipulations (e.g. Roll during Translation) are common on devices with such a form factor (cf. study 1 in [23]). In our work we correct this problem by combining physical manipulations with touch gestures, to ensure the robustness of TDome interactions.

The use of a touchscreen coupled with an input device is an approach already adopted in the design of the LensMouse [36]. Other mice have also proposed the use of multi-touch [6]. However, all of them coupled a 2D mouse with a

touchscreen, while the TDome uses a 6D mouse. This brings a new challenge in terms of device usability that we address in this paper through two user studies.

### TDOME OVERVIEW

TDome is a touch-enabled multi-DOF input device that embodies features and a form factor that facilitate MDE interactive tasks. This unique device results from the composition of a touchscreen with a dome-like Mouse, providing rotation, Roll, Translation and Lift-Up motions (6 DOF). The device also includes a camera that can sense the environment. We present an illustrative usage scenario with TDome prior to presenting its features.

### Usage scenario

Harry is an engineer working on a smart campus project that monitors data collected by multiple sensors on the university. To visualize and interact with the large datasets of energy consumption, the university has set up a multi-display environment composed of several displays, a large projection wall and two TDome devices.

As Harry enters the room to start his daily supervision of the energy data, he grabs one TDome and uses it to initialize the multi-display environment by simply pointing at each active display. He then selects the wall projection by rolling the device toward the wall. Harry decides to Spread the data visualisation across two displays: he selects both displays with TDome and transfers the visualizations from one to the other with a TDome gesture. As he wants to look closer at information on the second display, he grabs TDome and walks towards the display, using the device in mid-air to perform a zoom on the data for a closer look.

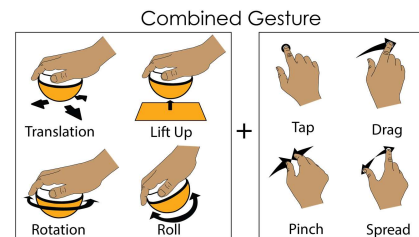
Later that day, Mary enters the room and grabs the second TDome. They have a meeting to explore the university map to mark points of interest. Harry and Mary take their personal smartphones and couple them with each TDome to benefit from personal interactions. Each smartphone shows a personal view of the projected map, which allows them to add and access personal annotations. Before ending, Harry wants to log onto the campus website and upload his annotations: he rolls TDome towards himself to display a virtual keyboard on the device's touchscreen and enter his personal password discreetly on the login page, displayed on the tabletop.

This scenario illustrates how TDome allows users to detect surroundings displays arrangement, select one display, move content between displays, reach to content at distant displays and perform personal interactions on TDome.

### Device Manipulation

Interacting with TDome requires the explicit combination of a physical manipulation with a tactile gesture on the touchscreen. The sequential combination of both actions acts as a delimiter whose accidental activation is unlikely, as demonstrated in our controlled evaluation (below). This approach reduces the risk of issuing a command after performing a physical manipulation inadvertently and

improves the robustness of the device. As illustrated in Figure 2, four different physical manipulations (Translations, Roll, Rotation and Lift-Up) can be combined with four different touch gestures (Tap, Drag, Pinch, Spread) for using TDome.



**Figure 2. TDome allows performing combined gestures, i.e. a physical manipulation followed by a touch gesture**

Initially, we favoured a one-handed interaction where the dominant hand was used to perform the physical manipulation on the device and the touch gestures on the display. But our preliminary tests revealed that some gestures were easier to perform in a bimanual mode, thus extending the touch vocabulary.

### TDome versions: small and large touchscreen

We implemented two design variations of TDome resulting from different device composition alternatives [24]: one with a small touchscreen inserted into the spherical shell (Small version) and one with a larger touchscreen laid on top of the spherical shell (Large version). As these two versions were meant to be complementary, we favoured the possibility of rapidly switching them as opposed to having two separate devices. This opens interesting possibilities, such as switching to the large touchscreen when a larger display is needed.

### IMPLEMENTATION

#### TDome core elements

We used the RPMouse [23] design guidelines to define the dimension of our device: a diameter of 8cm (~ 3.15in) was the easiest to handle and manipulate. As with the original RPMouse, we weighted the device with putty so that the device returns to its initial upright position when released (roly-poly toy principle).

Regarding the touchscreens, we implemented both the Small and Large versions. To restrict our device to the selected size, we had to limit the small screen size to less than 8cm. To create the Small version, we removed the bracelet from an Android smartwatch SimValley AW-414 (63g, 45x44x14mm, 28x28mm touchscreen) and enclosed the smartwatch into TDome. To implement the Large version, we used a Galaxy S4 smartphone (5 in, 134g, 137x70x8mm). We used the smartwatch's camera, which is situated on the edge of the watch, to provide TDome with a horizontal camera view. The camera has a 3MP sensor and a resolution of 1728x1728 pixels.

To support device modularity, the interchange of both touchscreens had to be easy and quick. We thus 3D printed two plastic adaptors that can be adjusted on a 3D printed base: the first one holds the watch while the second one fixes the phone using a magnet (Figure 3). The two plastic adaptors are very rapidly interchangeable. Altogether, the Small version, involving a smartwatch, weighted 207g in total and the Large version, involving a smartphone, weighted 297g. We used TCP sockets over a local Wifi network to connect the watch to the main computer.

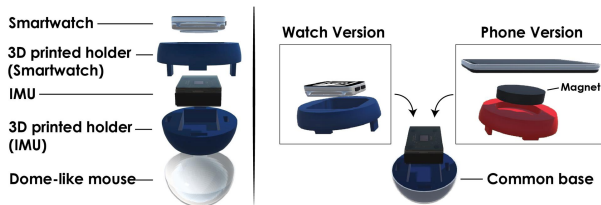


Figure 3. Arrangement of TDome elements for the Small version (left). Both TDome versions are rapidly interchangeable (right).

#### Physical manipulation detection

The spherical shell holds an x-IMU of x-io Technologies (48g,  $57 \times 38 \times 21$  mm) to detect the Roll and Rotation of the device in 3D. The IMU is composed of a triple-axis gyroscope, accelerometer and magnetometer. The refresh rate of the sensors goes up to 512Hz and we used Bluetooth to connect the IMU with the computer. The IMU offered an angular precision of  $1^\circ$ . We 3D printed a holder to fit the IMU in a horizontal position inside the TDome (Figure 3).

To detect the displacement of the device, we used an infrared bezel (Zaagtech, 42") that generated TUIO events. We implemented a filtering process to discard touch events that were detected when fingers touched the surface around the device. Thresholds were also empirically defined to avoid the detection of unwanted Translations, Rolls or Rotations: user's physical manipulations must reach a minimum amplitude to be detected (5cm for Translation,  $30^\circ$  for Roll,  $45^\circ$  for Rotation). Lift-Up was detected as soon as TDome was no longer in contact with the table.

#### SUITABILITY OF TDOME FOR MDES

One of the key characteristics of Multi-Display Environments (MDEs) is their heterogeneous nature. For the majority, MDEs are composed of displays offering different properties (size, orientation, input capabilities), which have led to a set of major interaction requirements [9,32]: *input redirection* (i.e. redirect input channels to different displays), *output redirection* (i.e. move content between displays), *physical relationship* (i.e. possess high-level information on the spatial layout of the displays), *reachability* (i.e. interact with a distant display) and *personal data management* (i.e. personal input and output interaction). TDome properties suit these interaction requirements.

#### Spatial sensing

TDome physical manipulations allow performing 3D pointing in the surrounding space. Combined with the on-board camera, it allows sensing the environment. This can be used to detect and locate nearby displays, creating a spatial layout of the MDE displays represented through a radar-view (*physical relationship*).

#### Input interaction

TDome allows up to 3 types of 2D pointing: by moving the device, by rolling it or by interacting with the touchscreen. These ranges of positioning facilitate *input redirection*. This also offers input that best suits a given display, such as a cursor for precise tasks, or touch input for coarser input.

#### Output interaction

The touchscreen display can be used as a visual buffer to move data among displays in MDEs (*output redirection*). It may also be useful to display a zoomed-in version of a selected area on a distant display (*reachability*). The built-in vibratory capabilities are an alternative to discretely provide the user with private information (*personal data management*).

Through the easy interchange of the Small and Large TDome versions, the user can adopt the most appropriate display for each task; e.g., to visualize large graphs, the user can choose the Large version, but to display the compact radar-view (i.e. a view of the MDE spatial layout), a smaller display is more appropriate (*output redirection*).

#### Mid-air interaction

Two of TDome's physical manipulations (Roll and Rotate) can be used in mid-air, thus facilitating physical displacements to interact with distant displays (*reachability*). It also offers more flexibility to the user to ensure the privacy for some of its tasks (*personal data management*).

#### Form factor

TDome's tilting capabilities facilitate orienting the device towards oneself for private input and output interaction (*personal data management*); and attaching their personal smartphone to TDome's base allows users to access their personal applications and data (*personal data management*).

#### TDOME MDE INTERACTION TECHNIQUES

We now introduce a set of proof-of-concept prototypes illustrating how the previous properties contribute to facilitate interaction in MDEs.

#### Display layout registration

To fulfil the physical relationship and arrangement requirement, we implemented a semi-automatic acquisition of the displays layout in the MDE. This technique allows detecting the displays and building a radar view interface of them, which can be later exploited to interact with the displays of the environment.

During the detection phase, TDome detects a QR code ascribed to each display (better recognition algorithms may not necessitate codes for detection as demonstrated by HuddleLamp [25]). The user orients TDome toward each display successively, so that the device’s on-board camera detects the QR codes (Figure 4 – left). Once the QR code is recognized, the user taps the touchscreen to terminate the identification: the detected display is assigned a position in the environment thanks to the incorporated IMU.



**Figure 4:** TDome’s on-board camera detects displays (left) and creates a radar-view of the spatial layout (center). Then the user can select a display by Rolling + Tapping towards it (right).

The user progressively creates a radar view describing the relative position of all detected displays, with TDome in its center (Figure 4 – center). The user can manually adjust the distance of each display to TDome on the radar view. Once created, the radar view can be used with a Roll + Tap on TDome to select a specific display, by rolling TDome in the direction of the display and tapping on the touchscreen to validate (Figure 4 – right).

**Input redirection**

One recurrent need in MDEs is to manage input redirection. In addition to changing focus from one display to another, TDome offers an input interaction that matches the input possibility to the display it is connected to. TDome can be used as a touch input device through its embedded touchscreen, as a mouse with its translation capability and as a 3D Mouse with its rotation and tilting capabilities depending on the input capability of the display it is redirected to.

For instance, to interact with a map on a distant touchscreen, the user can perform a Roll + Drag on TDome to pan, and a Lift-Up + Pinch on TDome to zoom. Both touchscreen gestures (Drag and Pinch) are the same than what would be used on the distant touch display. While using only TDome’s touchscreen gestures would be possible, using them in combination with the physical gestures (Roll or Translate) ensures a high recognition rate, prevents false positives, as demonstrated by our controlled study presented below and offers additional controls: the Roll angle may impact the panning speed.

**Moving content between displays**

We developed two interaction techniques to move content from one display to another. The Translation + Pinch/Spread technique combines a physical manipulation of TDome to select a display and a gesture on the

touchscreen to grab or place some content on the selected display (Figure 5). In our implementation, Translation + Pinch grabs the application of the screen selected by the translation’s direction, and displays it on the tabletop; while Translation + Spread sends the tabletop application to the screen situated in the translation’s direction. Usually, Pinch and Spread are two well-established gestures for zooming in map or photo applications. Therefore, we plan to design appropriate feedback to avoid any confusion.



**Figure 5:** A Translation + Spread gesture sends the tabletop content (left) to a secondary display (right).

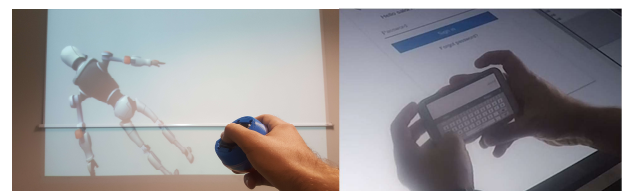
We implemented a second technique using the radar view on TDome to create a virtual information tunnel between two displays (Figure 6-left). The user creates the tunnel by sequentially selecting two displays on the radar view. Once the tunnel is defined, the user can move content along the tunnel with a Roll + Tap on TDome: rolling is performed in the spatial direction of the second display (i.e. a Roll to the right if the display is on the right of the first one); a Tap gesture finalizes the transfer.



**Figure 6:** Using the virtual tunnel technique to transfer information between displays.

**Mid-air interaction**

To support the reachability requirement, TDome provides support to interact with distant displays, i.e. beyond the user's reach. Given the size, shape and wireless design of TDome, the user can physically move to the distant display and perform mid-air interactions with TDome (Figure 7 - left).



**Figure 7:** Illustrating mid-air interaction with TDome (left) and privacy conservation when typing a password (right).

**Password input**

To preserve confidential information, the user can roll the device towards himself or lift the device to visualize and

input content privately. For instance, TDome's large touchscreen can be used as a private virtual keyboard to input a password on a surrounding display (Figure 7 – right). TDome can also be used to visualize a private detailed view of a public context.

#### Other techniques

Beyond effectively supporting essential interactions in MDE, TDome can be used for other common tasks such as controlling a pie menu on a distant display, supporting multi-clipboard copy and paste, and pointing on distant displays. We implemented all these interaction techniques using different combinations of physical manipulations and touch gestures.

#### Resulting challenges for TDome usage

These techniques illustrate how TDome contributes to the execution of relevant interactive situations in MDEs and how it is useful and sufficient to address major MDE interactions. Using a single device contributes to a more fluid interaction in MDEs by maintaining the user in the flow of her activity [4].

Informal tests also provided some early feedback on the importance of precision and on the required number of available gestures: a precise control of the device is important to perform spatial interactions, such as rolling to select a display; and the user requires a wide set of gestures to cover the multiple set of controls and interactions across displays. Therefore, conferring the highest usability level to TDome is essential to ensure MDEs will take full advantage of the device properties.

For these reasons, we first focused on exploring the usability of the device itself. To this end we performed a user experiment dedicated to identifying the set of most precise and robust TDome gestures.

#### EXPLORATORY STUDY ON TDOME GESTURES

The goal of this exploratory study was to inform the implementation of input gestures combining physical manipulations with touch input, by studying only their comfort and collecting initial user feedback. Ultimately we wanted to discard gestures that would be deemed too uncomfortable. While literature on physiology could be anticipatory, it would not help in identifying all the appropriate combinations of wrist gestures and multi-touch finger input. For this reason, we did not want to discard any gesture immediately and ran this exploratory study to reduce the initial gesture design space.

#### Protocol

We carried this exploratory study with 4 participants (all right-handed) from the local university. We instructed participants to manipulate the TDome with their dominant hand. During the experiment they were confronted with the two different versions of TDome (Small and Large). In both settings, they tested three physical manipulations (Roll and Translate in 8 different directions, Rotate in two directions, Lift-Up) in combination with four touch gestures (Tap,

Drag, Pinch, and Spread). Pinch and Spread gestures being more complex to perform, participants repeated these gestures twice: once with the dominant hand and once with the non-dominant hand (e.g. in a bi-manual setting).

Participants performed 2 TDome versions  $\times$  19 physical manipulations  $\times$  6 touch gestures = 228 combined gestures per participant. We asked participants to repeat each combined gesture 3 times, i.e. each participant performed 684 trials. We asked them to rate each gesture combination from 1 (comfortable) to 5 (uncomfortable) to help them verbalize their opinion and comment on their ratings. We report on their qualitative comments.

#### Results

Participants were very positive about performing the following gestures both with the Small and Large versions of TDome:

- Tap and Drag combined with any physical manipulation (Translation, Rotation, Roll or Lift-Up).
- Pinch and Spread in a bi-manual setting (one hand manipulates the rolling device while the other touches the display) when combined with a Translation, Rotation or Lift-Up.

However some other gestures seemed too uncomfortable to be performed:

- Performing Pinch and Spread with a single hand was always deemed very uncomfortable when combined with any physical gestures and for both TDome versions (Small and Large).
- Performing Pinch and Spread in a bi-manual setting in combination with a Roll gesture was perceived to be very uncomfortable.

We decided to remove these uncomfortable gestures (Pinch and Spread with a single hand or in combination with Roll) from our subsequent work.

#### CONTROLLED EXPERIMENT

The goal of this controlled experiment was to validate the feasibility of combined gestures, i.e. physical manipulation + touch input. We hypothesize that certain touch gestures could be difficult to perform on the Small version, on which certain combinations could lead to errors.

#### Combined Gestures

From the previous exploratory study, we decided to use two touch gestures with one hand: Tap and Drag. Gestures using two fingers, i.e. Spread and Pinch, were performed with two hands: one hand held the device while the other performed the touch gesture. These touch gestures were used in combination with a Translation, a Roll, a Rotation and a Lift-Up of TDome.

#### Task

Participants were required to perform each gesture, following a visual indication on a tabletop display. TDome was placed in an initial position at the center of the surface,



indicated by a visual feedback. We let users hold the device as they pleased. We asked participants to perform the gestures as fast as possible with high accuracy. We provided continuous visual feedback indicating the state of the device (position, Roll and Rotation) as well as touch gestures on the display. We provided them with knowledge of result and in case of error we indicated which gesture (physical manipulation and/or touch gesture) had been erroneously performed. Each trial started when the user pressed a button on the tabletop, which displayed the instructions, and ended when a combined gesture had been recognised.

### Participants

We recruited 12 participants (3 female), aged 27.5 years on average (SD=4.89) from the local university. 11 of them were right-handed and 3 of them took part in the exploratory study.

### Apparatus

We used the TDome implementation described earlier. The device was used on a tabletop display (96×72cm) of 102 cm height thus requiring the user to stand during the experiment. We used the display in an area limited to the size of the infrared bezel (42 inches, 1920×1080px).

### Design and protocol

The experiment followed a 2x4x4 within-subjects design, with Display (Small, Large), Physical manipulation (Translate, Roll, Rotate and Lift-Up) and Touch gesture (Tap, Drag, Pinch and Spread) as factors. We did not test the condition combining Roll with Pinch/Spread, as this combination appeared to be highly uncomfortable in our pre-study. We also decided to study one random translation direction to limit the experiment length: previous studies on RPM [36] showed that all translation directions were as easy to perform. For the other physical manipulations, participants performed eight Roll directions and two Rotations (left/right).

Our pre-study also showed that Pinch and Spread gestures seemed more difficult than Tap and Drag. Therefore, we paired Tap with Pinch, and Drag with Spread to balance the different blocks length and difficulty. Trials were grouped in four blocks: one block corresponded to one Display and two touch gestures (Tap/Pinch or Drag/Spread).

The four blocks were counterbalanced across participants using a 4x4 Latin Square. For each block, we ordered touch by difficulty: first Tap or Drag, then Pinch or Spread. For each set of trials corresponding to one touch gesture, the physical manipulations were ordered in a predefined way (Lift-Up, Translation, Roll and Rotation) because a random sorting would have made the instructions difficult to follow. Each combined gesture was repeated three times. Completing the four blocks took approximately 25 minutes.

The study started with a training set made of the same four blocks than in the experiment. The training consisted of 94

trials and took approximately 20 minutes. After the training, each participant performed 192 trials:

- 144 trials for the Tap and Drag: 2 Displays x 12 Physical Manipulations (1 Translation + 8 Rolls + 2 Rotations + 1 Lift-Up) × 2 Touch gestures × 3 repetitions.
- 48 trials for the Pinch and Spread: 2 Displays x 4 Physical Manipulations (1 Translation + 2 Rotations + 1 Lift-Up) × 2 Touch gestures × 3 repetitions.

We collected  $192 \times 12$  participants = 2304 trials in total, which took approximately 45 minutes for each participant.

### Collected Data

We logged all gestures from start to finish. We calculated success rates, completion time from instruction onset to validation, unintended touch gestures on the Display and amplitude of the physical manipulations. We classified errors in three categories according to the gesture that had been erroneously performed: physical, touch or both. Finally, we asked participants to rate each condition on a 1-5 Likert scale on perceived difficulty.

### RESULTS

A Shapiro-Wilk test established that the data was not normal and we could not normalize it. Therefore we used a Friedman test (we report  $\chi^2$  and p) to compare more than 2 conditions, and Wilcoxon tests otherwise (we report p value). Where appropriate, we used a Bonferroni correction. We first discuss the success rate for the Small and Large versions separately as a Wilcoxon test showed a significant effect of Display on success rate ( $p < .001$ ).

#### Success rate: Large version

When using the Large version, we found no significant effect of Touch gestures (Friedman:  $\chi^2=3.87$ ,  $p=0.2$ ) or Physical manipulations (Friedman:  $\chi^2=4.1$ ,  $p=0.2$ ) on success rate. Overall, success rate with the Large version was 94.44%. Errors were distributed among Physical Manipulations (2.52%) and Touch gestures (2.86%).

A Friedman test reveals a significant effect of Touch gesture on success rate when performing a Rotation ( $\chi^2=13.32$ ,  $p=.003$ ): a Wilcoxon test reveals a significant difference between Tap and Drag (81.94% vs. 98.61%;  $p=.022$ ) and between Tap and Pinch (81.94% vs. 97.22%;  $p=.045$ ).

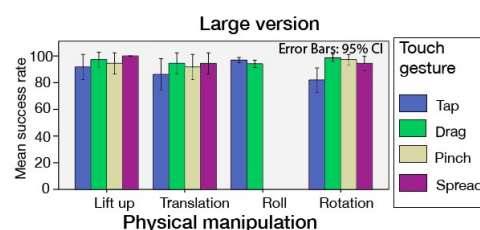
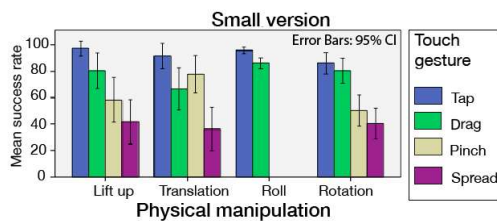


Figure 8. Mean success rate for each combination of Physical manipulation and Touch gesture when using the Large version.

We observed that when instructing participants to perform a Rotation + Tap combined gesture, 91% of the erroneously detected touch gestures are Drag gestures. Performing a Rotation induces a wrist distortion that may affect the user’s ability to precisely tap the display without swiping the finger: this may explain why a Drag is easier to perform than a Tap. Spread and Pinch are not affected by the wrist rotation since they are performed in a bi-manual setting (Figure 8).

**Success rate: Small version**

When using the Small version, a Friedman test revealed a significant effect of Physical Manipulation ( $\chi^2=17.46$ ,  $p<.001$ ) and Touch gestures ( $\chi^2=33.56$ ,  $p<.001$ ) on success rate. We analyse the success rates for each combined gesture, i.e. the combined Physical manipulation and Touch gesture.



**Figure 9. Mean success rate for each combination of Physical manipulation and Touch gesture when using the Small version.**

A Friedman test reveals a significant effect of Touch gestures on success rate when performing a Lift-Up ( $\chi^2=16$ ,  $p=.001$ ), a Translation ( $\chi^2=15.75$ ,  $p=.001$ ), a Roll ( $\chi^2=6.4$ ,  $p=.01$ ) or a Rotation ( $\chi^2=21.6$ ,  $p<.001$ ):

- Lift-Up: a Wilcoxon test reveals a significant difference between Tap and Spread (92.22% vs. 41.67%;  $p=.001$ ) and Tap and Pinch (92.22% vs. 58.33;  $p=.04$ ). Success rate with Drag is 80.56%.
- Translation: a Wilcoxon test reveals a significant difference between Tap and Spread (91.67% vs. 36.11%;  $p=.001$ ) and between Pinch and Spread (77.78% vs. 36.11%;  $p=.02$ ). Success rate with Drag is 66.67%.
- Roll: a Wilcoxon test reveals a significant difference between Tap and Drag (95.83% vs. 86.11%;  $p=.04$ ).
- Rotation: a Wilcoxon test reveals a significant difference between Tap (86.11%) / Drag (80.56%) and Spread (40.28%) / Pinch (50.0%;  $p<.02$ ).

**Completion time**

A Wilcoxon test shows no difference between the Small and the Large versions ( $p=.08$ ). Overall, it took participants 2.5 seconds to perform a combined gesture. While we found some differences across gesture combinations, all of them are compatible with the micro-interactions concept [1], i.e. fast interactions that take less than 4s completion: all times ranged between 2.1s and 2.7s.

**Unintentional touches**

We recorded unintended touches on the Small and Large versions. Overall, results were similar for both versions: we detected unintentional touches in 2% of the trials. These touches did not necessarily raise errors. The sequential use of a touch interaction after a physical gesture prevents from launching a command unintentionally.

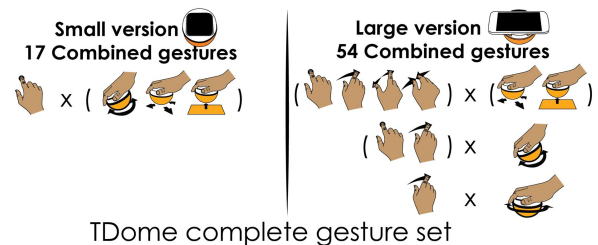
**Subjective feedback**

When considering the physical manipulations, results show that with the Small version, more than 50% of the participants found easy or very easy (4 or 5 on Likert scale) to perform a combined gesture involving a Roll, Translation or Lift-Up. In the case of the Large version, more than 75% of participants rated these gestures as easy or very easy.

When considering the touch gestures, we observed that with the Small version more than 50% of participants found difficult or very difficult (1 or 2 on Likert scale) to perform a combined gesture involving a Spread or Pinch. With the Large version, 60% or more of the participants found easy or very easy to perform combined gestures involving any kind of touch gesture.

**Summary**

Results reveal differences between the Small and Large versions (Figure 10). With the Small version, the experiment reveals that 17 combined gestures can be comfortably and efficiently performed: those based on the combination of a Roll (8 directions), a Translation (8 directions) or a Lift-Up with a Tap gesture (with a success rate of 95.83%, 91.67% and 92.22% respectively).



**Figure 10. Summary of the 17 (Small) + 54 (Large) combined gestures which offer a good usability and performance.**

With the Large version, the experiment reveals that 54 combined gestures can be comfortably and efficiently performed: 16 results from the combination of a Roll (95.49% success rate) with Tap or Drag gesture, 36 result from the combination of a Translation (91.67% success rate), or Lift-Up (95.83% success rate) with one of the four touch gestures (Tap, Drag, Pinch, Spread) and 2 results from the combination of a Rotation with a Drag (98.61% success rate).

**MAPPING TDOME GESTURES TO MDE TASKS**

We elicited user input through a user study to explore how the selected set of gestures from our previous experiment can be mapped to MDE interactive tasks.



**Overview and rationale**

We asked users to choose, for each TDome task, one gesture from the set of gestures selected in the controlled experiment.

**MDE tasks considered**

From our scenarios, we considered the following 7 tasks: pointing on a distant display; zooming on a distant display; displacing a window from one display to another (horizontal tunnel, vertical tunnel); sending a window from Tabletop/TDome (user position) to a distant display and vice-versa; selecting an icon on the radar view; panning and zooming a focused view of a distant context; and typing on a private keyboard.

**Participants**

12 (1 female) students and researchers from the local university volunteered for this study. They were aged 31.9 years on average (SD=9). Five of them took part in the previous studies.

**Procedure**

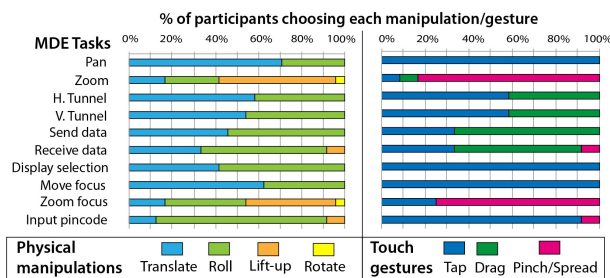
Participants were given the two TDome versions (Small and Large) and were situated in an MDE environment comprised of a tabletop, 4 displays and 1 video-projection. We familiarized our participants with TDome capabilities by showing them a video illustrating the combined gestures (without showing any interactive task). For each combined gesture, we asked participants to perform it themselves with both versions of TDome. Then, we asked participants to select and justify, for each task and each TDome version, which gesture they preferred. The session took about 15 minutes.

**Collected data**

Every user generated one sheet with a summary of the gestures chosen for each task and TDome version. We recorded users' verbal comments.

**Results**

Amongst all available combined gestures, only one was never used in our study (Lift-Up + Drag). Overall, participants took advantage of the gestures diversity to match the different tasks. The agreement scores [36] of the combined gestures (Physical manipulation + touch gestures) range between 0.3 and 0.6. These scores are in line with previous studies [36]. For the sake of clarity, we detail separately for each task the choice of physical manipulations and touch input.



**Figure 11. Percentage of users that chose each physical gesture on both versions of TDome (left) and each touch gesture on the Large version (right) for MDE tasks.**

*Physical manipulations*

Our results were similar for the Small and Large versions concerning which physical gesture to use. Thus we report both results together (i.e. 24 gestures per task).

Two physical gestures were used more often: Translate and Roll (Figure 11-left). For some tasks, one was preferred over the other: *Translation* for panning (17/24), or for moving a focus (15/24); *Roll* for private pincode input (19/24). For other tasks, such as redirecting data using the tunnel, output redirection or display selection, there was no clear preference for one of these two gestures.

The Lift-Up gesture was used for zooming 13 times (i.e. lifting up the device activates zoom mode). Rotation was used only one time for each of our zooming tasks.

*Touch gestures*

While only the Tap gesture is feasible on the Small version, users selected different gestures on the Large version according to the task (Figure 11-right). For instance, Pinch and Spread were preferred for zooming (10/12), and Drag was preferred for sending content from the tabletop to other displays (8/12). Taping was the preferred gesture for map panning (12/12), display selection (12/12) or pincode input (11/12).

**DISCUSSION AND PERSPECTIVES**

**TDome benefits**

We presented two versions of TDome: a Small version with an integrated touchscreen and a Large version based on attaching a smartphone. TDome's unique features offer several advantages to interact with MDEs:

- TDome supports performing multiple combined gestures involving a physical manipulation of the device (Translation, Roll, Rotation or Lift-Up) followed by a touch gesture on the touchscreen (Tap, Drag, Pinch or Spread). Such a combination prevents from unintended activations due to parasite touches on the touchscreen;
- The combined gestures from our final set represent good candidates to support rapid access to interactive commands;
- The two TDome versions are easily interchangeable and offer complementary functionalities: the Small version is useful to rapidly launch shortcuts, while the Large version offers a larger display area and supports multi-touch gestures (Pinch and Spread);
- Rolling TDome can be used to interact with multi-dimensional data through continuous physical gestures, as demonstrated earlier [23];
- The embedded display can be used as a personal display area to augment output visualisation, such as in overview + detail techniques [29];

- The embedded display can also be used to show feedback of the TDome interactions, such as displaying a copied object for the copy/paste technique.

#### Guidelines for mapping TDome gestures to MDE tasks

We propose a set of guidelines to map TDome gestures to MDE tasks based on our mapping study as well as on our experience developing TDome interaction techniques:

- TDome offers a diversity and large number of possible gestures of which users can take advantage as illustrated in our mapping study. Some of these gestures have natural mappings with MDE tasks, such as Rolling towards oneself to display private information, Pinch and Spread for zooming or dragging for sending data to another display. Appropriately combined with a physical manipulation or a touch gesture, these should become the “default” combined gestures with TDome on MDEs.
- While some mappings are obvious and have a large consensus, others are sometimes split between two types of gestures (usually Roll or Translation): this suggests that using TDome in an MDE could benefit from a device personalization step wherein the user defines his preferred mapping, especially for output redirection.
- Interacting in an MDE involves system tasks (i.e. tasks related to the environment, such as display selection) and application tasks. Since these tasks could be assigned to the same TDome gestures, there is a need for a mode switching gesture. The Lift-Up manipulation combined with touch input, is a good candidate as it was considered for switching between pan and zoom tasks in our mapping study

#### Memorability of a large number of gestures

TDome offers a large set of usable combined gestures. This diversity of available controls is particularly relevant to tackle tasks in MDEs. However, increasing the amount of controls might make them hard to memorize. The physical nature of these combined gestures can help cluster them according to the physical manipulation, as shown in our mapping study. Further experiments are required to identify how such clustering can improve gestures or command memorization.

#### Collaboration

MDEs are naturally designed to support collaboration. We can envision multiple TDome like devices, each controlled by the MDE’s users. However, in such cases input and output redirection mechanisms would need to be effectively controlled. Control mechanisms have been proposed by others [14] to handle synchronization, locking and input conflicts, and future iterations of TDome will adapt or build on such proposals.

#### Future work

In this paper we focused on the suitability of TDome capabilities for MDE tasks and the feasibility of its combined gestures. TDome interaction techniques still need

to be fine tuned and future work should compare their performance with a baseline for each MDE task. Theoretically, since TDome integrates the same capabilities as existing MDE devices, we hypothesise that it can perform similarly for each individual MDE task. For instance, TDome can perform translations like a mouse, and has the same touch and mid-air capabilities as a smartphone. Moreover, since TDome is a unique device that supports a range of core MDE tasks, it should improve the overall performance by reducing homing transition times and promoting the interaction flow. Therefore, beyond individual controlled comparisons, it would be interesting to carry a longitudinal study. We leave these studies for future work.

Beyond these aspects, we plan to focus on user expertise of TDome techniques: most menu or command techniques consider novice and expert modes as well as the transition from novice to expert [15]. In our work we focused on how the combined gestures are performed. It will be interesting to design techniques that efficiently support both novice and expert users and the transition from one group to the other, as done with the Marking Menus [37].

We also plan to investigate the extension of our physical manipulation gestures by adding thresholds. For instance, each Roll gesture could launch two different commands according to the Roll amplitude (under or over 42° according to our study results). Technical alternatives also need to be investigated to replace the infrared bezel used to detect the TDome translations. We are currently exploring the application of conductive paint on the external surface of TDome, which will allow using the device on any capacitive surface.

#### CONCLUSION

We presented TDome, a device designed for interactions in MDEs. We designed two TDome prototypes: a Small version with an integrated touchscreen and a Large version based on attaching a smartphone. We explored combined gestures involving a physical manipulation (Translation, Roll, Rotation or Lift-Up) followed by a touch gesture (Tap, Drag, Pinch or Spread) through a 3-steps process. First, an exploratory study focusing on comfort established that 60 combined gestures could be comfortably performed. Second, a controlled experiment evaluated the user's performance as well as the subjective perceived difficulty. Results revealed that the number of gestures that can be precisely and easily performed is 17 with the Small version, and 54 with the Large version. Finally, a user survey explored the mappings between these gestures and MDE tasks. Results show that some combined gestures are more prone to be used in specific tasks than other. In general, we find participants are able to match TDome features to MDE tasks

**ACKNOWLEDGEMENTS**

This work is funded by the neOCampus project of the University Paul Sabatier (Toulouse, France) and part of the AP2 project (research grant AP2 ANR-15-CE23-0001).

**REFERENCES**

1. Daniel L. Ashbrook. 2010. *Enabling Mobile Microinteractions*. Ph.D. Dissertation. Georgia Institute of Technology, Atlanta, GA, USA. Advisor(s) Thad E. Starner. AAI3414437
2. Ravin Balakrishnan, Thomas Baudel, Gordon Kurtenbach, and George Fitzmaurice. 1997. The Rockin'Mouse: integral 3D manipulation on a plane. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems (CHI '97)*. ACM, New York, NY, USA, 311-318. DOI=<http://dx.doi.org/10.1145/258549.258778>
3. Dominikus Baur, Sebastian Boring, and Steven Feiner. 2012. Virtual projection: exploring optical projection as a metaphor for multi-device interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1693-1702. DOI=<http://dx.doi.org/10.1145/2207676.2208297>
4. Benjamin B. Bederson. 2004. Interfaces for staying in the flow. *Ubiquity* 2004, September (September 2004), 1-1. DOI=<http://dx.doi.org/10.1145/1074068.1074069>
5. Hrvoje Benko and Steven Feiner. 2005. Multi-monitor mouse. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05)*. ACM, New York, NY, USA, 1208-1211. DOI=<http://dx.doi.org/10.1145/1056808.1056878>
6. Hrvoje Benko, Shahram Izadi, Andrew D. Wilson, Xiang Cao, Dan Rosenfeld, and Ken Hinckley. 2010. Design and evaluation of interaction models for multi-touch mice. In *Proceedings of Graphics Interface 2010 (GI '10)*. Canadian Information Processing Society, Toronto, Ont., Canada, Canada, 253-260.
7. Louis-Pierre Bergé, Marcos Serrano, Gary Perelman, and Emmanuel Dubois. 2014. Exploring smartphone-based interaction with overview+detail interfaces on 3D public displays. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services (MobileHCI '14)*. ACM, New York, NY, USA, 125-134. DOI: <http://dx.doi.org/10.1145/2628363.2628374>
8. Louis-Pierre Bergé, Emmanuel Dubois, and Mathieu Raynal. 2015. Design and Evaluation of an "Around the SmartPhone" Technique for 3D Manipulations on Distant Display. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction (SUI '15)*. ACM, New York, NY, USA, 69-78. DOI=<http://dx.doi.org/10.1145/2788940.2788941>
9. Boring, S. 2007. Interacting in Multi-Display Environments. In *DC Ubicomp'07*. ACM.
10. Sebastian Boring, Marko Jurmu, and Andreas Butz. 2009. Scroll, tilt or move it: using mobile phones to continuously control pointers on large public displays. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7 (OZCHI '09)*. ACM, New York, NY, USA, 161-168. DOI=<http://dx.doi.org/10.1145/1738826.1738853>
11. Sebastian Boring, Dominikus Baur, Andreas Butz, Sean Gustafson, and Patrick Baudisch. 2010. Touch projector: mobile interaction through video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 2287-2296. DOI=<http://dx.doi.org/10.1145/1753326.1753671>
12. Christophe Bortolaso, Matthew Oskamp, T.C. Nicholas Graham, and Doug Brown. 2013. OrMiS: a tabletop interface for simulation-based training. In *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces (ITS '13)*. ACM, New York, NY, USA, 145-154. DOI=<http://dx.doi.org/10.1145/2512349.2512792>
13. Tsung-Hsiang Chang and Yang Li. 2011. Deep shot: a framework for migrating tasks across devices using mobile phone cameras. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 2163-2172. DOI=<http://dx.doi.org/10.1145/1978942.1979257>
14. Olivier Chapuis, Anastasia Bezerianos, and Stelios Frantzeskakis. 2014. Smarties: an input system for wall display development. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2763-2772. DOI=<http://dx.doi.org/10.1145/2556288.2556956>
15. Andy Cockburn, Carl Gutwin, Joey Scarr, and Sylvain Malacria. 2014. Supporting Novice to Expert Transitions in User Interfaces. *ACM Comput. Surv.* 47, 2, Article 31 (November 2014), 36 pages. DOI: <http://dx.doi.org/10.1145/2659796>
16. Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. 2014. Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 1063-1072. DOI: <http://dx.doi.org/10.1145/2556288.2557130>
17. Ken Hinckley, Mike Sinclair, Erik Hanson, Richard Szeliski, and Matt Conway. 1999. The VideoMouse: a camera-based multi-degree-of-freedom input device. In *Proceedings of the 12th annual ACM symposium on User interface software and technology (UIST '99)*. ACM, New York, NY, USA, 103-112. DOI=<http://dx.doi.org/10.1145/320719.322591>

18. Steven Houben and Nicolai Marquardt. 2015. WatchConnect: A Toolkit for Prototyping Smartwatch-Centric Cross-Device Applications. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1247-1256. DOI: <http://dx.doi.org/10.1145/2702123.2702215>
19. Chang-Min Kim and Tek-Jin Nam. 2015. G-raff: An Elevating Tangible Block for Spatial Tabletop Interaction. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 4161-4164. DOI: <http://dx.doi.org/10.1145/2702123.2702444>
20. Miguel A. Nacenta, Dzmitry Aliakseyeu, Sriram Subramanian, and Carl Gutwin. 2005. A comparison of techniques for multi-display reaching. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. ACM, New York, NY, USA, 371-380. DOI=<http://dx.doi.org/10.1145/1054972.1055024>
21. Mathieu Nancel, Julie Wagner, Emmanuel Pietriga, Olivier Chapuis, and Wendy Mackay. 2011. Mid-air pan-and-zoom on wall-sized displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 177-186. DOI=<http://dx.doi.org/10.1145/1978942.1978969>
22. Mathieu Nancel, Olivier Chapuis, Emmanuel Pietriga, Xing-Dong Yang, Pourang P. Irani, and Michel Beaudouin-Lafon. 2013. High-precision pointing on large wall displays using small handheld devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 831-840. DOI: <http://dx.doi.org/10.1145/2470654.2470773>
23. Gary Perelman, Marcos Serrano, Mathieu Raynal, Celia Picard, Mustapha Derras, and Emmanuel Dubois. 2015. The Roly-Poly Mouse: Designing a Rolling Input Device Unifying 2D and 3D Interaction. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 327-336. DOI: <http://dx.doi.org/10.1145/2702123.2702244>
24. Gary Perelman, Marcos Serrano, Mathieu Raynal, Celia Picard, Mustapha Derras, and Emmanuel Dubois. 2016. DECO: A Design Space for Device Composition. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*. ACM, New York, NY, USA, 435-446. DOI: <http://dx.doi.org/10.1145/2901790.2901893>
25. Roman Rädle, Hans-Christian Jetter, Nicolai Marquardt, Harald Reiterer, and Yvonne Rogers. 2014. HuddleLamp: Spatially-Aware Mobile Displays for Ad-hoc Around-the-Table Collaboration. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces (ITS '14)*. ACM, New York, NY, USA, 45-54. DOI=<http://dx.doi.org/10.1145/2669485.2669500>
26. Ramesh Raskar, Greg Welch, Matt Cutts, Adam Lake, Lev Stesin, and Henry Fuchs. 1998. The office of the future: a unified approach to image-based modeling and spatially immersive displays. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques (SIGGRAPH '98)*. ACM, New York, NY, USA, 179-188. DOI: <http://dx.doi.org/10.1145/280814.280861>
27. Jun Rekimoto. 1997. Pick-and-drop: a direct manipulation technique for multiple computer environments. In *Proceedings of the 10th annual ACM symposium on User interface software and technology (UIST '97)*. ACM, New York, NY, USA, 31-39. DOI=<http://dx.doi.org/10.1145/263407.263505>
28. Jun Rekimoto and Masanori Saitoh. 1999. Augmented surfaces: a spatially continuous work space for hybrid computing environments. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI '99)*. ACM, New York, NY, USA, 378-385. DOI=<http://dx.doi.org/10.1145/302979.303113>
29. Houssein Saidi, Marcos Serrano, and Emmanuel Dubois. 2016. Investigating the effects of splitting detailed views in Overview+Detail interfaces. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '16)*. ACM, New York, NY, USA, 180-184. DOI: <http://dx.doi.org/10.1145/2935334.2935341>
30. Dominik Schmidt, Corina Sas, and Hans Gellersen. 2013. Personal clipboards for individual copy-and-paste on shared multi-user surfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 3335-3344. DOI: <http://dx.doi.org/10.1145/2470654.2466457>
31. Julian Seifert, Sebastian Boring, Christian Winkler, Florian Schaub, Fabian Schwab, Steffen Herrdum, Fabian Maier, Daniel Mayer, and Enrico Rukzio. 2014. Hover Pad: interacting with autonomous and self-actuated displays in space. In *Proceedings of the 27th annual ACM symposium on User interface software and technology (UIST '14)*. ACM, New York, NY, USA, 139-147. DOI=<http://dx.doi.org/10.1145/2642918.2647385>
32. Marcos Serrano, Barrett Ens, Xing-Dong Yang, and Pourang Irani. 2015. Gluey: Developing a Head-Worn Display Interface to Unify the Interaction Experience in Distributed Display Environments. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and*

- Services* (MobileHCI '15). ACM, New York, NY, USA, 161-171.  
DOI=<http://dx.doi.org/10.1145/2785830.2785838>
33. Nicole Sultanum, Sowmya Somanath, Ehud Sharlin, and Mario Costa Sousa. 2011. "Point it, split it, peel it, view it": techniques for interactive reservoir visualization on tabletops. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces* (ITS '11). ACM, New York, NY, USA, 192-201. DOI=<http://dx.doi.org/10.1145/2076354.2076390>
  34. Manuela Waldner, Ernst Kruijff, and Dieter Schmalstieg. 2010. Bridging gaps with pointer warping in multi-display environments. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries* (NordiCHI '10). ACM, New York, NY, USA, 813-816.  
DOI=10.1145/1868914.1869036  
<http://doi.acm.org/10.1145/1868914.1869036>
  35. Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '09). ACM, New York, NY, USA, 1083-1092.  
DOI=<http://dx.doi.org/10.1145/1518701.1518866>
  36. Xing-Dong Yang, Edward Mak, David McCallum, Pourang Irani, Xiang Cao, and Shahram Izadi. 2010. LensMouse: augmenting the mouse with an interactive touch display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '10). ACM, New York, NY, USA, 2431-2440.  
DOI=<http://dx.doi.org/10.1145/1753326.1753695>
  37. Shengdong Zhao and Ravin Balakrishnan. 2004. Simple vs. compound mark hierarchical marking menus. In *Proceedings of the 17th annual ACM symposium on User interface software and technology* (UIST '04). ACM, New York, NY, USA, 33-42.  
DOI=<http://dx.doi.org/10.1145/1029632.1029639>

# Exploring Smartphone-Based Interaction with Overview+Detail Interfaces on 3D Public Displays

Louis-Pierre Bergé, Marcos Serrano, Gary Perelman, Emmanuel Dubois  
 IRIT – University of Toulouse  
 118 Route de Narbonne, 31062 Toulouse, France  
 {Louis-Pierre.Berge, Marcos.Serrano, Gary.Perelman, Emmanuel.Dubois}@irit.fr

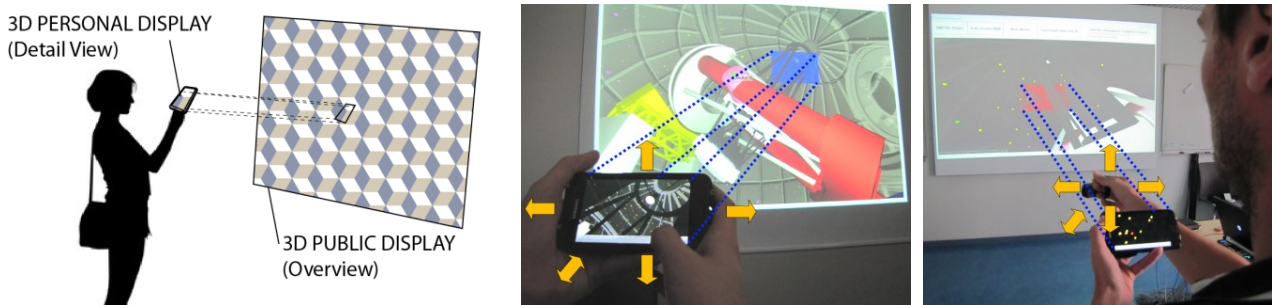


Figure 1: a) General setting of smartphone-based Overview+Detail interface on a 3D Public Display. We used two mid-air navigation techniques in a public installation to explore a 3D telescope visualization: b) Mid-Air Phone and c) Mid-Air Hand.

## ABSTRACT

As public displays integrate 3D content, Overview+Detail (O+D) interfaces on mobile devices will allow for a personal 3D exploration of the public display. In this paper we study the properties of mobile-based interaction with O+D interfaces on 3D public displays. We evaluate three types of existing interaction techniques for the 3D translation of the Detail view: touchscreen input, mid-air movement of the mobile device (Mid-Air Phone) and mid-air movement of the hand around the device (Mid-Air Hand). In a first experiment, we compare the performance and user preference of these three types of techniques with previous training. In a second experiment, we study how well the two mid-air techniques perform with no training or human help to imitate usual conditions in public context. Results reveal that Mid-Air Phone and Hand perform best with training. However, without training or human help Mid-Air Phone is more intuitive and performs better on the first trial. Interestingly, on both experiments users preferred Mid-Air Hand. We conclude with a discussion on the use of mobile devices to interact with public O+D interfaces.

## Author Keywords

Overview + detail; 3D interfaces; public display; mid-air interaction; personal displays

## ACM Classification Keywords

H.5.2. Information interfaces and presentation: Interaction.

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).  
*MobileHCI 2014*, September 23–26, 2014, Toronto, ON, Canada.  
 Copyright © 2014 ACM 978-1-4503-3004-6/14/09...\$15.00.  
<http://dx.doi.org/10.1145/2628363.2628374>

## INTRODUCTION

Public displays allow pedestrians to interact with 2D content such as city maps or tourism information [29]. In this context, existing systems have used smartphones as personal views (Detail) of the public display (Overview), leveraging multi-user access to one display [12]. Visualisation of 3D content on public displays is emerging to visualize scientific 3D data [37]; to explore culture heritage 3D scanned objects [1]; to teach history through virtual tours of 3D reconstructed historical sites [1]; to play public 3D games [42] or to navigate a city 3D map [44].

Most of these examples already include the use of a personal device to interact with the 3D content (to orient and position a slice plane [37] or to navigate in the 3D environment [44]) but few consider how to apply the Overview and Detail (O+D) paradigm using the smartphone. Using the O+D on mobile devices will provide the user with the ability to privately visualize details of the 3D environment while still taking advantage of the mobile device input to interact with the 3D content. However, the public context imposes certain constraints in terms of user's profiles (mainly beginners) and appropriate interaction techniques (which need to be easy to understand and perform).

In this paper we focus on the translation task, i.e. how to move the Detail view (displayed on the smartphone) on a 3D environment, the Overview (displayed on the public display). We report on the evaluation of three interaction techniques for controlling the translation of the Detail view: mid-air hand gestures around the mobile device like [27], mid-air movements of the mobile device itself and classical touchscreen pad. We performed two experiments. The first one is a controlled experiment with training: its goal is to

evaluate the three techniques in terms of performance, usability and user preference. The second experiment aims at studying the difficulty of performing the two mid-air gesture-based techniques in a situation imitating a public space, i.e. without the experimenter help or training.

Our contributions include: 1) An elicitation of the design space for mobile 3D O+D, 2) a controlled evaluation of three techniques for the 3D translation of the Detail view and 3) an evaluation of the difficulty to perform the mid-air techniques with no human help or training. We also report about an exhibition in public context for several days.

#### RELATED WORK

Our research is inspired by previous works on using mobile devices to visualize large datasets, to navigate 3D scenes and to interact with public displays.

##### Large information space visualization on mobile phone

To explore large information spaces such as web pages or maps, research on visualization proposes three solutions to separate focused and contextual views: Overview+Detail (spatial separation), Zooming (temporal separation) and Focus+Context (seamless focus in context) [12]. On desktop environments, previous research established that Overview+Detail (O+D) is worse than Zooming or Focus+Context (F+C) in terms of task completion time [4, 12]. However, users usually prefer O+D because it helps them build a mental model of the explored information [31]. In addition, display size has less impact on O+D than Zooming or F+C [34].

Recent implementations of these types of interfaces involved mobile devices [10] to support information search on a webpage [35], a map [9] or a scatterplot [40]. Results are mixed: O+D is faster with off-screen objects and maps but slower with scatterplots. And yet, participants prefer O+D for navigation on webpages [35]. They also are better in spatial recall of targets when using O+D [9]. Still, none of these previous works explored the use of focused or contextual views for 3D content.

Our goal is to explore interaction with a large screen providing the 3D contextual view and a mobile phone displaying the 3D focused view. F+C and Zooming interface are not in line with our setting because we have two separate views. Moreover, using F+C interface involve intentional distortion view such as fisheye or lens [12]. Using Zooming interface raises the risk of perturbing user's position and losing the user in the 3D volume. O+D interface seems to be good alternative for our context.

##### 3D navigation with a mobile phone

We focus our analysis on navigation techniques for 3D content [8] on smartphones and where the user controlled the navigation, i.e. active navigation. Existing solutions can be divided into three main types: using the touchscreen, moving the smartphone and interacting around the smartphone (around-device interaction).

Touchscreen is the standard input on mobile 3D games, usually through a rate-based pad [39]. Research works explored touchscreen input with one [5, 13, 18] or two fingers [10, 39]. To reduce finger occlusion, Hachet et al. extended the mobile phone with a 3-DOF elastic controller attached to the device [17].

Using built-in sensors or adding new ones, researchers proposed different techniques based on moving the smartphone. Hurst et al. explored two different metaphors [22]: fixed world, where the 3D environment does not move in relation with the real world, and "shoebox", where the 3D environment is fixed to the smartphone. The idea of pointing with the smartphone as if it was a magic lens has been used for augmented reality applications [1, 32] or to interact with virtual volumetric data [38]. A less direct solution based on tilting gestures has been compared with touchscreen input to navigate a virtual reality panoramic [21]. Results show that the gesture-based technique performed twice as well as the touchscreen input in an orientation task and was preferred by 80% of the users. More recent works investigated around-device interaction as a novel approach to interact with 3D content [36, 26, 27]. PalmSpace is a mid-air gesture technique to rotate 3D objects on the mobile display [27]. Around-device gestures have also been used to navigate 2D multi-scale maps [26].

In our research we compare these three different types of interactions in the context of 3D O+D public displays.

##### Public Display and mobile phone

To interact with public displays, researchers explored using multi-touch input [14] or mid-air gestures [43]. Another strategy inspired by the prevalence of mobile devices is the use of smartphones as remote controllers [28, 30]. For instance, touchscreen input is used to control a distant cursor [28], to select an item [7] or to pan and zoom [30]. Mid-air gesture with the phone has also been used for some of these tasks [6, 32].

Few works explored interaction with 3D content on public displays. Smartphones have been used to reveal hidden 3D content on augmented posters [16]. In the context of large displays (not necessarily in public context), Song et al. investigated the use of multi-touch interaction and mid-air gestures to explore and annotate 3D data [37].

An important aspect to consider on public displays is how to reveal the interaction technique to the user. This aspect has been barely addressed in the state of the art. Walter et al. [43] investigated different strategies to reveal mid-air gestures to interact with a public display. Their study indicates that users intuitively discover gestures by imitating or extending other user's gestures.

Our work is inspired by solutions using the mobile phone as the means to interact with public displays. For the task of interacting with a 3D O+D interface, existing interaction approaches need to be compared and evaluated. In particular, the public context imposes constraints on how to



reveal and learn these interaction techniques. The goal of our work is to deepen existing research on using mobile phones to interact with 3D interfaces on public displays.

#### INTERACTION TECHNIQUE DESIGN: APPROACH

Among the considerations addressed in the literature, we focus on the main properties for mobile-based interaction with 3D Overview+Detail (O+D) interfaces.

#### 3D Tasks

In our work we focus on the 3D navigation task and more precisely the *travel* subcategory defined by Bowman [8]. Generally the user controls 3 degrees of freedom (DOF) to translate the point of view and 3-DOF to rotate the point of view. Many metaphors have been explored to reduce these DOFs, such as the flying vehicle control metaphor or avatar metaphor [8]. In our work, we limit user control of the Detail view to a 3-DOF translation. This task is sufficient to explore public 3D content such as museum objects and it simplifies the task in a public setting, where interaction needs to be intuitive and straightforward.

#### Types of interaction techniques

Works presented in the literature report about different approaches to interact with 3D content. Among them using a touchscreen [18], moving the device [5] or around the device [27] have been implemented on mobile phone. We explore these three different approaches to control the position of the Detail view:

- **Mid-Air Hand** (Figure 2–a): the position of smartphone serves as a spatial reference. The position of the hand in this referential is mapped to the virtual position of the Detail view. We constrain the movement of the hand to the area behind the mobile phone. A virtual button on the mobile screen (de)activates this navigation mode.
- **Mid-Air Phone** (Figure 2–b): similar to [22], translations applied to the mobile phone translate the virtual position of the Detail view. As for Mid-Air Hand, a virtual button on the mobile display (de)activates this navigation mode.
- **Touchscreen** (Figure 2–c): inspired by commercial mobile 3D games, we use two rate-based joysticks to control the virtual position of the Detail view. The left circular joystick controls the 2D translation along the X and Y axis. The right cylinder joystick controls the 1D translation along the Z axis. Both pads can be used at the same time to control the 3-DOF navigation.

#### Mapping

Two ways of mapping a gesture with a 3D translation are illustrated in the literature: direct or indirect [26, 38]. In the direct mapping, the absolute position or gesture offset (in mid-air) is directly mapped to a 3D position on the virtual environment. Both the Mid-air Phone and the Mid-air Hand techniques are based on direct mapping. In the indirect mapping, the gesture controls the velocity and direction of

the camera movement in the virtual environment. The Touchscreen technique is based on indirect mapping.

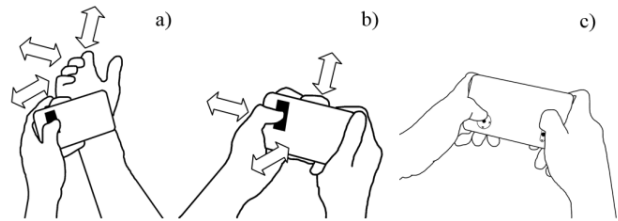


Figure 2: Three types of techniques: a) Mid-Air Hand, b) Mid-Air Phone and c) Touchscreen.

#### Mid-air gestures reference

Another factor concerning mid-air gestures is whether these gestures relate to a reference position in the real world. We can divide gestures into absolute or relative. Absolute gestures are not linked to a particular position in the real world. For instance, the Mid-Air Phone is absolute: when the user presses the button the virtual camera moves according to the absolute displacement of the phone in the real world. Relative mid-air gestures are linked to a real-world reference object. The Mid-Air Hand technique is relative as it uses the phone as a reference: the movements of the hand are linked to the position of the phone and limited in space around this reference. Previous work suggests that using a reference is useful for mid-air input without visual feedback [15] and for 3D interaction [20].

#### O+D visualization

To design the representational aspect of O+D interfaces, a set of parameters can be adjusted [10]: 1) the overlap of O+D views (usually not overlapping), 2) the relative size of both views (Overview is usually smaller than Detail view, as in PowerPoint or Google Maps), 3) the volume of space displayed by the Detail view, 4) the number of Detail views and 5) the feedback of the Detail view's position in the Overview (typically a polygonal outline or a shaded area).

In our work, we fixed these visual parameters in order to focus on input interaction. Thereby, the two views are non-overlapping and are displayed on two different displays: the Overview on the Public Display and the Detail view on the smartphone. As opposed to usual O+D, the Detail view is smaller than the Overview. In public context, this setting allows a personal view and multi-user access to the public display. The Detail view displays the 3D environment with no depth limitation. In terms of feedback, a coloured 3D pyramid on the Overview indicates the position of the Detail view: one color is assigned to each user.

#### Revealing message

Another important factor on public displays is how to inform the user of the interaction to perform. This has already been explored for full-body interaction with a large public display [43]. Different types of feedback (text, icon, video...) were explored. In our work we designed textual and image-based revealing messages to assess the suitability of studied interactions to public context.

### 3D environment: occlusion and orientation

Previous work on the use of O+D interfaces mainly refers to large 2D information spaces. In a 3D environment, interaction with the Detail view can be difficult due to occlusion and orientation issues. 3D elements can occlude the target or the feedback of the Detail view in the Overview. Without feedback, it may be hard or almost impossible for the user to situate the Detail view position in the 3D environment. To study this factor, our first experiment includes two conditions: with and without occlusion objects.

In next sections we report on two experiments to explore these properties. The first study evaluates the three types of gestures with previous training and with different occlusions. The second study evaluates the two mid-air techniques without training or indications from the experimenter.

### SYSTEM DEVELOPMENT

In this section we describe the implementation of the 3D O+D environment and the three interaction techniques illustrated in Figure 2.

#### 3D environment

Our implementation runs synced 3D content both on a mobile device and on a large display. We used a Samsung Galaxy S2 smartphone (6.6x12.5x0.8cm, 116 gr., 4.3" screen) running Android 4.1.2. The large display was a 24" monitor with a resolution of 1920x1080px. To implement all 3D content, we used a C++ open source engine based on OpenGL, Irrlicht [23]. The same C++ 3D code executed on the smartphone and on the large display: to run it on Android we compiled it with a JNI-based tool. The 3D scene on the smartphone ran at 20fps and the scene on the large display ran at 300fps. This difference of frequency was barely perceivable and no user commented on this.



Figure 3. Tracker on hand and on smartphone (left) and user performing the Mid-Air Hand technique (right).

#### Mid-air gestures tracking

To prototype the mid-air gestures we used two Polhemus Patriot Wireless trackers (7x3x2.5cm, 79.4gr.) [33]. One tracker was attached to the back of the smartphone using Velcro touch fastener. For the Mid-Air Hand technique, the tracker was attached to the hand using a glove and Velcro (Figure 3). We consider in the discussion section envisioned solutions to implement the mid-air tracking.

The trackers communicate wirelessly with a magnetic receptor, USB-connected to the computer running the 3D scene on the large display. The computer and the mobile phone communicated wireless through sockets using the IVY-bus library [24]. We filtered data using the 1€ filter [11]. The tracking system runs at a maximum frequency of 50Hz. Overall, the latency was negligible and no user commented on this.

### EXPERIMENT 1: TRANSLATION OF THE DETAIL VIEW

The goal of this experiment is to evaluate the comparative performance of the three techniques presented in the previous section for the 3D translation of the Detail view.

#### Task and mapping

We asked users to reach a target on a 3D scene using the Detail view. A sphere represented the target. Virtual walls limited the 3D scene to a cubic area. We divided the cube into a matrix of 2x9 equal smaller cubes. We placed the target in the center of one of the 18 small cubes (Figure 4-left). The Detail view, displayed on the mobile screen, is initially situated on the front-center position (red cube in Figure 4-left). The Detail view does not show the entire 3D scene and so the user needs to look at the large display to find the target.

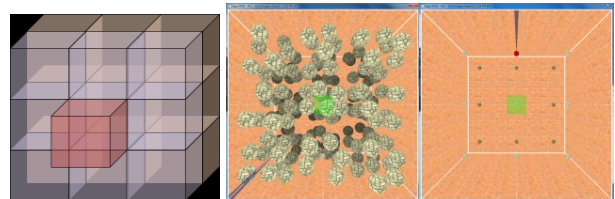


Figure 4: The 3D environment was divided into 18 areas (left), and displayed with or without occlusion objects (center/right).

The participant moves the Detail view until it is close enough to the target: on the Detail view, the target color changes to indicate the user reached it. We studied two occlusion conditions: with occlusion objects (10 spheres randomly distributed inside each of the 18 smaller areas) and without occlusion objects (Figure 4-center, right). We added a feedback on the target position: a ray attached to the target, parallel to the depth axis of the square area.

#### Apparatus

We used the apparatus described in the previous section.

#### Participants

We recruited 12 participants (3 female) aged 28.6 years on average (SD=5.2). Two of them had previously played 3D games on mobile phones and 10 of them had used mid-air interaction (Wiimote or Kinect).

#### Design and procedure

The experiment followed a 3x2 within-participant design, with *Technique* (Mid-Air Hand, Mid-Air Phone or Touchscreen pad) and *Occlusion* (True or False) as factors. Three blocks were run for each technique, the Techniques factor being counterbalanced by means of a 3X3 Latin Square. Trials in a block were grouped by the Occlusion

factor, which was always ordered by increasing difficulty (first without Occlusion, then with Occlusion). Each block of trials required 18 selections per Occlusion factor. The 18 targets were randomly ordered. Each subject performed 3 techniques x 3 blocks x 2 occlusions x 18 targets = 324 trials. On average, the experiment lasted 97 minutes, with 1 min. 21 sec. of training for each *Technique*. The training period was performed before each technique and consisted of a minimum of 5 successful trials.

#### Collected data

We logged all tracking data as well as touch events from the touchscreen. We measured trial completion time from stimulus onset to target reached. Beside time, we collected user preference through a set of techniques: usability via the System Usability Scale questionnaire (SUS) [3] and attractiveness via the AttrakDiff questionnaire [19]. AttrakDiff [19] informs on the attractiveness of a technique according to three distinct dimensions: the pragmatic quality (PQ) indicates whether the user can achieve his goals; the hedonic quality (HQ) indicates to what extent the technique enhances the possibilities of the user; finally the attractiveness (ATT) gives an idea on how the user values each technique based on its quality and engagement.

#### EXPERIMENT 1: RESULTS

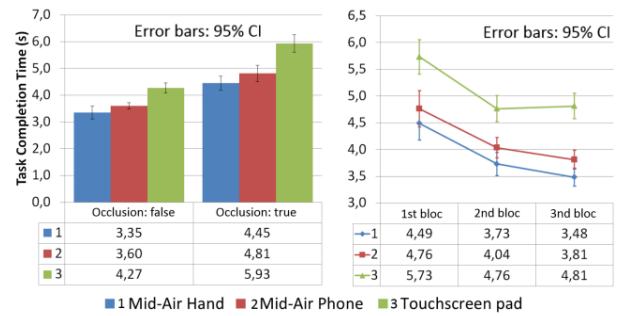
We collected 324 trials/user x 12 users = 3888 trials in total.

#### Quantitative results

##### Task completion times

A Shapiro-Wilk test shows that task completion time data does not follow a normal distribution ( $p < 2.2 \times 10^{-16}$ ). We did not find any data transformation that would allow us to use parametric tests. Our statistical analysis is thus based on non-parametric tests. Figure 5 (left) summarizes the task completion times for each technique with and without occlusion. Three different Wilcoxon tests reveal that the occlusion factor has a significant effect on task completion time for each interaction technique: +32.8% (+1.1s,  $p < 0.01$ ) for Mid-Air Hand, +33.6% (+1.21s,  $p < 0.01$ ) for Mid-Air Phone and +38.9% (+1.66s,  $p < 0.01$ ). As expected, the completion time is longer with occlusion.

For the condition without occlusion, a Friedman test reveals a significant effect of the interaction technique on task completion times ( $\chi^2(2)=7.17$ ,  $p=0.028$ ). A post-hoc test using Wilcoxon with Bonferroni correction shows a significant difference between Mid-Air Hand and Touchscreen Pad ( $p=0.032$ ). Without occlusion, only one direct mapping technique is significantly faster than the indirect mapping technique. For the condition with occlusion, a Friedman test reveals a significant effect of the interaction technique ( $\chi^2(2)=15.17$ ,  $p < 0.01$ ). A post-hoc test using Wilcoxon with Bonferroni correction shows a significant difference between Mid-Air Hand and Touchscreen Pad ( $p < 0.01$ ) and between Mid-Air Phone and Touchscreen Pad ( $p=0.016$ ).



**Figure 5: Task completion times (in s) for each technique according to the Occlusion factor (left) and learning effect for each technique (right).**

#### Learning effect

Friedman tests reveal a significant effect of the block order on task completion time for each different interaction technique (Mid-Air Hand:  $\chi^2(2)=10.67$ ,  $p < 0.01$ ; Mid-Air Phone:  $\chi^2(2)=8.17$ ,  $p=0.017$ ; Touchscreen:  $\chi^2(2)=18$ ,  $p < 0.01$ ). A post-hoc test using Wilcoxon with Bonferroni correction showed a significant difference between the first and the last block for all techniques (Figure 5-right). This confirms a learning effect. Completion time improves 22.5% for Mid-Air Hand ( $p=0.003$ ), 20.0% for Mid-Air Phone ( $p=0.05$ ) and 16.0% for Touchscreen ( $p < 0.0001$ ). Further studies are required to establish if a longer use of these techniques would increase the observed improvement as suggested in Figure 5 (right), especially for the Mid-Air Hand and Mid-Air Phone techniques.

#### Qualitative results

Three aspects have been considered in the qualitative evaluation: usability, attractiveness and user preference.

##### Usability evaluation

A SUS score was computed for each technique [3]: 81.04 (SD=12.94) for the Mid-Air Hand, 77.50 (SD=18.59) for the Mid-Air Phone and 72.5 (SD=22.69) for the Touchscreen. A Friedman test did not reveal any significant effect of the interaction techniques on the SUS score ( $\chi^2(2)=0.13$ ,  $p=0.94$ ). According to [3] the usability of the three techniques can be rate as “good”.

##### Attractiveness

To measure the attractiveness of the three techniques we relied on the Attrakdiff method [19]. We summarize in Figure 6-a, the results of the Pragmatic Quality (PQ) and Hedonic Quality (HQ) dimensions. According to the Attrakdiff report, Mid-Air Hand is rated as “desired”. More precisely, with regards to PQ, the technique is very pragmatic and assists the user optimally. With regards to HQ the report establishes the technique is very hedonic: the user identifies with the technique, which motivates and stimulates him. On the other hand, Mid-Air Phone is rated as “task-oriented”. With regards to PQ and HQ dimensions, the Attrakdiff report concludes there is room for improvement in terms of usability and user’s stimulation. Finally the Touchscreen technique is rated as “neutral”:

there is also room for improvement in terms of usability and stimulation.

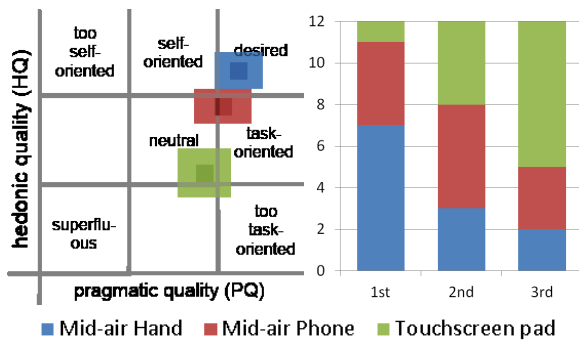


Figure 6: a) Portfolio generated using the AttrakDiff method, and b) user ranking of the three techniques.

In addition the overall user's impression of Mid-Air Hand and Mid-Air Phone is that they are very attractive (Attractiveness  $ATT > 1$ ). For Touchscreen the overall impression is moderately attractive ( $ATT = 0.5$ ).

#### User preference

We asked participants to rank the three techniques in order of preference. 7 participants rated Mid-Air Hand as the preferred technique, 4 the Mid-Air Phone and 1 the Touchscreen pad. 7 participants ranked Touchscreen pad in last position. For the statistical analysis, we marked the most preferred technique with one and the least preferred technique with three. A Friedman test did not reveal any significant effect on the mark representing the user preference on interaction technique ( $\chi^2(2)=5.17, p=0.075$ ).

Finally the most frequently mentioned positive comments refer to an accurate, funny and intuitive technique for Mid-Air Hand (P1, P3); an intuitive and easy to use technique for Mid-Air Phone (P3, P5); and a familiar technique for Touchscreen (P6, P12). The most frequently mentioned negative comments relate to the weight of the smartphone and the fatigue for Mid-Air Hand and Phone (P1); the lack of accuracy with distant targets and the loss of reference with quick movement for Mid-Air Phone (P3, P10); and the lack of accuracy, the two-handed aspect and the difficulty to combine movements in 3-DOF for Touchscreen (P6, P5).

#### Summary

Our study reveals that techniques based on direct mapping (Mid-Air Hand and Phone) are better than those based on indirect mapping (Touchscreen) for controlling the 3D translation of a Detail view. The study also reveals Mid-Air Hand scores better in terms of attractiveness and user preference, although there is no significant difference concerning SUS score. An interesting result of our study is that Touchscreen, i.e. the most common technique, is the worst in terms of performance, of perceived attractiveness and of user preference. These results are very encouraging and lead us to further explore the two mid-air techniques in a second experiment: the goal is to better investigate their use in the context of interaction with a public display.

## EXPERIMENT 2: UNCOVERING THE GESTURES

The goal of this second experiment is to evaluate the difficulty of performing the two mid-air techniques in usual public context, i.e. without training and without human explanation. We thus study the impact of revealing strategies.

#### Task and apparatus

The task is the same as in the previous experiment: reach a target in a 3D space. We decided to display the 3D space without occlusion objects.

The apparatus is the same than in previous experiment: one sensor was attached to the smartphone while the other was attached to the hand.

#### Revealing strategies

Inspired by [43], we explore different revealing strategies. In the context of smartphone-based interaction with a public display, we focus on personal feedback on the smartphone. We study two different strategies to reveal the mid-air gestures: image (Figure 2) or a text explaining the gesture to perform. The text to explain the Mid-Air Hand was: "To move the view in the 3D scene, move your right hand in mid-air behind the smartphone while pressing the on-screen button". The text to explain the Mid-Air Phone was: "To move the view in the 3D scene, move the smartphone in mid-air while pressing the on-screen button".

#### Participants

We recruited 24 participants (4 females) from the local university, aged 27.54 years on average ( $SD=5.44$ ). 8 of them had experience with 3D on mobile devices and 22 with mid-air gestures (Wiimote or Kinect). None of them participated in the previous experiment and we ensured they had not heard of it.

#### Design

To evaluate the different revealing strategies, our experiment followed a 2x2 between-participants design with Technique (Mid-Air Hand or Mid-Air Phone) and Revealing strategy (Text or Image) as factors. Every participant performed the two Techniques using one of the two Revealing strategies only. We counterbalanced the order of the Technique factor across participants: half of the participants started with one technique and half with the other. For every technique, each participant performed 8 selections on randomly selected targets (all at the same distance from the initial position).

#### Procedure

We decided to conduct a controlled experiment in order to exclude the imitation effect of public context [43], i.e. to reveal the gestures to future participants. Moreover the apparatus implied that participants could not perform the tasks with their own mobile device, excluding a large field study in a public space.

The study was performed in the presence of the interviewer. Participants were equipped with the sensors and explained the task (reach a target) without describing the interaction



techniques. We informed them of the contextual revealing message (text or image) that would describe the technique. Participants watched the revealing message and then performed the task eight times for each technique. In case the participant took more than 2 minutes to understand the technique (i.e. perform the first trial), we ended the technique's block and marked it as a failure. This time (120 s) has been identified as the maximum time a user will try to perform an interaction in public spaces [43].

To diminish the influence of the sensor on the hand (that could partially reveal the nature of the Mid-Air Hand gesture), both sensors were attached during all the experiment. We did not inform users of what these sensors were used for.

#### Collected data

We logged all tracking data and measured time to complete the task from stimulus onset (including the time to read the help message). We measured user preference and perceived difficulty using a 5-points Likert scale.

### EXPERIMENT 2 : RESULTS

Some participants failed to use some techniques since they were unable to understand the gesture. We first report on success rate and on data collected before analyzing quantitative and qualitative results.

#### Success rate and data collected

Among the 24 participants, 2 of them (8.3%) did not understand the Mid-Air Hand gesture and 1 (4.15%) the Mid-Air Phone gesture. These 3 participants were unable to understand in less than 2 minutes the technique, which was the first one to be used in the experiment. The Revealing strategy of these three failed conditions was 1 Text and 1 Image for the two Mid-Air Hand failures, and 1 Image for the Mid-Air Phone failure. Once these participants failed to perform the technique (always the first one), we asked them to perform the second technique while still providing the same revealing message: they all succeeded.

Consequently we collected a different number of trials for these three participants than for the others. For the 3 participants that failed to perform one technique we collected 8 trials (1 technique x 1 revealing strategy x 8 repetitions) x 3 users = 24 trials. For the 21 other participants, we collected 16 trials (2 techniques x 1 revealing strategy x 8 repetitions) x 21 users = 336 trials. In total we collected 360 trials.

#### Quantitative results

##### Task completion time

A Shapiro-Wilk test shows that the task completion time data do not follow a normal distribution ( $p < 2.2 \times 10^{-16}$ ). None of the data transformations we tried allowed us to use parametric tests. The statistical analysis is thus based on non-parametric tests. We first compare for each technique the task completion time between trials when the technique was performed first or second. A Mann-Whitney test indicates that the order of techniques does not have a

significant effect on task completion time ( $Z=6.62$ ,  $p=0.2755$  for Mid-Air Hand used first and  $Z=6.0$ ,  $p=0.7956$  for Mid-Air Phone used first). Thus we use all collected trials (including the 3 participants that failed one technique) for our analysis.

A Mann-Whitney test reveals the overall difference between the two techniques is not statistically significant ( $Z=9.0$ ,  $p=0.38$ ). The average task completion time for Mid-Air Hand is 24.54s ( $SD = 34.85$ ) and for Mid-Air Phone 21.50s ( $SD = 26.57$ ). We refine this analysis to distinguish the results obtained for each of the 8 trials performed by participants (Figure 7). For the first trial a Mann-Whitney test reveals a significant difference between the two techniques ( $Z=4.73$ ,  $p=0.04$ ), Mid-Air Hand being slower (95.55s  $SD = 53.71$ ) than Mid-Air Phone (63.39s  $SD = 43.99$ ). But for the other seven trials the difference is not significant (Figure 7). We note that the results obtained for these seven trials match the measures observed during the training session of experiment 1 (on average 81 second for five successfully repeated trials).

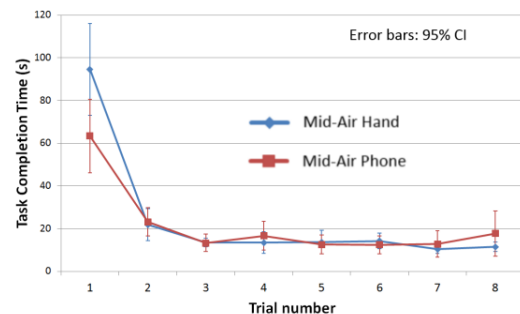


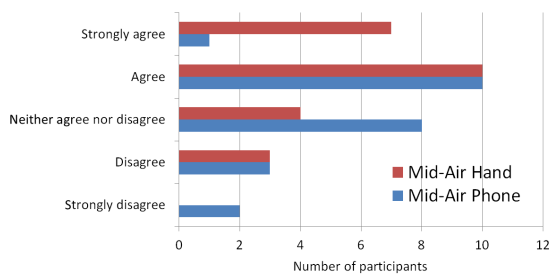
Figure 7: Task Completion Time (in s) for the 8 trials

#### Effect of the revealing strategies

A Mann-Whitney test reveals there is no significant effect of the revealing strategy factor ( $Z=4.71$ ,  $p=0.35$  for Mid-Air Hand and  $Z=5.08$ ,  $p=0.50$  for Mid-Air Phone).

#### Qualitative results

Both techniques get similar results in terms of perceived difficulty: 58% of participants find the Mid-Air Hand easy to use ("agree" or "strongly agree") vs. 63% for Mid-Air Phone. A Mann-Witney's test shows that there is no significant effect of the technique on the perceived difficulty expressed on the 5-Likert scale question ( $Z=-0.13$ ,  $p=0.90$ ). Interestingly, user preference produces different results for both techniques: 71% of participants like the Mid-Air Hand ("agree" or "strongly agree") whereas only 46% like the Mid-Air Phone technique (Figure 8). A Mann-Witney's test shows a significant effect of the technique on the overall user's rating expressed on the 5-Likert scale question ( $Z=2.23$ ,  $p=0.026$ ). This last result permits to reinforce the trend highlighted in the experiment 1 with regards to the hypothesis H3: user prefers interacting with Mid-Air Hand than with Mid-Air Phone.



**Figure 8: Likert scale results on the user preference for the two interaction techniques**

### Summary

Overall, results confirm that mid-air gestures can effectively be used to interact with O+D interface on 3D public display: 91.7% of participants have successfully used the Mid-Air Hand technique and 95.8% the Mid-Air Phone without any training or human explanation. These percentages would probably rise in a public context since users would be able to imitate other participants as observed in [43], thus enhancing the overall understanding of the gesture to perform.

A surprising outcome of our study is that the Mid-Air Hand gesture is more difficult to understand and to perform at first than the Mid-Air Phone. Not only the success rate is higher for the Mid-Air Phone, but it also allows a faster interaction during the first trial. However, our results also reveal that after the first trial, both techniques are comparable in terms of task completion time. Interestingly our study shows that despite the initial difficulty, participants preferred the Mid-Air Hand technique.

### DISCUSSION

Based on the results of our two experiments we draw a set of design guidelines for smartphone-based interaction with Overview+Detail (O+D) interfaces on 3D public displays.

#### Gesture type and spatial mapping

One of the main findings of our experiments is that mid-air gestures are more efficient and preferred than touchscreen input for 3D interaction. Previous works on mid-air interaction with 2D content showed mixed results: some found mid-air interaction to perform as well as touchscreen [26], while others found touchscreen input to perform better [30]. Our results indicate that mid-air interaction is the best solution for interacting with 3D content. This result can be explained by the straightforward mapping between the gestures and the 3D translation.

#### Mid-air gestures reference

Our study establishes that with some training relative gestures (Mid-Air Hand) are as efficient and more preferred than absolute gestures (Mid-Air Phone). This result is in line with previous work on the use of spatial references for gestures [15,20]. One drawback is that the length of user's arms limits the Mid-Air Hand interaction. To overcome this limitation, the physical space behind the smartphone could be split into two areas: hand movements in the closest area

would drive a position control interaction while movements in the furthest area would drive a rate-based control interaction, as explored in [41].

#### Gestures according to expertise

Our study also highlights that Mid-Air Hand performance depends on the user's expertise. To understand the gesture and perform it correctly, a novice user with only a text or an image as instructions is facing more difficulties than with the Mid-Air Phone (Figure 7). However, after the first trial users' performances are similar. The novelty of the Mid-Air Hand interaction, which is still far from being established, can explain this result.

#### Gesture information

A direct implication of the previous finding is that designers of 3D public applications should pay attention to the message used to reveal the interaction. Based on the experiments reported in [43] on 2D public displays and gestures, our second experiment compared the use of text and image revealing messages. As opposed to [43], we did not observe any significant differences. Given the unusual interaction, a more concrete instruction (e.g. video based or a combination of image and text) might be more appropriate. These results motivate us to keep investigating in the future the use of other revealing messages to explain how to physically operate a 3D mid-air input interaction.

#### Differences between our study and public context

Our second experiment was a controlled study representing a worst-case scenario. This experiment consisted in asking a user to perform a gesture never seen before. In a public place like a museum, the visitor is not alone and chances are s/he will observe others interacting with the system. The user will then benefit from an imitation effect [43] and the Mid-Air Hand performance will probably enhance.

To validate the interest of mid-air gestures for interacting with 3D O+D interfaces, we have employed the two mid-air techniques (Mid-Air Hand and Mid-Air Phone) in a concrete case study (Figure 1-b, c). The 3D scene, projected on a public display in the local university hall, represents a large telescope. The goal is to explore the different parts of the telescope and understand how it works. During two days, a large and varied audience (approx. 100 visitors composed of students, teachers and external public) virtually explored the dome of the telescope. This in-situ evaluation enriched the feedback from our experiments and permitted us to identify some limitations about the techniques.

#### Limitations

The addition of a sensor on the mobile phone and the hand makes them slightly heavier. As a result, several participants pointed out some muscle fatigue when using the Mid-Air Phone and Hand techniques. The size and weight of the mobile device should be limited to take fully advantage of these techniques. In the future, we will consider a better use of the embedded sensors to address

this problem and the use of vision tracking to detect the hand position on the Mid-Air Hand technique.

To perform the Mid-Air Hand technique, we used the back area of the device as in [27]. This area was established after a preliminary study aimed at selecting the optimal width, height and depth of the mid-air interaction area. During the experiments, some participants wished to perform larger gestures. It should be possible for a user to calibrate his hand movement or to adjust the interaction area. In the future, we plan to consider other areas such as the side of the device as in [26].

Finally, the in-situ installation revealed a limitation when selecting an object in the 3D scene with the Mid-Air Hand technique. Newcomers were facing problems when trying to validate the selection with the hand handling the mobile phone instead of the hand behind the phone. This sometimes resulted in very difficult thumbs motions, especially when the target was close to the border of the screen. Using the hand behind the device is neither optimal, as it induces mid-air clutching. Designing different selection procedures, such as finger gestures (pinching for instance) or device actions (such as pressing a physical button), should be considered and offered as an alternative to touchscreen input.

#### CONCLUSION AND FUTURE WORK

In this paper, we explored the design space of mobile-based interaction with Overview+Detail (O+D) interfaces on 3D public displays. We evaluated three mobile-based techniques from the literature to navigate in a 3D scene: mid-air gesture around the mobile device (Mid-Air Hand), mid-air gesture with the mobile device (Mid-Air Phone) and touchscreen input. Our two controlled experiments show that Mid-Air Hand and Mid-Air Phone perform better than touchscreen input. However, the Mid-Air Phone gesture is easier to understand than the Mid-Air Hand gesture in usual public conditions, i.e. without training and with only a text message or an image as revealing information.

In the future we plan to increase the degrees of freedom of the navigation task by integrating the 3D rotation. For instance the Mid-Air Hand technique could be extended using gestures similar to [27]. To better implement mid-air solutions, we plan to remove the additional tracker: we will explore the combination of integrated sensors (gyroscope, accelerometer) and camera-based detection. Finally, we will extend our research to consider the selection and manipulation of the 3D objects in the 3D scene.

#### ACKNOWLEDGMENTS

We would like to thank the OMP-TBL members and all the participants who generously shared their time for the study. We are also grateful to the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

#### REFERENCES

1. 3D Ancient Wonders, archeological reconstruction online virtual museum. 2014. <http://www.3dancientwonders.com/>. Accessed: 2014-05.
2. Alessio, P., & Topol, A. 2011. A Public 3D Visualization Tool for the Musée des Arts et Métiers de Paris. *Entertainment Computing – ICEC 2011*, 6972, 136–142.
3. Bangor, A., Kortum, P.T., and Miller, J.T. 2008. An Empirical Evaluation of the System Usability Scale. In *Proc. of IJHCI*, 24, 6, 574–594.
4. Baudisch, P., Good, N., Bellotti, V. and Schraedley, P. 2002. Keeping things in context: a comparative evaluation of focus plus context screens, overviews, and zooming. In *Proc. of CHI '02*, ACM, 259-266.
5. Benzina, A., Dey, A., Toennis, M. and Klinker, G. 2012. Empirical evaluation of mapping functions for navigation in virtual reality using phones with integrated sensors. In *Proc. of APCHI '12*, ACM, 149–158.
6. Boring, S. et al. 2009. Scroll, tilt or move it: using mobile phones to continuously control pointers on large public displays. In *Proc. of OZCHI '09*, ACM, 161–168.
7. Boring, S, Baur, D., Butz, A., Gustafson, S. and Baudisch, P. 2010. Touch projector: mobile interaction through video. In *Proc. of CHI '10*, ACM, 2287-2296.
8. Bowman, D., Kruijff, E., LaViola, J. and Poupyrev, I. 2004. *3D User Interfaces: Theory and Practice*. Addison Wesley Longman, Redwood City, CA, USA.
9. Burigat, S., Chittaro, L., and Vianello, A. 2012. Dynamic visualization of large numbers of off-screen objects on mobile devices. In *Proc. of MobileHCI '12*, ACM, 93-102.
10. Burigat, S. and Chittaro, L. 2011. On the effectiveness of Overview+Detail visualization on mobile devices. *Personal and Ubiquitous Computing*. 17, 371–385.
11. Casiez, G., Roussel, N., and Vogel, D. 2012. 1 € filter: a simple speed-based low-pass filter for noisy input in interactive systems. In *Proc. of CHI '12*, ACM, 2527-2530.
12. Cockburn, A., Karlson, A., and Bederson, B. 2008. A review of overview+detail, zooming, and focus+context interfaces. *ACM Computing Surveys* 41, 1, 1–31.
13. Decle, F. and Hachet, M. 2009. A Study of Direct Versus Planned 3D Camera Manipulation on Touch-Based Mobile Phones. In *Proc. of MobileHCI '09*, ACM, 32–35.
14. Fu, C, Goh, W. and Allen, J. 2010. Multi-touch techniques for exploring large-scale 3D astrophysical simulations. In *Proc. of CHI '10*, ACM, 2213–2222.
15. Gustafson, S., Bierwirth, D., and Baudisch, P. 2010. Imaginary interfaces: spatial interaction with empty hands and without visual feedback. In *Proc. of UIST '10*, ACM, 3-12.



16. Grubert, J. Grasset, R. and Reitmayr, G. 2012. Exploring the design of hybrid interfaces for augmented posters in public spaces. In Proc. of *NordiCHI '12*, ACM, 238-246.
17. Hachet, M. et al. 2008. 3D Elastic Control for Mobile Devices. *IEEE Computer Graphics and Applications*. 28, 4, 58–62.
18. Hachet, M. et al. 2009. Navidget for 3D interaction: Camera positioning and further uses. *International Journal of Human-Computer Studies*. 67, 3, 225–236.
19. Hassenzahl, M., Burmester, M., and Koller, F. AttrakDiff. 2014, <http://attrakdiff.de/index-en.html>.
20. Hinckley, K., Pausch, R., Goble, J., and Kassell, N. 1994. A survey of design issues in spatial input. In Proc. of *UIST '94*, ACM, 213-222.
21. Hürst, W. and Bilyalov, T. 2010. Dynamic versus static peephole navigation of VR panoramas on handheld devices. In Proc. of *MUM '10*, ACM, Article 25.
22. Hürst, W. and Helder, M. 2011. Mobile 3D graphics and virtual reality interaction. In Proc. of *ACE '11*, ACM, Article 28.
23. Irrlicht Engine - A free open source 3D engine. 2014. <http://irrlicht.sourceforge.net/>. Accessed: 2014-02.
24. Ivy software bus: <http://www.eei.cena.fr/products/ivy/>. Accessed: 2014-02.
25. Jacob, R. J. K., & Sibert, L. E. 1992. The perceptual structure of multidimensional input device selection. In Proc. of *CHI '92*, ACM. 211–218.
26. Jones, B., Sodhi, R., Forsyth, D., Bailey, B., and Maciocci, G. 2012. Around device interaction for multiscale navigation. In Proc. of *MobileHCI '12*, ACM, 83-92.
27. Kratz, S. et al. 2012. PalmSpace: continuous around-device gestures vs. multitouch for 3D rotation tasks on mobile devices. In Proc. of *AVI '12*, ACM, 181-188.
28. McCallum, D.C. and Irani, P. 2009. ARC-Pad: absolute+relative cursor positioning for large displays with a mobile touchscreen. In Proc. of *UIST '09*, ACM, 153-156.
29. Müller, J., Jentsch, M., Kray, C., and Krüger, A. 2008. Exploring factors that influence the combined use of mobile devices and public displays for pedestrian navigation. In Proc. of *NordiCHI '08*, ACM, 308-317.
30. Nancel, M., Wagner, J., Pietriga, E., Chapuis, O., & Mackay, W. 2011. Mid-air pan-and-zoom on wall-sized displays. In Proc. of *CHI '11*, ACM, 177-186.
31. Nekrasovski, D. et al. 2006. An evaluation of pan & zoom and rubber sheet navigation with and without an overview. In Proc. of *CHI '06*, ACM, 11-20.
32. Ouilhet, H. 2010. Google Sky Map: using your phone as an interface. In Proc. of *MobileHCI '10*, ACM, 419-422.
33. Polhemus Patriot Wireless - Polhemus. 2014. <http://www.polhemus.com/motion-tracking/all-trackers/patriot-wireless/>. Accessed: 2014-01.
34. Rønne Jakobsen, M. and Hornbæk, K. 2011. Sizing up visualizations. In Proc. of *CHI '11*, ACM, 1451-1460.
35. Roto, V. et al. 2006. Minimap: a web page visualization method for mobile phones. In Proc. of *CHI '06*, ACM, 35–44.
36. Serrano, M., Hildebrandt, D. and Irani, P. 2014. Identifying Suitable Projection Parameters and Display Configurations for Mobile True-3D Displays. In Proc. of *MobileHCI '14*, ACM, 9 pages.
37. Song, P., Goh, W.B., Fu, C.-W., Meng, Q., and Heng, P.-A. 2011. WYSIWYF: exploring and annotating volume data with a tangible handheld device. In Proc. of *CHI '11*, ACM, 1333–1342.
38. Tangible Media Group. <http://tangible.media.mit.edu/project/tether>. Accessed: 2014-01.
39. Telkenaroglu, C. and Capin, T. 2012. Dual-Finger 3D Interaction Techniques for mobile devices. *Personal and Ubiquitous Computing*. 17, 7, 1551–1572.
40. Thorsten B, et al. 2006. Usability of overview-supported zooming on small screens with regard to individual differences in spatial ability. In Proc. of *AVI '06*, ACM, 233-240.
41. Tsandilas, T., Dubois, E., and Raynal, M. 2013. Modeless Pointing with Low-Precision Wrist Movements. In Proc. of *Interact '13*, Springer, 494–511.
42. Vajk, T., Coulton, P., Bamford, W., & Edwards, R. 2008. Using a Mobile Phone as a “Wii-like” Controller for Playing Games on a Large Public Display. *International Journal of Computer Games Technology*, 2008, 1–6.
43. Walter, R., Bailly, G., and Müller, J. 2013. StrikeAPose: revealing mid-air gestures on public displays. In Proc. of *CHI '13*, ACM, 841-850.
44. Zhai, Y., Zhao, G., Alatalo, T., Heikkilä, J., Ojala, T., & Huang, X. 2013. Gesture interaction for wall-sized touchscreen display. In Proc. *UbiComp '13*, ACM, 175-178.

# Investigating the Effects of Splitting Detailed Views in Overview+Detail Interfaces

Houssein Saidi, Marcos Serrano and Emmanuel Dubois

University of Toulouse - IRIT

Toulouse, France

{first\_name.last\_name}@irit.fr

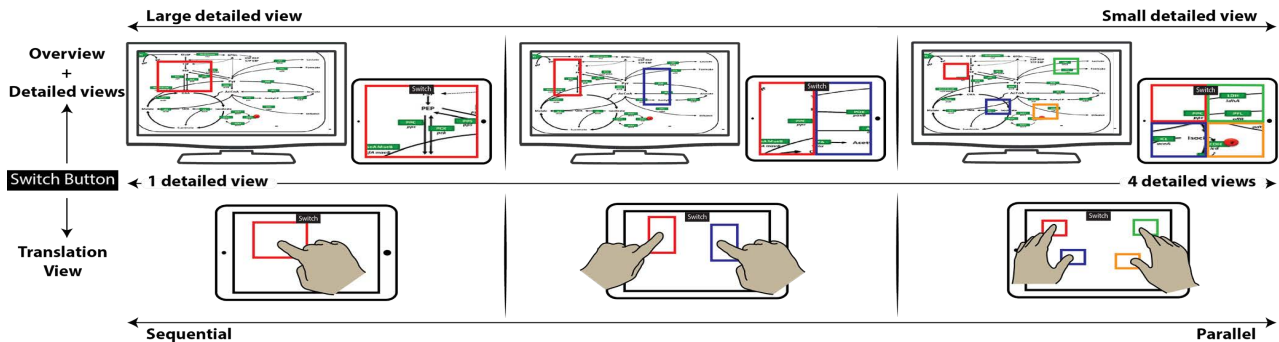


Figure 1. Arrangement of the 1, 2, 4 split-views configurations (top) and expected input control of the position of the views (bottom)

## ABSTRACT

While several techniques offer more than one detailed view in Overview+Detail (O+D) interfaces, the optimal number of detailed views has not been investigated. But the answer is not trivial: using a single detailed view offers a larger display size but only allows a sequential exploration of the overview; using several detailed views reduces the size of each view but allows a parallel exploration of the overview. In this paper we investigate the benefits of splitting the detailed view in O+D interfaces for working with very large graphs. We implemented an O+D interface where the overview is displayed on a large screen while 1, 2 or 4 split views are displayed on a tactile tablet. We experimentally evaluated the effect of the number of split views according to the number of nodes to connect. Using 4 split views is better than 1 and 2 for working on more than 2 nodes.

## Author Keywords

Interaction techniques; graph; multi-device; multi-surface; overview and detail; multi-view.

## ACM Classification Keywords

H.5.2. Information interfaces and presentation: User Interfaces – Graphical user interfaces.

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

MobileHCI '16, September 06-09, 2016, Florence, Italy  
© 2016 ACM. ISBN 978-1-4503-4408-1/16/09...\$15.00  
DOI: <http://dx.doi.org/10.1145/2935334.2935341>

## INTRODUCTION

Overview+Detail (O+D) interfaces are a well-known approach for data visualization and manipulation [9]. These interfaces reach their limits when it comes to work on multiple regions of the overview simultaneously, e.g. connecting distant nodes of very large graphs for example. Moving the detailed view repeatedly from one region to another is tedious and interaction complexity increases with the number of regions to work on [10, 11].

To address this situation, several techniques have been designed in single or multi-display configurations to support the use of more than one detailed view simultaneously [4, 5, 9, 12]. Earlier work had established a set of rules for working with multiple views [2]: the “rule of diversity” recommends the use of one view per information type and the “rule of parsimony” suggests using multiple views minimally. However none of these works has investigated the optimal number of detailed views to use, most existing techniques use 2 or 4 views. The optimal number of detailed views that will benefit complex tasks is thus still an open question.

In this paper we compare the use of different number of detailed views to interact with very large graphs in an overview + detail setting composed of a large screen and a mobile device (tablet). Our work is inspired by our collaboration with biologists using molecular interaction maps (MIMs maps [13]): these graphs can become extremely large (the Alzheimer MIM map contains 1347 nodes [13]) and complex to read or edit.

Our work aims at answering two questions: 1) are multiple detailed views better than one to interact with large graphs? and 2) what is the optimal number of detailed views needed

to perform tasks with multiple graph nodes? Answering these questions is not obvious: using a single detailed view constrains the user to translate the view sequentially to each interesting region of the graph whereas using several split view allows parallel access to different locations of the graphs.

To answer these questions, we implemented an interface based on the O+D scheme. Our interface supports the simultaneous use of up to 4 detailed views independent from each other. The overview (the overall graph) is displayed on a large screen while the detailed views are displayed on a single tablet: we hereafter refer to them as the split views. Deploying O+D interfaces on multiple displays has been shown to improve data visualization and manipulation [7,15].

We experimentally compared three values for the number of split views (1, 2 or 4) in a node connection task, where the user is asked to create a link between 2, 3 or 4 nodes. These types of multi-node links are usual in large graphs such as MIMs [13].

Our contributions are 1) a study on the effects of varying the number of detailed views in O+D interfaces and 2) a discussion on the limits of multi-view O+D interfaces.

## RELATED WORK

Several techniques have been designed to support interaction with large graphs, such as using the topology of links [14] or touch gestures to detect edge interactions [17]. However, these techniques do not allow working on different areas of the graph at the same time, as opposed to multi-display and multi-focus techniques.

### Multi-display systems

Multi-display systems [1,6,7,15] consist in combining several displays, usually tablets, large displays and tabletops, to extend the overall interaction space; it has been proven to be useful to interact with large contexts such as geographical data [1]. Multi-display systems have been used in an overview+detail configuration [4,8]. Rashid et al. [15] found that for searching on large maps, a multi-device approach was better than a simple mobile one. Cheng et al. [7] showed that, in an overview+detail multi-surface technique, moving the position of the detail in a miniaturized view was preferred over other techniques. In our work we apply this approach to multi-view interaction.

### Multi-focus techniques

The use of multiple focused views has been proposed to allow working simultaneously on multiple regions of large contexts [9,5,12]. Polyzoom [12] allows multi-scale and multi-focus exploration in 2D visual spaces by offering the user the possibility to create several hierarchies of zoomed views. Melange [9] uses a distortion-based technique that offers the possibility to bring together two regions of a large space by folding them. SpaceFold [5], inspired by Melange, introduces a multi-touch interaction technique to improve

the manipulation of the folds. These previous works inspire the multi-focus technique we implemented for our study.

## APPLICATION CONTEXT: MIM GRAPHS

This work was originally inspired by our collaboration with biologists carrying research on cancer. They archive knowledge in graphs called molecular interaction maps (MIM [13]). These MIM graphs contain several types of nodes (molecules, protein, etc.) and connections. There is no limit to the number of nodes that can be connected by one connection and each connection can also be connected to other connections, e.g. genes playing the role of catalysts of this connection. As research on cancer progresses, results are added to existing MIM maps, which grow extremely large, making them difficult to read and edit (see Figure 2). As a consequence of this growth, connected nodes can be located far apart from each other. In this context, using split-views would allow users to work on distant regions of MIM maps simultaneously.

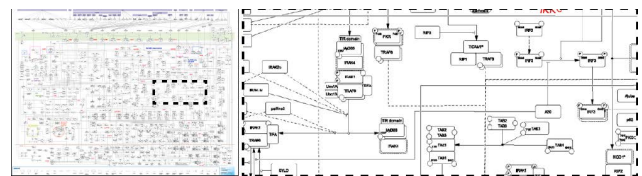


Figure 2. MIM map [13] (left) and detail (right) illustrating the density and complexity of such graphs.

## INTERFACE DESIGN

We designed and implemented an O+D visualization interface that consists of a large screen to display the contextual information and a tablet to show a magnified version of selected region(s) of the large space. We describe the three main views of our interface (overview, split views and translation view) and we analyze our design with the rules for multiple views defined by Baldonado [2].

### Split views

Our technique allows the user to have up to four independent split views at the same time (Fig.1), offering a detailed view on a graph region. It's also useful in supporting tasks requiring focusing on different places of the overview. We implemented three configurations for the multiple views on the tablet: 1-view, 2-views and 4-views. Using split views allows to *decompose* (R3) [2] the complex graph rendering.

With the 1-view technique, the split view occupies the entire tablet display; with 2-views, each view occupies half; and with 4-views a quarter. For all of them, the zoom level is always the same, which means that as the number of views augment, the information displayed by each view decreases. This design conforms to the *rule of consistency* (R7) as the overall detailed area size is consistent over the 3 versions of our technique and when several focus are displayed their relative size is consistent as well. It also presents different conditions of *space/time resource allocation* (R5): *sequential* for 1-view, and *side-by-side* for 2-views and 4-views.

Swipe gesture inside one of the split views moves the underlying graph in the same direction: this behavior is *consistent* (R7) with regular map interactions on mobile devices. Finally, when the user selects a node in one of the split views, appropriate feedback is provided so that *user's attention* (R8) is focused on the appropriate view.

### Overview

The overview displays the entire graph on a large display. The ratio between the overview size and the split views size is 9 for the 1-view configuration (overview is 9 times bigger), 18 for 2-views and 36 for 4-views. These ratios were chosen to explore the effect of a zoom factor bigger than 30 (threshold identified in [18]). A contour color is applied to the split views on the tablet and to its representation on the overview to help the user establish the *relationship between the points of view* (R6) (Fig.1).

### Translation view

Positioning the split views relies on the use of the translation view on the tablet, which is activated when the user presses the black button “switch” displayed on the tablet (see Fig. 1). The translation view provides a representation of the position of the 1, 2 or 4 split views on the overview. In the translation view, each split view position is represented using a **view icon**. Given the density of the graphs, displaying a miniature of it on the tablet would be useless. Therefore, the view icons are displayed on a black background. By looking at the overview, the user can use *multiple* (R1) view icons in *complementarity* (R2) for selecting multiple nodes.

The user can adjust the position of one or several view icons simultaneously by direct touch manipulation as recommended in [7]. Using two hands and the multi-touch screen, the user can theoretically translate 4 view icons at the same time. Closing the translation view restores the split views. In our configuration, no zoom is allowed: this ensures a higher *consistency* over the split views (R7) [1].

### USER STUDY

Using our multi-view technique, we conducted a controlled experiment to evaluate the effect of using multiple detailed views (1, 2 or 4) when connecting various number of nodes (2, 3 or 4) situated on different areas of large graphs.

### Task

Participants were asked to create a connection between 2, 3 or 4 nodes. The overview displayed only the nodes to connect on a white background. To connect several nodes, participants had to select them by touching each node in the split views displayed on the tablet. Selecting one node required translating one of the split views displayed on the tablet so that the node becomes visible. On each trial, participants could translate each of the split views with swipe gestures directly in the split view or through the manipulation of its corresponding view icon in the translation view. Selection was validated with a single tap on the node, which was then highlighted in blue. Before

each task, the position of the split-views were reset to a default position.

### Node positions

To define the position of the 2, 3 and 4 nodes to connect, we decided to fix their distance from the center of the overview and change their relative distance as well as their distribution. We used eight absolute positions corresponding to the intersection of an ellipse positioned at the center of the overview with horizontal, vertical and diagonal axes. The ellipse shape is used so that the positions of the nodes are spread across the width and height of the tablet. We selected 10 combinations of these positions for each number of nodes, equilibrating the number of neighbor nodes (i.e. on consecutive positions) and the cases where all nodes were far from each other with the cases where nodes were close to each other.

### Participants

We recruited 12 participants (4 females) from our local university. They were 26 years old on average (SD 4.7) and 11 of them were right-handed. All participants had used touchscreen tablets before. No specific skills were required.

### Apparatus

The experimental apparatus consisted of a multi-device setting involving one PC and one tablet. The PC had a 23 inches display, showing the overview (1920x1080px). Nodes on the overview measured 15x37px. The tablet was a 10.5 inches Samsung galaxy tab S (2560x1600px). Nodes on the split views (i.e. the targets to touch) on the tablet measured 40x157px. On the translation view, each view icon measured 826x526px for 1-view configuration, 413x526px for 2-views configuration and 413x263px for 4-views configuration. A Dlink DIR-615 router was used to establish a wireless connection between the workstation and the tablet. We placed the tablet on a desk and allowed users to interact with both hands, a usual configuration in multi-display settings to avoid fatigue during long interactions and to benefit from multi-touch input [16]. The tablet rested on its cover at a 60° angle and in the same field of view than the large display, which has been shown to be paramount in multi-display environments [6]. Participants sat at 1m from the display and we ensured that there were no light reflections on the tablet.

### Experimental Design

The experiment followed a 3x3 within-subject design with number of split views (NViews factor: 1V, 2V or 4V) and number of nodes to connect (NNodes factor: 2N, 3N or 4N) as factors. The NViews factor was counterbalanced by means of a 3x3 Latin square: three blocks were run, one for each value of the NViews factor. Trials in a block were grouped by the NNodes factor. Each subject performed 3 NViews x 3 NNodes x 10 predefined Node Positions x 3 repetitions = 270 trials. The training consisted of one block for each value of the NViews factor (36 trials in total). The experiment lasted 60 minutes on average.

## Procedure and instructions

To begin a trial, the participant pressed a "start" button displayed in the center of the tablet. Between each block, the user was informed via an information screen that he was about to start another condition. Participants were asked to finish each trial as quickly as possible using any number of hands or fingers. They were told they could take a break if required between trials. At the end of the experiment, participants were asked to fill a System Usability Scale questionnaire (SUS).

## Collected Data

We logged all touch events from the screen tablet. We measured trial completion time from stimulus onset to screen release, the number of actions to complete each trial and the number of switches between overview and split views on the tablet. We also logged the number of view icons translated simultaneously, i.e. the number of fingers performing a view icon translation at the same time.

## RESULTS

We used a Shapiro-Wilk test to determine the normality of collected data. Our data could not be normalized, so we used a non-parametric Friedman test to compare more than 2 conditions and Wilcoxon tests otherwise. When needed we used the Bonferroni correction.

### Completion time

Friedman tests reveal a significant effect of the NViews on completion time for each number of nodes (2N:  $\chi^2(2)=22$ , 3N:  $\chi^2(2)=22$ , 4N:  $\chi^2(2)=22$  with  $p<.01$ ). A Wilcoxon test confirms a significant difference between 1V (8652ms) and 2V (6904ms) ( $Z= -2.98$ ,  $p<.01$ ), and between 1V and 4V (6311ms) ( $Z=-3.05$ ,  $p <.01$ ). Overall, using 2V and 4V was respectively 20% and 35% faster than using 1V (Figure 3). There is no significant difference between using 2V (4487ms) and 4V (4902ms) when connecting 2 nodes, but using 4V (7015ms) was 15% faster than 2V (8112ms) when connecting more than two nodes (3 nodes:  $Z= -3.06$ ,  $p<.01$ , 4 nodes:  $Z=-3.06$ ,  $p<.01$ ).

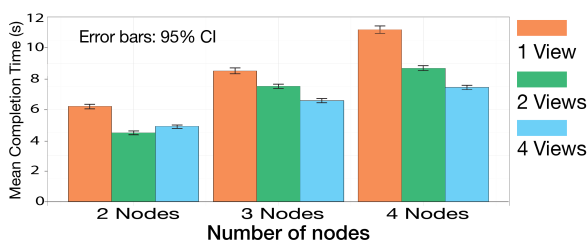


Figure 3. Trial completion time per number of nodes and number of views.

### Switches between Translation and Detailed view

A Friedman test reveals a significant effect of the NViews on the number of switches between the Translation view and the Detailed view ( $\chi^2(2)=18$ ,  $p<.01$ ). A Wilcoxon test reveals a significant difference between 1V and 2V ( $Z=-2.98$ ,  $p<.01$ ), between 1V and 4V ( $Z=-3.06$ ,  $p<.01$ ) and between 2V and 4V ( $Z=-3.06$ ,  $p<.01$ ). The number of

switches decreases with the NViews: 2.2 on average for 1V, 1.6 for 2V and 1.0 for 4V (see Figure 4-Left).

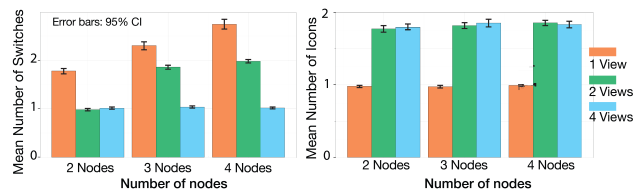


Figure 4. Left: Number of switches between the translation and detailed views. Right: Nb. of icons moved at the same time Simultaneous icons translation

A Friedman test reveals a significant effect of the NViews on the number of view icons translated simultaneously (i.e. the number of fingers moving an icon at the same time in the translation view) ( $\chi^2(2)=22$ ,  $p<.01$ ). A Wilcoxon test reveals a difference between 1V and 2V ( $Z= -2.93$ ,  $p<.01$ ), and between 1V and 4V ( $Z=-3.06$ ,  $p<.01$ ). For 1V, the number of icons used at the same time is slightly under 1 (0.99) because the user could pan inside the split view, without using the translation view (i.e. no icon translation).

Interestingly, we found no difference between the number of view icons translated simultaneously in 2V and 4V, even though users could employ their two hands to translate the view icons. In these conditions, whatever the number of nodes to connect, the average number of view icons translated was very similar (2V: 1,82; 4V: 1,83), even when more than 2 nodes had to be connected (see Fig. 4-Right).

We could expect users to move 3 or even 4 icons simultaneously by using a bimanual multi-touch gesture under the 4V condition. This actually happened, but in low proportion: over the 1080 trials done with 4V, 20% were performed moving only one view icon at the same time, 77% moving two icons at the same time, 2% (22 trials) moving three and 0.5% (6 trials) moving four icons (the rest 0.5% of trials did not involve moving any icon). The same user did 15 of these 22 trials (75%) performed with 3 fingers. Five participants did the other 7 trials: they tried the gesture one or two times but did not use it any longer. The analysis of the 6 trials done with four fingers raises similar results: one subject did it 2 times, and four users tried it once. Instead, moving simultaneously two icons seemed affordable for most participants. We observed that most of these bi-touch gestures were done with one finger of each hand in a bimanual coordinated gesture.

### SUS Scores and User preference

SUS scores reveal that the 1V and 4V conditions were deemed good (75 and 80 respectively) while the 2V was deemed excellent (86). Interestingly, when asked, users preferred the 4V condition for the tasks where they had to work on more than two nodes while opinions were mixed for the task with two nodes only: some participants liked having four views at hand, other disliked having smaller views than under the 2V condition.

## DISCUSSION AND PERSPECTIVES

In this paper, we studied the effects of splitting the detailed view in an overview+detail interface to work on large graphs. Split views were displayed on a single tablet and we evaluated three multi-view configurations: one detailed view (1V), two split views (2V) and four split views (4V). Overall, results show that using two or more split views is significantly faster than using only one detailed view. Results reveal that using 4 split views is only better than 2 split views for working on more than 2 regions of the graph.

An interesting finding of our experiment is that, when using 4 split views, users did not take full benefit of bimanual multitouch interaction to translate several view icons at the same time. Most of them (77%) used a sequential approach, first using one finger of each hand to move two icons, and then moving the two remaining view icons.

While previous work on symmetric bimanual interaction (where each hand is assigned an identical role) has already highlighted its benefit in some settings [3,19], we are only aware of one work [20] exploring symmetric bimanual *multitouch* interaction (each finger performs a pointing gesture on a different target). In this previous work, up to 47% of the trials for some tasks were performed using multiple fingers in a bimanual setting. In contrast, our results indicate that symmetric bimanual multi-touch input is hard to perform. We believe these results are dependent on the task and we need to further explore the factors influencing symmetric bimanual multi-touch interaction.

Given our findings, we plan to investigate two design questions. First, we plan to explore how to improve bimanual multitouch interaction to facilitate the translation of several split views at the same time. One idea could be to study combinations of fingers that can be moved synchronously and to help the user in employing these fingers. Second, as most participants used only one finger of each hand, we will consider other potential uses of the remaining fingers: for example additional fingers might act as modifiers to bring split views together, or to move views to specific positions such as corners.

## ACKNOWLEDGEMENTS

This work is funded by the neOCampus project of the University Paul Sabatier and part of the AP2 project (research grant AP2 ANR-15-CE23-0001).

## REFERENCES

1. Abad, Z., Anslow, C., and Maurer, F. 2014. Multi Surface Interactions with Geospatial Data: A Systematic Review. In ITS '14. ACM, pp. 69-78.
2. Baldonado, M., Woodruff, A., Kuchinsky, A. 2000. Guidelines for using multiple views in information visualization, In AVI'00. ACM, pp.110-119.
3. Balakrishnan, R. and Hinckley, K. 2000. Symmetric bimanual interaction. In CHI '00. ACM, 33-40.
4. Baudisch, P., Good, N., Stewart, P. 2001. Focus plus context screens: combining display technology with visualization techniques. In UIST'01. ACM, pp.31-40.
5. Butscher, S., Hornbæk, K., Reiterer, H.. 2014. SpaceFold and PhysicLenses: simultaneous multifocus navigation on touch surfaces. In AVI'14. ACM, pp. 209-216.
6. Cauchard, J., et al. 2011. Visual separation in mobile multi-display environments. In UIST'11, ACM, pp. 451-460.
7. Cheng, K., Dingyun Zhu, Tablet interaction techniques for viewport navigation on large displays. In CHIEA'14. ACM, pp. 2029-2034
8. Cockburn, A., Karlson, A., and Bederson, B.B. 2008. A Review of Overview+Detail, Zooming, and Focus+Context Interfaces. ACM CSUR, 41, 1.
9. Elmqvist, N., Nathalie Henry , Yann Riche , Jean-Daniel Fekete. 2008. Melange: space folding for multi-focus interaction. In CHI'08. ACM, pp. 1333-1342
10. Furnas, G. W. 1986. Generalized fisheye views, In CHI'86. ACM, pp.16-23.
11. Furnas, George W. 2006. A fisheye follow-up: further reflections on focus + context. In CHI'06. ACM, pp. 999-08.
12. Javed, W., Ghani, S., Elmqvist, N. 2012. Polyzoom: multiscale and multifocus exploration in 2d visual spaces. In Proc. of CHI'12. ACM, pp. 287-296
13. Kohn KW. Molecular interaction map of the mammalian cell cycle control and DNA repair systems. Mol Biol Cell. 1999;10(8):2703-2734.
14. Moscovich, T., Chevalier, F., Henry, N., Pietriga, E., and Fekete, J-D. 2009. Topology-aware navigation in large networks. In CHI '09. ACM, 2319-2328.
15. Rashid, U., Miguel A. Nacenta, and Aaron Quigley. 2012. The cost of display switching: a comparison of mobile, large display and hybrid UI configurations. In Proc. of AVI '12. ACM, 99-106.
16. Rodrigues, F., Seyed, T., Maurer, F., and Carpendale, S. 2014. Bancada: Using Mobile Zoomable Lenses for Geospatial Exploration. In ITS '14. ACM, 409-414.
17. Schmidt, S., Nacenta, M., Dachsel, R., and Carpendale, S. 2010. A set of multi-touch graph interaction techniques. In ITS '10. ACM, 113-116.
18. Shneiderman, B. 1997. Designing the User Interface: Strategies for Effective Human-Computer Interaction (3rd ed.). Addison-Wesley, Boston, MA, USA.
19. Moscovich, T., Hughes, J.F. 2008. Indirect mappings of multi-touch input using one and two hands, In CHI'08. ACM, pp. 1275-1284.
20. Geyer, F., Höchtl, A., and Reiterer, H. 2012. Harnessing the benefits of bimanual and multi-finger input for supporting grouping tasks on interactive tabletops. In NordiCHI'12. pp. 496-499.



# Investigating Text Legibility on Non-Rectangular Displays

Marcos Serrano\*

University of Toulouse - IRIT  
Toulouse, France  
[Marcos.Serrano@irit.fr](mailto:Marcos.Serrano@irit.fr)

Anne Roudaut\*

University of Bristol  
Bristol, United Kingdom  
[Anne.Roudaut@bristol.ac.uk](mailto:Anne.Roudaut@bristol.ac.uk)

Pourang Irani

University of Manitoba  
Winnipeg, Canada  
[Irani@cs.umanitoba.ca](mailto:Irani@cs.umanitoba.ca)

\*both authors contributed equally to this work

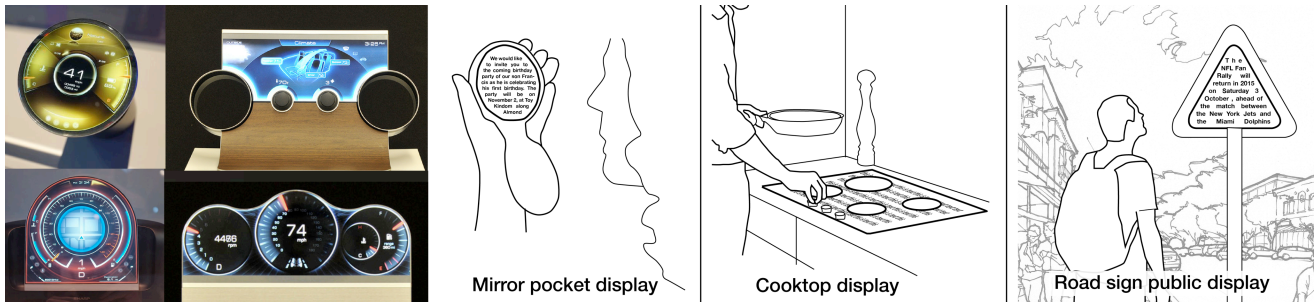


Figure 1. Left: examples of freeform displays developed by Sharp. Right: freeform display usage scenarios collected during two focus groups that illustrate the diversity of shapes that can hold text, such as: circular mirrors for private notifications, shapes with holes such as a cooktop displays for recipes or the back of triangular road signs as public displays.

## ABSTRACT

Emerging technologies allow for the creation of non-rectangular displays with unlimited constraints in shape. However, the introduction of such displays radically deviates from the prevailing tradition of placing content on rectangular screens and raises fundamental design questions. Among these is the foremost question of how to legibly present text. We address this fundamental concern through a multi-part exploration that includes: (1) a focus-group study from which we collected free-form display scenarios and extracted *display shape properties*; (2) a framework that identifies different *mappings* of text onto a non-rectangular shape and formulates hypotheses concerning legibility for different *display shape properties*; and (3) a series of quantitative text legibility studies to assess our hypotheses. Our results agree with and extend upon other findings in the existing literature on text legibility, but they also uncover unique instances in which different rules need to be applied for non-rectangular displays. These results also provide guidelines for the design of visual interfaces.

## Author Keywords

Freeform display; non-rectangular display; visual design guidelines; text legibility.

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

CHI'16, May 07-12, 2016, San Jose, CA, USA  
© 2016 ACM. ISBN 978-1-4503-3362-7/16/05...\$15.00  
DOI: <http://dx.doi.org/10.1145/2858036.2858057>

## ACM Classification Keywords

H5.2 [Information interfaces and presentation]: User Interfaces - Screen design.

## INTRODUCTION

It is commonly accepted that interactive devices should have rectangular screens and, by proxy, rectilinear interfaces. Surprisingly, the first CRT displays were designed as circles but were later changed to fit the rectangular shape of perforated strips used to record old movies. Almost a century later, we are still using rectangular displays!

Recent breakthroughs in semiconductors and display technologies (e.g. OLEDs, IGZO<sup>1</sup>) enable the design of displays with varying shapes and topologies (Figure 1-left). However, such novel form factors challenge many of the fundamental principles and guidelines that have been accumulated over the past decades for presenting and interacting with content. To support the practical adoption of such form factors, we need to rethink our understanding of how we display and interact with associated content.

Here we tackle the core concern of how to display text on a screen that is non-rectangular. Reading text is fundamental to many tasks including visually scanning, flicking through a document for specific content or displaying icons. However, running a large study comparing text legibility on multiple shapes is difficult because of the high dimensionality of possible topologies. To address this issue, we selected a self-contained subset of shapes by means of a qualitative study with end-users to seek descriptions of some compelling and practical usage scenarios of free-form

<sup>1</sup> Sharp manufactures free form LCDs using Indium Gallium Zinc Oxide thin-film transistor in the backplane of flat-panel displays.

displays in our everyday life. More specifically, we built on information obtained from participants in focus groups in two countries to extend the diversity of scenarios usages of free-form displays. From these, we compute several *display shape properties* using an algorithm inspired from [37].

We first propose a framework that identifies the possible *mappings* between text content and a shape. Supported by earlier work on text legibility, we then formulate 10 hypotheses to predict how these *mappings* impact text legibility on displays with various *shape properties*. Finally, we verify our hypotheses through four controlled experiments. The results indicate that text legibility performance can be affected by shape properties, such as with shapes having holes, or in certain interactive conditions, such as when text is scrolling in a non-rectangular display. We found that presentation strategies can mitigate these problems, such as by arranging text into columns around a display's holes or by using variable text size for scrolling on a non-rectangular display.

Our contributions are: (1) a set of compelling *display shape properties* derived from scenarios created by focus group participants; (2) a framework identifying the possible *mappings* of text onto free-form shapes; (3) a series of guidelines for the design of text content on free-form displays based on the results of our quantitative studies.

We believe that our work will not only be relevant for the growing field of organic user interfaces but also to the topic of information visualization and to the question of how to display text or labels on non-rectangular shapes.

#### RELATED WORK

Our work is motivated by recent developments in non-rectangular display technologies as well a text legibility. However the literature on text legibility is quite extensive and thus we only cover relevant work in the section "Mapping text onto shapes" to introduce our hypothesis.

#### Electronic systems

For practical reasons (wiring components, mechanical stability, and production yields) traditional liquid crystal displays (LCD) are manufactured as rectangular objects. Only recently, Sharp [19] introduced technologies to design arbitrary 2D display shapes. Embedding thin-film transistors (Indium Gallium Zinc Oxide) in the backplane of flat panel displays allows them to be 'cut' into any desired 2D shape. Deformable displays, such as Organic Light Emitting Displays (OLED) and electrophoretic displays (E-ink), are also promising technologies for free-form shape (or non-rectangular) displays. Although most technologies use a rectangular base, it is possible to create other 2D topologies [31]. For instance LG has created the first circular Plastic-OLED [43]. To our knowledge such technologies, while still in their infancy, are limited to 2D shapes.

#### Optical systems

Projections can be used to create non-rectangular displays. In Sphere [3], a projector at the base of a sphere projects pre-distorted or flattened objects on the sphere's inner surface. In [11], the same principle of back projection is applied to a humanoid face. Projection mappings can also take place on arbitrary surfaces [22] or surfaces with pre-computed geometries [34].

While these approaches can display content on many types of shapes, they are cumbersome due to the size of the projectors as well as the need for an unobstructed path between the projector and the surface. PAPILLON [6] addresses these issues with 3D printed optics. The display surface is constructed using a 3D light array of pipes that directs images from a source (e.g. a LCD) to the surface. As a result a smaller distance is required between the display sources and the surface.

#### Multifaceted systems

These systems use display primitives such as Facet [28], TUISTER [8], DataTiles [35] or Tilt Displays [1] to compose larger displays. In some of these systems, actuators facilitate the dynamism in display shape properties [1]. Multi-faceted systems have also found practical applications in larger scale public displays [25], wherein several rectangular displays are combined in different shapes to investigate the effect on passer-by interaction. Another interesting example of a multi-faceted system is D20 [33], which is a prototype of an icosahedron-shaped handheld digital device that has a triangular display on each of its faces.

#### FOCUS GROUPS: DISPLAY SHAPE PROPERTIES

The goal of this initial study is to collect usage scenarios of free-form displays in order to generate *display shape properties* that will inform our choices of shape categories in further quantitative studies.

#### Goal

As previously highlighted, the high dimensionality of geometrical topologies makes it difficult to design an experiment to compare all the potential variables that define a free-form display. We thus aimed to identify only a subset of relevant shapes. To achieve this, a first strategy may be to systematically study the geometrical features of shapes, e.g. using *shape resolution* features [36]. However this approach might lack ecological validity. Another approach could be to focus on a particular case. However results would have been too specific and not generalizable. We opted for a more general and user-centered approach and thus brainstormed with end-users to capture a subset of compelling shapes in terms of displaying and interacting with content. This approach constrains the potentially large set of shapes, but it also provides the added advantage of allowing us to focus our set of initial studies on this subset.

**Task and procedure**

We ran two focus groups in 2 countries (France and the UK) to maximize the diversity of scenarios we could collect, and to avoid cultural biases (albeit, both countries are dominated by “Western” culture). Eight university students (2 females) with an average age of 27 took part in the French study, while the UK study was made up of 12 computer science students (3 females) between the ages of 20 and 24.

We began the focus groups by presenting the concept of free-form display. We then asked participants to brainstorm in groups of 3-4 and to create cards describing their ideas. Each card contained specific text fields: what, where, and a description along with a blank area for an illustration of their idea with a shape. We ended the session with an open discussion where participants could generate more cards.

We collected 62 cards depicting 41 shapes once redundancies were eliminated. Most were 2D, and 3D ones were represented using 2D likenesses that corresponded to the user’s point of view (e.g. a circle for a sphere). The 3D shapes consisted of simple geometries such as spheres, cylinders, or semi-cylinders (e.g. the arm rest of a seat).

**Results**

To analyze the shapes, we used a clustering algorithm similar to the one proposed in [37]. Figure 2 shows the 41 shapes clustered in 10 groups. From this figure we can observe a set of *display shape properties*:

- **Symmetry:** overall there are slightly more symmetrical shapes (25) than non-symmetrical.
- **Curvature:** group A contains the largest number of shapes that are ovoid in nature, for example car side mirrors, purses, sinks, or oval tables. Group B, C and E are also quite round in nature (ovals, circles, cylinders) but these groups also contain some shapes with sharp boundaries such as a triangle (road signs), miniature house shape, a tee shirt, or a cooktop. Our algorithm grouped these shapes because their overall distribution of points is similar, although boundaries differ. On the other hand, group F contains shapes with rather smooth curvature but with more intricate patterns such as a hand, a piece of a jigsaw puzzle, or a humanoid shape.
- **Porosity:** we found several shapes with holes in groups A (bathroom elements), E (electrical plugs), G (glasses), and H (cooktop). In group F we also observed shapes with long concavities such as a hand or a jigsaw puzzle piece. We can liken these concavities to open holes as they will disturb the display of the text in a similar way.
- **Length-Width ratio:** groups D, I, and J contain shapes with a low length to width ratio, meaning that a good number of samples included long and thin shapes such as pencils, faucets, chair arms, or belts.

In addition to the observations in Figure 2, we also noted these additional properties:

- **Orientation:** we observed that some displays had particular orientations (not reflected in Fig.2 as our

algorithm is rotation invariant). This is quite often the case with shapes that have a small Length-Width ratio. For instance the pen where the display on the casing is at an angle from the user’s point of view. Another example is the handle of a frying pan.

- **Scrolling:** many scenarios involve a display content that needs to be scrolled to access more information, such as the case of the cooking jar or the umbrella.
- **Environmental conditions:** finally, we also observed that some scenarios involve specific environmental conditions. This is most typical in the case of the cooktop where the display is close to heating elements, but it is also applicable in the case of bathroom elements that are in contact with water. For those cases, other design considerations must be taken that we have not explored in this paper.

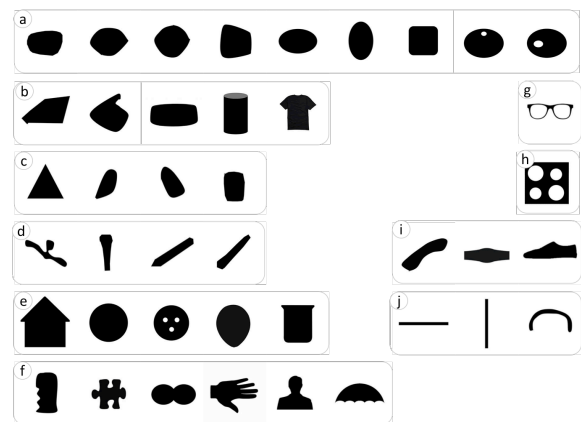


Figure 2. Cluster of shapes from the qualitative study.

**MAPPING TEXT ONTO SHAPES**

We propose a framework that aims at presenting different *mappings* of text content onto arbitrary shapes. We then draw on relevant text legibility work and formulate hypotheses to predict how the *mappings* affect legibility when displayed with different *display shape properties*.

**Mappings framework**

The framework describes three axes with increasing levels of abstraction (examples in Fig. 3). This list is non exhaustive as we only consider text mappings that relate to readability, e.g. we dropped cases with upside down text.

- **Layout:** this axis describes the general text layout, which can be *continuous* or by *block*. For example, the CHI Proceedings layout is in blocks (formatted on two columns). We could have also considered the case where the layout is not continuous (e.g. random), but this would clearly disturb text readability.
- **Token size:** this axis describes the size of the tokens, which can be *constant* or *variable*. E.g. the fisheye menu [2] illustrates the case variable. It is important to note that many deformations are possible.
- **Line alignment:** this axis describes the line alignments in which the text fits. It could be *linear*, i.e. horizontal, or oriented parallel lines, or what we call *tangential*, i.e.

following the shape. More precisely, text could follow a vector field around the shape boundary. This is typically the case in calligrams<sup>2</sup>. Note that various vector fields can be generated resulting in different text alignments as shown in [29].

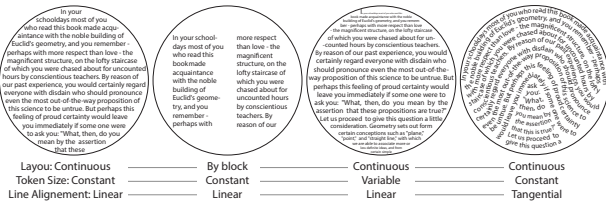


Figure 3. Example of text mappings.

**Hypotheses**

To formulate hypotheses around our central question of how non-rectangular displays affect text legibility, we rely on the existing knowledge on text legibility. For each hypothesis we cite relevant related work (these references are not exhaustive given the huge body of knowledge on text legibility). In the following, the *return sweep* refers to the action of visually scanning a line and returning to the start of the next one.

**Layout**

Because we are familiar with reading text that is aligned to the left [16,38,44], we can assume that return sweeps will be more difficult when the text is not left aligned:

- H1. Text will be more legible on shapes allowing a straight left alignment than it will with non-straight ones.

However, certain shapes can make the return sweep more predictable. This can be the case in shapes with low curvature where the left-alignment does not change abruptly. It can also be the case with symmetrical shapes: we know from prior work that the type of right-alignment does not affect the readability (e.g. ragged vs. justified). Nevertheless, if the right-alignment and left-alignment are symmetrical, it may be easier to perform the return sweep:

- H2. Non left-aligned text will be more legible on shapes with low curvature than it will on shapes with high curvature.
- H3. Non left-aligned text will be more legible on symmetrical shapes than it will on non-symmetrical ones.

We also know that line length is correlated with higher readability [13,14,39,40] and thus we can assume that:

- H4. Continuous layout will be more legible than broken layouts.

However, this might not be true when the topologies disrupt the flow of a continuous layout [15], thus decreasing the line length. This is typical of shapes with holes or long concavities. In such cases we can assume that:

- H5. For shapes with holes or long concavities, Broken Layout text will be more legible than Continuous Layout.

**Token Size**

Larger tokens produce larger retinal images [4,40] and we can thus assume that a constant token size will outperform a variable token size.

<sup>2</sup> A calligram is a phrase arranged to create a visual image.

- H6. Text will be more legible if token size is constant as opposed to variable.

On the other hand, changing the token size can help create text that is spatially stable when scrolling, and we can assume that variable token size will improve legibility [9,14,18,30].

- H7. Text will be more legible if token size is variable as opposed to constant when the text is undergoing scrolling line by line.

**Line alignment**

Large line spacing between lines improves readability based on the assumption that it makes it easier for users to detect lines [5,24]. We can thus assume that it will be harder to predict the start of the next line in the case of a tangential line alignment. In addition, shapes with low changes in curvature will create less abrupt changes, making the text more legible.

- H8. Text will be more legible if the line alignment is linear rather than tangential.
- H9. For tangential line-alignment, text will be more legible if the shape change in curvature is low instead of high.

In addition, we know that text orientation decreases readability after a certain angle [17,20,26,40,42]. However, there are cases where it could be more advantageous to orient the text in order to increase the line length. This is typical in thin and long oriented shapes.

- H10. For thin and long shapes, text will be more legible if the line alignment follows the main axis of the shape rather than the horizontal.

**Summary**

We have proposed 10 hypotheses that predict how the text mappings affect legibility when displayed with different display shape properties: layout, token size and line alignment. Our hypotheses rely on existing knowledge on text legibility, but also extend it as yet we are unaware of any study investigating text legibility on non-rectangular shapes. In the following of this paper we carry a set of quantitative studies on different display shapes to validate or invalidate our predictions.

**QUANTITATIVE STUDIES OVERVIEW**

All the studies are designed to examine specific hypotheses: study 1 is to verify H1-3; study 2 to verify H4-5; study 3 to verify H6-7; and study 4 is designed to verify H8-10.

**Reading Task**

Reading tasks need to be carefully designed so they bear resemblance on how we commonly read (a trait that is left missing from many readability studies [12]). Two primary task options exist. In one case, the post-reading comprehension of users is evaluated using procedures such as the Nelson-Denning reading test [7]. However, this test is designed primarily for gauging reading deficiencies. A second approach consists of seeking spelling mistakes or finding specific words. Such tasks promote skimming.

We adopted a task similar to that of Jankowski et al. [21], used successfully in a number of studies [23,4,10]. The task introduces word substitution errors, which forces participants to read and comprehend sentences. Incomprehensible sentences need to be flagged for errors.

Such a task has been considered to tease apart many realistic reading traits [21] as subjects must read the entire passage to recognize substituted words. The new words are common words that are placed grossly out of context. We also designed the test to ensure that native and non-native speakers have no difficulties in identifying such errors.

### Text length

We focused on short text (150 to 170 words) as a result of our brainstorming sessions. Using longer texts may have shown more differences in results, but small passages are ecologically valid and in line with the scenarios we gathered. Moreover our text length was similar to that in previous studies [21], where texts were 150 words long.

### Participants

A total of 37 people (8 female) with normal or corrected-to-normal vision took part in our experiments. Eleven of them participated in two studies (on different dates) and each study involved 12 participants. Participants were aged 24.7 years on average (SD=6.4) and 29 of them were native speakers while all others were fully proficient. 19 participants were undergraduate students, 12 were PhD students, and 6 were senior researchers from a local university. Participants reported that they read from a computer screen for 8.9 hours/day on average. Table 1 shows participant details for each study.

	Participants (female)	Age (SD)	Native speakers	Reading time/day
Exp. 1	12 (4)	26.4 (4)	9	9 H
Exp. 2	12 (1)	20.5 (5)	9	9.25 H
Exp. 3	12 (1)	24.2 (8)	9	8 H
Exp. 4	12 (4)	27.6 (4)	11	9.4 H

Table 1. Participants involved in each experiment.

### Apparatus, implementation and font

We used a 21.5-inch iMac with a 1.4GHz dual-core Intel Core i5 and a 1920x1080 display. The operating system was OSX. The participants sat in a lab illuminated by overhead fluorescent lights. They were positioned 1 meter from the display to ensure no glare appeared on the screen. The graphical rendering and input interaction was developed using Processing. The application loads images (white shape on black background) and calculates the shape's area by using a standard ray-casting technique. The application then computes valid text lines from several parameters, such as interline spacing and margins. For each study trial, the application loads text from an XML file and automatically fills the shapes using the appropriate values for layout, line alignment, and token size. We used sans serif Helvetica because sans-serif fonts are easier to read on screens [21]. We carried out informal tests to define typographical values: default text size was set to 16 pixels.

### Measurements of legibility

As in prior work [4,21], we measured text legibility by both examining reading time and reading accuracy. We recorded (a) reading completion time, as well as (b) the number of errors identified in each passage. We also developed a questionnaire to collect the participants' subjective

impressions of our various text rendering styles. Our questionnaire probed user's perception of the aesthetics [27], as aesthetics can be critical to the degree of enjoyment associated with a task [41]. We selected 4 aesthetic labels: (a) chaotic/clean, (b) boring/interesting, (c) ugly/beautiful, and (d) non-aesthetic/aesthetic. We also asked participants to rate how easy or difficult it was to read the text passages.

### Procedure

Our interface was similar to that used in [21]. The reading task began as the participant clicked on the "Start" button and ended when the user clicked on the "Done" button. Users were presented with a series of passages and for each clicked the "Start" button, read the passage, clicked on any erroneous word substitutions, and then clicked "Done". They were asked to read the text passages "as accurately and as quickly as possible" and to read them only once. Clicking on a word substitution caused the application to highlight the word in green if the word had indeed been replaced or in red if the word was from the original text, giving participants feedback on the correctness of their actions as they went. The participants were instructed to keep questions and comments for the breaks between passages. To avoid boredom and eye-strain effects the users were told that they can rest during the breaks. Participants were not told how many substituted words were introduced in the passages. We controlled these to be 4 to 6 errors per passage so that users would not familiarize themselves to the exact number. After being presented with all passages, users were given the questionnaire to rate all renderings.

### Statistical analysis

We used a Shapiro-Wilk test to determine the normality of the data. If the data was normal or could be normalized, we used a Univariate ANOVA test (we report F and p). If not, we used a Friedman test (we report  $\chi^2$  and p) to compare more than 2 conditions, and Wilcoxon tests otherwise (we report p value). If needed, we used a Bonferroni correction.

### EXPERIMENT 1: LEFT AND RIGHT ALIGNMENTS

The goal of this study is to support or refute H1-3. We compared shapes with different left or right text alignments.

- H1. Text will be more legible on shapes allowing a regular left alignment than it will with a non-regular one.
- H2. Non left-aligned text will be more legible on shapes with low curvature than it will on shapes with a high curvature.
- H3. Non left-aligned text will be more legible on symmetric shapes than on non-symmetric ones.

### Shapes and text

As shown in Figure 2, we tested 16 shapes by combining 4 left alignments and 4 right alignments (symmetrical from the right). The 4 alignments were linear, circular, sharp and irregular. We chose these shape properties in relation to the findings of the quantitative study showing shapes with or without *symmetry* as well as shapes with different *curvature*. We used 24 different paragraphs from the novel *The Stranger* by Albert Camus. Passages had the same length, with 150 words on average (SD = 1.1). 22 paragraphs were repeated three times and 2 paragraphs



were only showed during training. Different words were replaced each time.

**Experimental design**

A repeated measures within-subject design was conducted. The independent variables were Left alignment (linear, circular, irregular, sharp) and Right alignment (linear, circular, irregular, sharp). The combination of both factors created 16 shapes as shown in Figure 4. The presentation of variables was counterbalanced among participants by means of a Latin Square. All of them performed 4 trials for each condition. At the beginning of the experiment, the participants performed 4 practice trials on a random shape that was not used in this experimental design. In summary, the design was: 4 Left Alignments × 4 Right Alignments × 4 trials = 64 trials (35min on avg.) per participant. Our sessions lasted approximately 50min, including a training session (with 4 passages and 4 randomly selected shapes) to familiarize the participant with our interface and procedure.

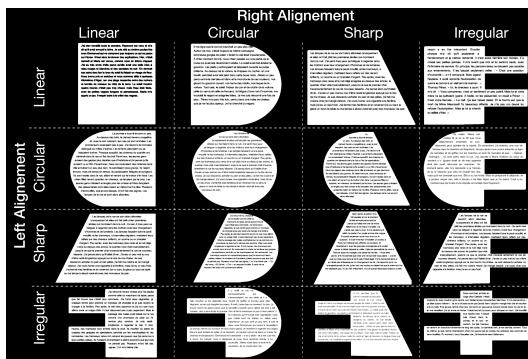


Figure 4. Shapes used in experiment 1 to test the combination of left and right alignments (text darkened for this figure).

**Results and discussion**

*Reading time and accuracy*

Quite surprisingly, we did not find that left or right alignment had a significant effect on task completion time (Figure 5). The observed power of the test was 0.79. On average, reading 150 words took 31.3 seconds (95% CI [30.5, 32.1]). We also did not observe that left or right alignment had an effect on reading accuracy. On average, reading accuracy (i.e. the percentage of words found) was 95.6%, (95% CI [94.9, 96.3]). Although non-significant results do not prove anything, we see these results as promising because displays with considerably different shapes did not impact the text legibility performance of the representative participants.

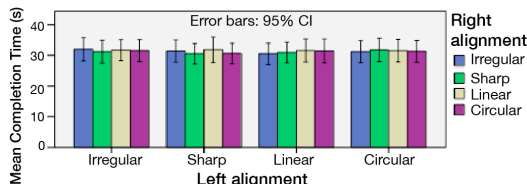


Figure 5. Reading time for each combination of left and right alignment.

*Subjective results*

Participants rated certain shapes as easier to read than others ( $\chi^2(15)=48, p<.01$ ). Each shape with Irregular left or right alignment was rated significantly different ( $p<.02$ ) from its symmetric counterpart (i.e. Irregular-Linear from Linear-Linear). Overall, shapes involving combinations of Linear, Circular and Sharp alignments were rated Easy by at least 75% of participants, compared to only 45.8% for irregular alignments (Table 2).

EASY	right straight	right circular	right sharp	right irregular
left straight	83,33%	75,00%	91,67%	58,33%
left circular	75,00%	75,00%	66,67%	50,00%
left sharp	58,33%	75,00%	91,67%	50,00%
left irregular	41,67%	41,67%	41,67%	41,67%

Table 2. Percentage of participants that rated each condition as Easy (5 or higher on 7 points Likert Scale).

Questions regarding the aesthetic properties produced similar outcomes: 71% of participants considered conditions involving Linear, Circular, and Sharp alignments to be clean, beautiful, aesthetic, and interesting, compared to only 21.4% for conditions involving Irregular alignment.

*Discussion*

We observed no significant differences in reading rates or accuracy despite users’ subjective preference for shapes without irregular boundaries. This study does not support H1, H2 or H3 in terms of performance times. However symmetric shapes are perceived to be easier to read compared to shapes with irregular alignment, which supports H3 and partially H1 in terms of user’s perception.

**EXPERIMENT 2: LAYOUT COMPARISON**

The goal of this study was to test H4 and H5. We compared different text layout on various shapes with or without a hole.

- H4. Continuous layout will be more legible than broken layout.
- H5. For shapes with holes or long concavities, Broken Layout text will be more legible than Continuous Layout.

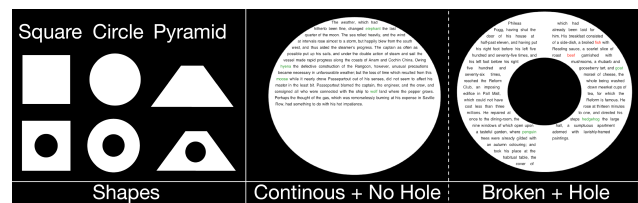


Figure 6. Shapes used in Experiment 2 and illustration of Holes and Layout factors with the Circle shape.

**Shapes and texts**

We evaluated the 12 conditions combining three shapes with or without a hole and with two text layouts (Figure 6). We used the same shapes as study 1, except we explored porosity instead of symmetry this time. We dropped the cases of shapes with extremely irregular borders because they were really disliked in study 1, and we used a round hole because it appeared frequently in our qualitative study. We used 54 different paragraphs from the novel *Around the World in Eighty Days* by Jules Verne. Passages were all 170 words. (SD = 0.3), and no paragraphs were repeated.



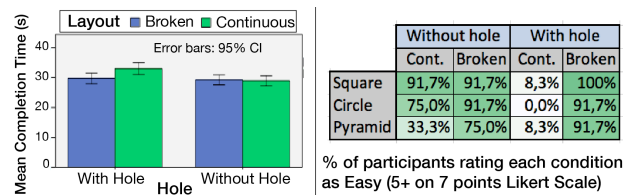
**Experimental design**

A repeated measures within-subject design was conducted. The independent variables were Shape (Square, Circle, Pyramid), Holes (With, Without) and Text Layout (Continuous, Broken). The variables were counterbalanced among participants. All of them performed 4 trials for each condition. At the beginning, participants performed 4 practice trials. In summary, the design was: 3 Shapes × 2 Holes × 2 Text Layout × 4 repetitions = 48 trials (~25 min) per participant. The experiment lasted ~40min.

**Results and discussion**

*Reading time and reading accuracy*

Figure 7 shows the results on completion time. Our tests indicate a statistically significant effect of Holes ( $F_{1,11}=6.3, p=.012$ ) and an interaction effect between Holes and Layout ( $F_{1,11}=3.9, p=.047$ ). There was no effect of the Shape factor, which is consistent with results from experiment 1. The observed power of the test was 0.871. On average, reading text on a shape with a hole was 8% slower than on a shape without a hole. A post-hoc comparison reveals no significant difference between layouts without a hole, but there is a significant difference when a hole is present ( $p=.01$ ): continuous layout is 10% slower than a broken layout. There were no effects on reading accuracy. On average, reading accuracy was 77.3% (95% CI [75.6, 79.2]).



**Figure 7. Left: Reading time for the conditions with or without hole for the two layouts. Right: Perceived reading easiness.**

*Subjective results*

We found significant differences in user ratings among conditions ( $\chi^2(11)=97, p<.01$ ). For each shape, ratings on continuous layout with hole were significantly different from all other conditions ( $p<.01$ ), except for the Pyramid shape without hole and with a continuous layout. On average, the three conditions involving a hole and a continuous layout were only rated Easy by 5% of participants. More than 75% of users rated all other conditions Easy (except the Pyramid one cited earlier).

Results regarding aesthetic properties followed the previous tendency on the Hole-Broken vs. Hole-Continuous division. Most participants (58%) found the Hole-Broken condition to be clean, interesting and aesthetic, even if half of participants found it ugly. The Hole-Continuous condition was perceived as chaotic, ugly, boring (except for the circle shape), and non-aesthetic by 70.8% of participants. Conditions without hole were mostly perceived as clean (61%) but not interesting (25%).

*Discussion*

Our study cannot confirm H4 as we found no overall effect of the Layout factor. However the study reveals that, when using a continuous layout, shapes with holes are more difficult to read than shapes without holes, which confirms H5. In this particular condition, the hole cuts sentences and readers have to locate the second part of the sentence after the hole, which is tedious. Using a broken layout neutralizes the negative effect of the hole, both in reading time and in perceived difficulty.

**EXPERIMENT 3: TOKEN SIZE AND SCROLLING**

The goal of this study is to support or refute H6 and H7. We compared different token sizes on different shapes. We also wanted to compare the impact of continuous scrolling vs. page scrolling on text legibility [32] in free-form shapes.

- H6. Text will be more legible if token size is constant rather than variable.
- H7. Text will be more legible if token size is variable instead of constant when the text is scrolling line by line.

**Shapes and texts**

We tested 12 conditions combining three shapes, two scrolling techniques, and two token sizes (Figure 8). We chose the same shapes used in study 2 (square, circle and pyramid) except that this time we did not investigate the porosity but instead included the scrolling feature as identified from our focus group results. Token resizing was done using a linear function with the line size as variable in order to have the same amount of text on each line. As a result, the size of the square shape text was always constant, even for the resized condition. When dynamically scrolling resized text, each line moved up or down (Figure 8). Unlike the case of the square, dynamically scrolling non-resized text on the circle or pyramid shapes involved words being repositioned for every scroll movement (Figure 8). For page scroll with resized text, circle and pyramid shapes were partially filled to avoid text smaller than 11px, as users could not scroll or zoom to see the smaller text. Participants scrolled using the mouse wheel. We used 40 different paragraphs taken from book synopses on Wikipedia. Passages were 200 words on average (SD = 0.5), and no paragraphs were repeated.



**Figure 8. Shapes used in exp. 4 to test the combination of Scroll type and Token Size. Blue boxes and arrows illustrate how a word would move when text is dynamically scrolled.**

**Experimental design**

A repeated measures within-subject design was conducted. The independent variables were Token Size (constant, variable), Shape (square, circle and pyramid) and scrolling type (dynamic scrolling or page scrolling). The presentation of variables was counterbalanced among participants. All of them performed 3 trials for each condition. At the

beginning of the experiment, the participants performed 8 practice trials. In summary, the design was: 2 Token Sizes × 3 Shapes × 2 Scrolling Types × 3 trials = 36 trials (~40 min) per participant. The experiment lasted ~1 hour.

**Results and discussion**

*Reading time and accuracy*

Results on completion time indicate a statistical significant effect of Shape ( $F_{2,22}=6.5, p=.031$ ), Token Size ( $F_{1,11}=12, p<.01$ ) and Scrolling ( $F_{1,11}=35, p<.01$ ), as well as an interaction between Token Size and Scrolling ( $F=14.6, p<.01$ ) and between all three factors ( $F_{2,22}=3.8, p=.02$ ). The observed power of the test was 1. A Post-hoc test reveals no effect of the scroll factor on the square shape. We found no effect of page scroll for the circle and pyramid shapes among token sizes. However, dynamic scrolling on the circle and pyramid with constant token size was respectively 17.4s and 14.3s slower than with resized text size (Figure 9). There were no effects on reading accuracy, which was on average 90.06%, (95% CI [88.67, 91.45]).

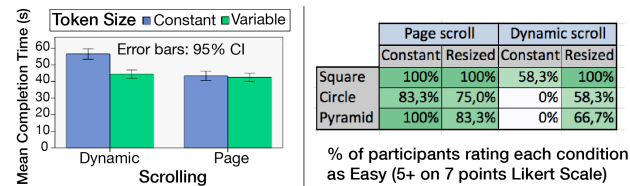


Figure 9. Left: Reading time for the scrolling techniques with constant or variable token size. Right: Reading easiness.

*Subjective results*

We found significant differences in user ratings among conditions ( $\chi^2(7)=69, p<.01$ ). There was no difference among ratings between constant and resized text for page scroll. For Circle and Pyramid, constant token size under continuous scroll was significantly different from all other conditions ( $p<.01$ ): it was never rated Easy. At least 58% of participants rated all other conditions Easy (Figure 9). Concerning aesthetic properties, constant token size with dynamic scroll were considered to be ugly, chaotic, non-aesthetic, and boring by 73% of participants, while constant token size with page scroll was considered as beautiful, aesthetic, and clean by at least 62.5% of participants. The variable token size condition with dynamic scroll evoked the same outcome, and some participants pointed out that the Pyramid reminded them of the ‘Star Wars’ crawl. Results on variable size with page scroll were mixed.

*Discussion*

Our results interestingly reveal that shapes with constant text size are not any easier to read than those with variable sized text. Furthermore, when scrolling, variable text size makes text easier to read on non-rectangular displays. Our findings support H7 but not H6.

**EXPERIMENT 4: LINE ALIGNEMENT**

The goal of this study was to examine H8-H10. We compared different line alignments on different shapes.

- H8. Text will be more legible if the line alignment is linear rather than tangential.
- H9. For tangential line-alignment, text will be more legible if the shape change in curvature is low rather than high.
- H10. For thin and long shapes, text will be more legible if the line alignment follows the main axis of the shape instead of the horizontal.

**Shapes and texts**

We tested 8 conditions, shown in Figure 10, combining four shapes with two line alignments. We chose different shapes than those we used in the three previous studies, as we wanted to investigate different shape properties highlighted in our focus groups. In particular, we wanted to explore extreme Curvature, Length-Width Ratio (thin shape), and Orientation. We used 40 different paragraphs from book synopses on Wikipedia. Passages were 150 words on average (SD = 0.4), and no paragraphs were repeated.

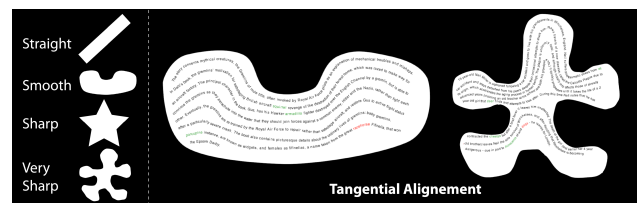


Figure 10. Left: Shapes curvatures used in exp. 3. Right: Illustration of Tangential alignment with two shapes.

**Experimental design**

A repeated measures within-subject design was conducted. The independent variables were Line Alignment (continuous or tangential) and Shape Curvature (straight, smooth, sharp, very sharp). The presentation of factors was counterbalanced among participants. All of them performed 4 trials for each condition. At the start of the experiment, participants performed 8 practice trials. In summary, the design was: 2 Line Alignment × 4 Shape Curvature × 4 repetitions = 32 trials (~40 min) per participant. The experiment lasted ~50 min.

**Results and discussion**

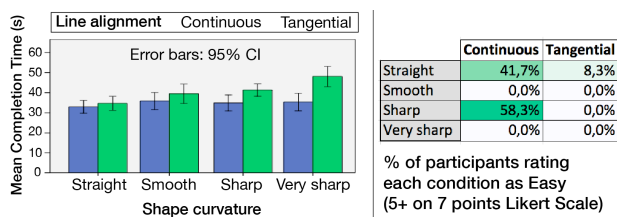
*Reading time and accuracy*

Results on completion time indicate a statistically significant effect for both Layout ( $F_{1,11}=18.1, p<.001$ ) and Shape ( $F_{3,33}=4.9, p=.002$ ), as well as an interaction between Shape and Layout ( $F_{3,33}=2.7, p=.042$ ). The observed power of the test was 0.99. Concerning Shape, a pairwise comparison only reveals a significant difference between Straight and Very Sharp shapes ( $p<.01$ ), with Straight being read an average of 7.8s faster than Very Sharp shapes (Figure 11). Concerning Layout, results reveal that Continuous layout is read an average of 6.1s faster than the Tangential layout. Post-hoc comparison shows that the Very Sharp curvature is read significantly slower (12.2s on average) if its layout is Tangential rather than a Continuous ( $p<.01$ ). There is no significant difference between layouts for each of the other shape curvatures. Layout and Shape had no effect on reading accuracy. On average, reading accuracy was 84.18%, (95% CI [82.1, 86.2]).

**Subjective results**

We found significant differences in user ratings among conditions ( $\chi^2(11)=83, p<.01$ ). Straight and Sharp shapes with continuous alignment were rated significantly different from all other conditions ( $p<.01$ ). Other shapes, except Straight with Tangential alignment, were never rated Easy (Figure 11). All users assigned the Very Sharp shape with tangential alignment a rating of 7 (very hard). Participants disliked the Smooth shape with Continuous layout due to sentences being cut (similar to holes in study 2).

Results regarding aesthetic properties show that only 10% of participants rated study 4 conditions as clean, beautiful, aesthetically pleasing, or interesting (except for the Sharp and Very Sharp tangential conditions, rated interesting by 54%).



**Figure 11. Left: Reading time for the different shapes with continuous or tangential layouts. Right: Perceived easiness.**

**Discussion**

Our third study reveals that shapes with continuous layout are easier to read than shapes with tangential layout in terms of reading speed, which confirms H8. Our study also partially confirms H9, as tangential text on the straight shape is easier to read than on the very sharp shape. Our results refute H10, as the straight shape was not read any slower with continuous than with tangential layout.

**FINAL DISCUSSION**

Here we discuss the implications of our findings, as well as current limitations and possibilities for future work.

**Guidelines for mapping text onto free-form shapes**

We investigated different mappings of text content onto free-form shapes based on a framework we defined using three axes: layout, token size, and line alignment. From our studies we can provide a set of design guidelines for optimizing text legibility on non-rectangular displays:

- Both left and right irregular alignments should be avoided, as text in these are perceived to be difficult to read and overall not aesthetic. Instead, symmetric shapes are preferred.
- Shapes with circular or sharp alignments are acceptable for presenting text: they are perceived to be easy to read, and overly clean, beautiful and interesting.
- If the shape contains a hole, text should be displayed using a broken layout with two columns around the hole to prevent any impact on reading performance.
- Shapes without holes are perceived to be less interesting than with holes. Thus, using holes in freeform shapes is

not only a solution to context requirements (such as the cooktop), but also an aesthetic feature to explore.

- To use dynamic scrolling on non-rectangular shapes, text should be resized so that each line contains the same amount of text. Otherwise, use page scroll with constant text size.
- While resizing text for dynamically scrolling is perceived as beautiful and clean, resizing text with page scrolling raises mixed results. Some users disliked it because of display space loss and of varying interline spacing. Thus, resizing text should be limited to dynamic scrolling.
- Shapes with continuous line alignment where lines are cut by the shape curvature should be avoided as they are perceived to be difficult to read and non aesthetic. This is similar to the effect of holes on continuous text. Even though tangential alignment does not affect reading performance on linear shapes, continuous text should be preferred as it reduces the perceived difficulty.
- Text on very sharp shapes should be avoided, as text on these is harder to read than on linear shapes. If used, such shapes should be filled with continuous text rather than tangential that impacts reading performance.

**Novel non-rectangular display usages and future work**

It is not clear that non-rectangular displays will replace traditional displays as the latter benefit from decades of interface optimizations. Aside from the highly publicized example of using non-rectangular displays in cars (Fig. 1-left), our focus groups revealed a broad range of usage contexts. In most cases, existing artifacts having non-rectangular features were suggested for text augmentation. Some examples included placing text on road signs, kitchen cooktops, pocket mirrors, puzzle pieces, bike handles, shoes, drink cans, and electric plugs, among others.

While our work represents a first step in identifying text legibility concerns on non-rectangular displays, other interactive tasks need to be carefully investigated. For example, certain shapes do not provide the necessary space for effectively flicking document content. Long and narrow objects could perhaps enable other flicking mechanisms, such as using the edges of the shape to displace content. Aside from navigating through a document, presentations including images alongside textual documents also present novel challenges. For instance, should images be cropped as one scrolls through a document, or should images use variable shapes to fit the contained display? Such questions merit further investigation and could impact the manner in which traditional artefacts, such as rectangular images, get re-engineered for non-rectangular displays.

Furthermore, there are other challenges to explore. In particular, our scope is limited to text and further work could be achieved to extend it to other UI contents such as 1D (in particular data with no carriage return such as timelines), 2D (such as maps), or 3D content. We would also like to use an eye-tracking device to further explore how free-form shapes affect text-skimming patterns.

**Limitless shapes, limitless rules**

We collected 41 different shapes across both focus groups, illustrating the diversity of shapes that could be augmented with text. From these, display shape properties emerged: symmetric displays tend to be chosen more often (Exp. 1, 2 and 3); few shapes with low length-to-width ratio were selected, i.e. they were long and thin (Exp. 4); some shapes contain holes (Exp. 2); and there is a large diversity of curvature, ranging from smooth to sharp boundaries (Exp. 1 and 4).

We observed no significant differences among different shapes in reading rates or accuracy despite users' subjective preference for shapes without irregular boundaries (Exp. 1). Further experiments are required to identify why our experimental setting did not impact reading performance, by using longer texts for instance. However we believe free-form display manufacturers should consider both reading performance as well as perceived difficulty, which means avoiding shapes with irregular alignments.

**Comparison to previous reading studies**

Our study methodology is based on the one adopted by Jankowski et al. [21]. It is thus interesting to compare our results in terms of variance among subjects (reported with CIs in our paper and SDs in [21]). Trends in both studies are consistent: in comparison to Jankowski et al's SD values, our study results show a larger deviation, on average 10s (31% of mean task completion time in our paper vs. 13% in their paper) probably due to our smaller number of participants (12 vs. 20). To check whether our population sample was large enough, we looked at the observed power of our studies (provided above for each study). The power was always above .80, which indicates enough statistical power (except for study 1 where we found no statistical difference on completion time).

**Limitations**

Our work is a first exploration of the multiple factors that can affect reading on free-form displays and is not intended to be exhaustive. We limited our study to certain shapes related to everyday use that were gathered through focus groups, but many different types of scenarios exist. Moreover, identifying shapes based on other factors, such as geometric properties would have resulted in different shapes being selected for the studies. More work is needed to explore the effects of other shapes and line alignments, other types of holes, or other resizing methods. Our studies are naturally limited by our typographical settings, and other fonts, text sizes, interline spacing or margin sizes could bring novel results. Obviously, testing all such combinations is not possible through an initial exploration such as ours, and so we leave these as possibilities for future investigation. Finally, our controlled setting does not simulate real reading conditions: reading on a public display or on a cooktop display implies reading from different angles or distances and using various display sizes. Many such conditions exist, and our results pave way for further in-depth explorations.

**CONCLUSION AND FUTURE WORK**

Emerging technologies are enabling the creation of non-rectangular displays. However, the introduction of such displays creates unprecedented challenges for designers who have to rethink new ways of creating user interfaces. Included among these challenges is the foremost concern of how to legibly present textual content, which is a chief concern in our paper. Our results agree with and extend upon other findings in the existing literature on text legibility, but they also uncover unique instances in which different rules need to be applied for non-rectangular displays. Finally, we mostly focused on output but there is much work to accomplish toward understanding how we can most effectively interact with free-form displays. We hope that our work will generate new research directions that will help to fill the bigger research agenda of free-form displays.

**ACKNOWLEDGMENTS**

A Leverhulme Trust Early Career Fellowship funded part of this work. We thank all study participants for their valuable time. We thank Brendan Li for his support in running the studies. We acknowledge the Canada Research Chairs award to the third author for supporting this work.

**REFERENCES**

1. Alexander, J., Lucero, A., and Subramanian, S. 2012. Tilt displays: designing display surfaces with multi-axis tilting and actuation. In *MobileHCI '12*. ACM, 161-170.
2. Bederson, B. 2000. Fisheye menus. In *UIST '00*. ACM, 217-225.
3. Benko, H., Andrew D. Wilson, and Balakrishnan, R. 2008. Sphere: multi-touch interactions on a spherical display. In *UIST '08*. ACM, 77-86.
4. Bernard, M., Chaparro, B., Mills, M., and Halcomb, C. 2003. Comparing the effects of text size and format on the readability of computer-displayed Times New Roman and Arial text. *IJHCS*. 59, 6, 823-835.
5. Bouma, H. 1980. Visual Reading Processes and the Quality of Text Displays. In *Ergonomic Aspects of Visual Display Terminals*, Grandjean and Vigliani, 101-114.
6. Brockmeyer, E., Poupyrev, I., and Hudson, S. 2013. PAPILLON: designing curved display surfaces with printed optics. In *UIST '13*. ACM, 457-462.
7. Brown, J.A., Fishco, V.V., & Hanna, G. 1993. *Nelson-Denny Reading Test*. Riverside Publishing.
8. Butz, A., Grob, M., and Krüger, A. 2004. TUISTER: a tangible UI for hierarchical structures. In *IUI '04*. ACM, 223-225.
9. Chevalier, F., Dragicevic, P., Bezerianos, A. and Fekete, J.Dchi 2010. Using text animated transitions to support navigation in document histories. In *CHI '10*. ACM, 683-692.
10. Darroch, I., Goodman, J. Brewster, S. and Gray, P. The Effect of Age and Font Size on Reading Text on Handheld Computers. In *INTERACT 2005*.



11. Delaunay, F., Greeff, J., and Belpaeme, T. 2010. A study of a retro-projected robotic face and its effectiveness for gaze reading by humans. In HRI '10. IEEE Press, 39-44.
12. Dillon, A. Reading from paper vs. screens: A critical review of the empirical literature. *Ergonomics*, 1992.
13. Dyson, M. 2004. How physical text layout affects reading from screen *Behaviour & Information Technology* Vol. 23, Iss. 6, 2004.
14. Dyson, M., and Haselgrove, M. 2001. The influence of reading speed and line length on the effectiveness of reading from screen. *Int. J. Hum.-Comput. Stud.* 54, 4, 585-612.
15. Dyson, M., Kipping, G. 1997. The legibility of screen formats: are three columns better than one?, *Computers & Graphics*, 21, 703-712.
16. Gregory, M. and Poulton, E.C. 1970. Even versus Uneven Right-hand Margins and the Rate of Comprehension in Reading. *Ergonomics*, Volume 13, Issue 4, pages 427-434.
17. Grossman, T., Wigdor, D., and Balakrishnan, R. 2007. Exploring and reducing the effects of orientation on text readability in volumetric displays. In CHI '07. ACM, 483-492.
18. Hansen, W. J. and Haas, C. 1988. Reading and writing with computers: a framework for explaining differences in performance. *Communications of the ACM*, 31, 1080-1089.
19. <http://sharp-world.com/igzo/>
20. Huey, E. 1898. Preliminary Experiments in the Physiology and Psychology of Reading. *The American Journal of Psychology*, vol. 9, pp. 575-586.
21. Jankowski, J., Samp, K., Irzynska, I., Jozwicz, M., and Decker, S. 2010. Integrating Text with Video and 3D Graphics: The Effects of Text Drawing Styles on Text Readability. In CHI '10. ACM, 1321-1330.
22. Jones, B., Benko, H., Ofek, E., and Wilson, A. 2013. IllumiRoom: peripheral projected illusions for interactive experiences. In CHI '13. ACM, 869-878.
23. Jorna, G.C. and Snyder, H.L. Image Quality Determines Differences in Reading Performance and Perceived Image Quality with CRT and Hard-Copy Displays. *Human Factors*, 1991.
24. Kolers, P., Duchnicky, R. and Ferguson, D. 1981. Eye movement measurement of readability of CRT displays. *Human Factors*. 23(5), 517-27.
25. Koppel, M., Bailly, G., Müller, J., and Walter, R. 2012. Chained displays: configurations of public displays can be used to influence actor-, audience-, and passer-by behavior. In CHI '12. ACM, 317-326.
26. Koriat, A. and Norman, J. 1985. Reading Rotated Words. *Journal of Experimental Psychology; Human Perception and Performance*, vol. 11, number 4, pp: 490-508.
27. Lavie, T. and Tractinsky, N. Assessing dimensions of perceived visual aesthetics of web sites. *IJHCS*, 2004.
28. Lyons, K., Nguyen, D., Ashbrook, D., and White, S. 2012. Facet: a multi-segment wrist worn system. In *UIST '12*. ACM, 123-130.
29. Maharik, R., Bessmeltsev, M., Sheffer, A., Shamir, A., and Carr, N. 2011. Digital micrography. In *SIGGRAPH '11*. ACM, Article 100, 12 pages.
30. O'Hara, K. and Sellen, A. 1997. A comparison of reading paper and on-line documents. In *CHI '97*. ACM, 335-342.
31. Simon Olberding, Michael Wessely, and Jürgen Steimle. 2014. PrintScreen: fabricating highly customizable thin-film touch-displays. In *UIST '14*. ACM, 281-290.
32. Piolat, A., Roussey, J.-Y. and Thunin, O. 1997. Effects of screen presentation on text reading and revising. *IJHCS*, 47, 565-589.
33. Poupyrev, I., Newton-Dunn, H., and Bau, O. 2006. D20: interaction with multifaceted display devices. In *CHI EA '06*. ACM, 1241-1246.
34. Ramakers, R., Schöning, J., and Luyten, K. 2014. Paddle: highly deformable mobile devices with physical controls. In *CHI EA '14*. ACM, 191-192.
35. Rekimoto, J., Ullmer, B., and Oba, H. 2001. DataTiles: a modular platform for mixed physical and graphical interactions. In *CHI '01*. ACM, 269-276.
36. Roudaut, A., Karnik, A., Löchtefeld, M., and Subramanian, S. 2013. Morphees: toward high "shape resolution" in self-actuated flexible mobile devices. In *CHI '13*. ACM, 593-602.
37. Roudaut, A., Reed, R., Hao, T., and Subramanian, S. 2014. Changibles: analyzing and designing shape changing constructive assembly. In *CHI '14*. ACM, 2593-2596.
38. Schriver, K. 1997. *Dynamics in Document Design: Creating Text for Readers*. John Wiley & Sons, Inc., New York, NY, USA.
39. Spencer, H. 1968. *The Visible Word*. London: Royal College of Art.
40. Tinker, M. 1972. Effect of angular alignment upon readability of print. *Journal of Educational Psychology*, vol. 47, pp. 358-363.
41. Tractinsky, N., Shoval-Katz, A., and Ikar, D. What is beautiful is usable. *Interacting with Computers*, 2000.
42. Wigdor, D. and Balakrishnan, R. 2005. Empirical investigation into the effect of orientation on text readability in tabletop displays. In *ECSCW'05*. Springer, 205-224.
43. [www.lgdnewsroom.com](http://www.lgdnewsroom.com)
44. Zachrisson, B. 1965, *Studies in the Legibility of Printed Text*. Almqvist & Wiksell, 225 pages.

# Visual Composition of Graphical Elements on Non-Rectangular Displays

Marcos Serrano\*

University of Toulouse - IRIT  
Toulouse, France  
*Marcos.Serrano@irit.fr*

Anne Roudaut\*

University of Bristol  
Bristol, United Kingdom  
*Anne.Roudaut@bristol.ac.uk*

Pourang Irani

University of Manitoba  
Winnipeg, Canada  
*Irani@cs.umanitoba.ca*

## ABSTRACT

Graphical user interfaces are composed of varying elements (text, images, etc.) whose visual arrangement has been relatively well established in the context of rectangular interfaces. The advent of non-rectangular displays questions this knowledge. In this paper we study how traditional content layouts can be adapted to fit different non-rectangular displays. We performed a first qualitative study where graphic designers fitted text and images into different non-rectangular displays. From the analysis of their output we generalize and adapt ten composition principles that have been proposed in the literature for rectangular displays. We evaluate the revised principles through a paired comparison questionnaire where 57 participants compared pairs of layouts. Using the Bradley-Terry-Luce model to analyze our data we show that some results contradict current conventions on visual design for rectangular displays. We then extracted the most interesting cases and conducted a follow up study with additional shapes to investigate how the principles generalize. From these results we propose a set of guidelines for designing visual content for non-rectangular displays.

## Author Keywords

Freeform display; non-rectangular display; visual design guidelines.

## ACM Classification Keywords

H5.2 [Information interfaces and presentation]: User Interfaces - Screen design.

## INTRODUCTION

Emerging technologies allow for the creation of non-rectangular displays with unlimited constraints in shape<sup>1</sup>. These displays can be applied to a wide variety of usage contexts, ranging from in-vehicle [18] and wearable displays [1] to public displays [31]. In these applications non-rectangular displays meet different needs for which traditional displays are not well-suited at. For instance, a

single non-rectangular display can replace current instrument panels on car dashboards. Non-rectangular displays will also facilitate inserting displays on non-rectangular objects, furniture and/or urban architecture. In the context of mobile computing, non-rectangular displays can adopt shapes which will better fit wearable devices or replicate existing jewellery. These applications correspond to a market in expansion that could impact millions of users in the near future [9,19]. With this eminent adoption comes the urgent challenge of rethinking the way we present content on non-rectangular displays [31].

Presenting content on rectangular displays is, in contrast, reasonably well understood. For instance Galitz [15] proposed ten aesthetic composition guidelines extracted from tacit knowledge that visual designers have accumulated over years of experience (e.g. balance, proportion or unity). There is also strong evidence to support the role of aesthetics in interface design and especially how they impact perceived usability [8,21,32] and visual search performance [30].

However non-rectangular displays challenge these fundamental composition principles. Recent work [31] demonstrated that different ways of presenting text on non-rectangular displays affect usability (reading performance and subjective preference). We therefore postulate that it will also be the case for presenting more complex content types combining images and texts, such as online newspapers. Our goal is thus to investigate if the established composition principles for traditional displays generalize to non-rectangular displays, and if not how they can be adapted for designers to create the layout of the next generation of display content.

To this end, we first performed a qualitative study that consisted in asking graphic designers to map traditional web content onto non-rectangular shapes. The analysis of the results showed that graphic designers' inner sense for composing layouts matches existing composition principles (simplicity, sequentiality, economy and unity) but that some revisions (balance, regularity, proportion, predictability) are needed; for these, we discuss how they can be generalized to non-rectangular displays and propose a set of hypotheses.

We then performed a pairwise experiment with 57 participants to test the revised composition principles. Using a Bradley-Terry-Luce model to associate subjective

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

\* Both authors contributed equally to this work



metrics to different visualizations we found that vertical symmetry is significantly better than with other axes. Among other results we found that adapting the grid orientation to the shape (radial for circle, oriented for triangle) makes the layout visually more pleasing. And placing the menu at the bottom of some shapes (e.g. circle) is better.

We also performed a follow-up pairwise experiment with additional shapes to investigate how the principles generalize. The study confirmed our findings on grid layout (adapting the layout to the shape can improve its clarity), on element shapes (shaping elements to the shape of the display is visually attractive) and on menu position (menu on top of the display is not always preferred).

Our contributions are (1) a set of visual composition guidelines for non-rectangular displays that extend existing guides on rectangular displays using webpage productions gathered in a qualitative study with graphic designers; (2) a pairwise experiment which assesses these guidelines; (3) a follow-up pairwise experiment investigating how the principles generalize to other shapes.

#### RELATED WORK

Our work relates to page layout, visual design principles and visualizations on non-rectangular interfaces.

##### Page layout guidelines

In graphics design, *page layout* consists in arranging visual elements on a page. Page layout has deeply evolved with printing and editing technologies [25]: physically assembling characters in letterpress printing (Renaissance), gluing papers in paste-up techniques to create pages that are later photographed (mid-20th century), and finally using Wysiwyg publishing software with the advent of computers. All these stages have influenced page layouts, which are most often based on rectangular grids [12,26].

Page layout involves design elements and choices, such as margins, images size and position. In grid design, these choices are based on predominant grid styles that have evolved over time, and on principles of page construction. These principles apply rules, such as the golden section [17], to divide a page in pleasing proportions. These rules are generally implemented in modern editing software. Using these rules, theoretical work in the field of graphic design has analyzed the proportioning systems of works of art, buildings or products with diverse shapes [11].

##### Visual Design principles

Visual layout guidelines have been to a large extent based on Gestalt Laws [10,16]. These laws mainly describe how visual elements are perceived as grouped based on principles of Proximity, Similarity (color, shape and size), Common fate, Good Continuation and Connectedness. These principles have been widely appreciated in various fields including architecture but also widely used in GUI design and information visualization [13,14,20,37].

Galitz [15] proposed 10 visually pleasing composition guidelines for GUI design: Balance, Symmetry, Regularity, Predictability, Sequentiality, Economy, Unity, Proportion,

Simplicity, and Groupings. These principles summarize tacit knowledge developed by graphical designers over decades of practice. These principles encompass many page layout and visual perception principles (e.g. Gestalt).

Ngo et al. [27,28] proposed to use the 14 aesthetic features inspired from Galitz [15] and to associate a metric. However the work only focuses on a few aspects defined by Galitz while leaving out others. E.g. the authors did not consider the color or the shape of the elements, which are known to have an impact on the way we look for information. There are also inconsistencies in some metrics. E.g. Galitz defined “sequentially” as a way to arrange elements to guide the eye through a UI (e.g. bigger, colored saturated before small, black and white, unsaturated). However Ngo’s metric measures if most of the big elements are in the upper part (vs. bottom and lower right). For these reasons our work uses Galitz’ [15] principles because they clearly define and encompass all the important factors.

##### Interface aesthetic and usability

Norman, a father’s of *usability design*, argued “good design means that beauty and usability are in balance” [29]. Work by Kurosu et al. [21] on the influence of aesthetics on usability showed a strong correlation between the two [21]. The results were confirmed by other studies [8,32]. Later work has investigated the effect of aesthetics on inherent usability, i.e. performance [30, 35]. Van Schnaik et al. [35] discovered that aesthetically pleasant web pages reduce error rate. Brewster et al [30] evaluated some of Ngo’s aesthetic metrics [27,28] and demonstrated their effect on search performance. In our work, we perform the first evaluation of non-rectangular layouts using aesthetic measures such as beauty, clarity and symmetry [15].

##### Non-rectangular interfaces

Previous work on tabletop interfaces explored the use of non-rectangular graphical widgets to avoid hand occlusion or improve user collaboration. For example Cotting et al [6] proposed mapping rectangular widgets into bubbles having physical properties and that can be elastically deformed. Their work mostly focused on these elastic properties and how to adapt the content by using display warping.

Work has been carried on specific interfaces for round smartwatches. Xu et al. presented two different prototypes based on the use of LEDs around the bezel [36]. The first prototype used 4 LEDs to render four applications while the second one used 12 LEDs to show time and app data. Commercial round smartwatches use different approaches to present information. Apple Watch uses a circular honeycomb grid to present applications. This grid uses a fisheye view to provide context and focus: only items in the center can be tapped. What we learn from the above work and current interfaces, is that significant effort is needed to adapt UIs from rectangular to circular shaped displays.

Recently, Serrano et al. investigated text legibility on non-rectangular displays [31]. They presented scenarios in which non-rectangular displays reveal useful information.

They also conducted four studies to assess ten hypotheses on reading performances of text displayed on free-form displays. Their work shows that text reading performance is affected if not carefully designed for, and the authors provides a number of guidelines for text justifications based on display shapes [31]. Our work builds on these previous studies on non-rectangular displays and investigates more complex content types combining images and texts.

#### ELICITATION STUDY

To begin our exploration on the visual composition of graphical elements on non-rectangular displays, we first captured how graphical designers tacitly organize visual layouts. We gathered qualitative probes that we could use to generate new hypotheses for the visual composition of elements on free-form displays. We asked five graphic designers to compose webpages on non-rectangular shapes. This task relies heavily on creativity and thus we designed this study to be in the form of homework. Designers had a week to do the task wherever they wanted, thus avoiding us to interfere with any creative processes that might emerge from their environment. We then analyzed the productions to better understand which choices they made.

#### Design task

We gave participants a webpage with all associated content. We asked them to fill this content into four shapes: a circle and a triangle, with or without hole (see Figures 1, 2 and 3). These shapes come from prior work [35] and are based on geometrical shapes, which will make it easier to classify or extend our results. The initial rectangular area and the target non-rectangular shape had the same area (450cm<sup>2</sup>), to ensure all content could fit. To facilitate the task and ensure that all designers had the same working environment, we gave designers a vector file ready to use: the webpage content (text, images, colors) was ready to be manipulated and the four final shapes were included in the file.

We choose a news webpage as it combines a variety of graphical elements. To limit the length of the task and increase the results quality we picked a subset of shapes based on previous studies revealing that they could be used in upcoming scenarios [31] (e.g. free-form displays on road signs, circular pocket mirrors, non-rectangular coffee tables or in vehicles as well as displays with holes such as embedded displays in cooking hobs).

#### Participants

We recruited 5 professional graphic designers, aged 27.8 years on average (SD=6) from 3 different countries (USA, France and UK). They worked on graphic design for 6 years on average (SD=3). Their areas of expertise were print (5/5), web (4/5), and UI (2/5) design. We offered a 50£ voucher prize to a randomly chosen participant.

#### Procedure

We first informed designers about the project. Designers were given one week to complete their task. They could take all the time needed to perform it. They could use their preferred tool: all of them used Illustrator.

#### Collected data

We collected all resulting graphical designs. We also asked them to fill a questionnaire to gather information such as the time they took or if they had any specific strategy.

#### Questionnaire results

Designers completed the task in 90 min (SD=37). We asked them how satisfied they were on a 5-point Likert scale. They rated between neutral and positive the results on the circle, rather neutral with the triangle without hole, but were not satisfied with designs on the triangle with hole. These results are in line with the perceived difficulty. All designers found that shapes with hole were more difficult to fill. D4 stated that the triangle with hole “lacked symmetry”. D3 argued, “shapes with holes are more difficult as you hardly can place a unit of content”. D1 said, “much space was lost with the hole”, even though all shapes had the same area. (4/5) designers agreed that the easiest shape was the circle without hole. D4 suggested, “circular shapes could actually be an advantage in creating an interesting layout”. D3 preferred the triangle without hole as she felt “you can separate icons from content”.

#### Productions results

We choose to perform their analysis in the next section to correlate our observations with previously established principles on rectangular displays.

#### GENERALISATION OF COMPOSITION PRINCIPLES

We present in detail the composition principles proposed by Galitz [15] and discuss how the definition can be adapted to non-rectangular displays. We use the productions of the elicitation study as probes and highlight hypotheses that we evaluate in the studies presented in the rest of the paper.

##### 1. Balance/Symmetry

*Original definition* [15]: Balance (stability) means providing an equal weight of elements on each side of the horizontal or vertical axis. Heavier elements are the larger ones, with dark colours or unusual shapes. Galitz does not explicitly mention what regular shapes are. Our assumption is that rectangular topologies (including rounded corners) are implied because they represent most of the current GUI layouts. Symmetry is a subcase of Balance where the Balance is respected for both horizontal and vertical axis.

*Generalization*: There are two aspects that can change the way we define Balance when moving to non-rectangular displays: (1) the symmetrical axes of the display and (2) the definition of a “regular” shape. We have no reason to think that other aspects of the original definition should change (for the color and the size of the elements).

For the symmetrical axes, it is tempting to say that the balance should use natural axes of symmetry of the displays (e.g. medians in triangle) rather than the horizontal/vertical axes. However, examples produced by our designers suggest that balance should follow the vertical axis (Fig.1). This may stem from the fact that in Western cultures we read from left to right. The only example deviating from this is shown in Figure 2-left. We can see that designer (D1)

used a radial composition, however we can note that s(he) used circular shapes for the components as well, thus emphasizing the Balance with the shape of element.



**Figure 1. Balance:** productions tend to follow a vertical symmetry axis with only one exception for the case of the hollow circle shown in Figure 2-left.

Most element shapes in the productions were rectangular except for the designs show in Fig.2. In these two designs the elements shapes were directly related to the shape of the display (the designed cut the elements in circle or triangle). For the circular designs it could also mean a desire to copy the shape of the hole itself. None of the productions are fully symmetrical, although there is rather a strong trend for vertical symmetry. Thus we do not think that full symmetry necessarily needs to be followed. This corroborates principles on rectangular displays where it is often advised to “break” the symmetry.



**Figure 2. Balance:** although most elements were rectangular, two examples use different shapes, but their topologies are directly related to the shape of the display.

*Summary:* In summary Balance should not change much from the original definition. The balance should still be done around a horizontal symmetry and should be done for all the difference color, size and shape of elements. We can hypothesize that:

- H1. Balance using vertical axis is better than using some of the natural axes of symmetry.
- H2. Balance using vertical axis is better than using all of the natural axes of symmetry.
- H3. Balance using natural symmetry axes is better if element shapes follow display or hole shapes.

**2. Regularity**

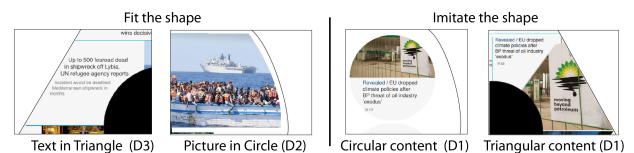
*Original definition* [15]: Regularity (consistency) means providing consistently spaced horizontal and vertical alignment points and spacing, as well as using similar element sizes, shapes and colors overall.

*Generalization:* Similar to Balance, there are two aspects that can potentially change the way we define Regularity when moving to non-rectangular displays: (1) the alignment axes which can be more than just horizontal/vertical and (2) the definition of a “regular” shape. We do not have reasons to believe that the second part of the definition should change (similar elements sizes and colors).

Concerning the alignment axes, it is possible to imagine different layouts and to deviate from the rectangular grid. In fact only 10/20 productions used a rectangular grid, 4 used a radial alignment and 6 used a tangential alignment (aligned with one or more edges of the screen). Interestingly the 4 designs using a radial alignment (Fig.2) had a hole suggesting using the hole as part of the design. Also the 6 using tangential alignment are triangles (Fig.3) suggesting that using the edges of the display shape as alignment point is not only obvious but also a good technique. E.g. D4 said “I tried to use the shape to my advantage in terms of following the lines. Also keeping a grid pattern made it easier to place items”. We did not observe any design using more than one concentric alignment (e.g. 2 concentric circles in the circular display or smaller triangles inside the triangular one). This may be due to the size and amount of elements we provided (not enough elements to use two concentric alignments), so we cannot conclude anything. However, we note that on the design with circles, the designer choose to decenter the elements to avoid a vertical symmetry. This corroborates the fact that full symmetry is not necessarily desirable.



**Figure 3. Regularity:** designs following a tangential alignment are made on triangular shapes only.



**Figure 4. Designers** reshaped content elements to fit the shape (left) or to enhance aesthetics by imitating the shape (right).

Concerning the regularity of element shapes we observed that designers reshaped the elements for two main purposes: to fit the shape and for aesthetics. Note that, in our instructions, we did not suggest that elements (images, text) could be reshaped, as we did not want to influence the designers’ creativity and the resulting designs. To fit the shape, designers aligned text with the shape bezel and/or cut the images (Fig. 4-Left). Out of the four designers, one never reshaped content (D4) and kept the original rectangular shape of texts and images. One of the designers, D1, reshaped content to imitate the shape: texts and images were made circular in the Circle shapes, and triangular in the Triangle shape (Fig. 4-Right). While this approach allows filling the Triangle efficiently, it seems to be mainly aesthetic in the case of the Circle. It is possible that, although images are cut, the regularity is preserved thanks to perception clues (e.g. Gestalt continuity).

*Summary:* Regularity should differ from the original definition regarding the alignment and the regularity of the shape of elements. We hypothesize that:

- H4. It is better to keep regularity via a rectangular grid than via different grid shape.
- H5. It is better to keep regularity over keeping shape aspect (rather cutting it on the edges).

### 3. Predictability

*Original definition* [15]: Predictability (conventions) means providing conventional orders or arrangements to ensure that viewing a screen (or part of it) enables one to predict how another will look (or the rest of it). Predictability is also enhanced through design Regularity.

*Generalization:* This guideline rather links the way several pages are designed and thus we have no reason to think that this should change with the shape of the device. To a large extent we can also couple this definition with some of Nielsen's guidelines of "consistency across platform", e.g. the fact that headers and menus are always at the top of a webpage. While all designers decided to keep the regular menu position at the top of the shape for the Circles (Fig. 1-2 left), most of them inversed the position for the Triangles (Fig. 1-2 right). This change is rather surprising, given that it goes against traditional web page layouts.

*Summary:* In summary the predictability should stay the same than for rectangular display except for the placement of the main elements such as menus. We hypothesize that:

- H6 Menus should be placed at the widest horizontal portion.

### 4. Proportion

*Original definition* [15]: Proportion means using aesthetically pleasing proportions for components of the screen. E.g. some rectangular shapes are more aesthetically pleasing than others [24] (square, square root of two, golden rectangle, square root of three or double square). Other guidelines also recommend using non-regular grids that follow a golden number layout [15].

*Generalization:* Common sense would favor the use of the same pleasing ratio for other shapes than the rectangle, e.g. golden ovals. More interestingly it is interesting to see how the golden number layout would match with non-rectangular content. Figure 7 illustrates how it would work with a circle and a triangle. Interestingly none of the productions seem to use such a layout but this may be due to the fact that they had to reshape an existing interface with a specific number of items. It would be interesting to see how they perform with a non-given interface (free to choose which elements to place).

*Summary:* The pleasing proportion for elements should not change but it could change for the overall layout that could for instance follow the golden number ratio. We could hypothesize that "using a golden number layout is better than a random one if Regularity and Simplicity are kept similar". However we will not investigate this issue further as it is hardly applicable to an informational webpage design and rather fit to more artistic content.

### 5. Simplicity

*Original definition* [15]: Simplicity (opp. complexity) means optimizing the number of elements on a screen, within limits of clarity, as well as minimizing the alignment points, especially horizontal or columnar. Several researchers have proposed to use information theory to provide a metric to this principle [1,33,34].

*Generalization:* We do not think that the definition should change with the number of elements, however the measure of alignment points should change to better reflect the fact that the element could be aligned differently (as already explained earlier in Regularity). Instead of using row and column we could thus use radial and tangential alignment.

*Summary:* We postulate that Simplicity should only be slightly different from the original definition regarding the alignment lines count. We could hypothesize that "the smaller the number of alignment points the better, whether these points are aligned on a grid or on a radial/tangential form". However, it is obvious that this guideline will have an impact on visual search (as the number of on screen elements increases) and thus we will not investigate it.

### 6. Sequentiality

*Original definition* [15]: Sequentiality means arranging elements to guide the eye through the screen in an obvious, logical, rhythmic, and efficient manner. The eye tends to be attracted by certain features (brighter, isolated or bigger elements, graphics, highly saturated colors) and then move to others. Sequentiality is enhanced through Grouping.

*Generalization:* We have no reason to believe that this should change. The definition of "unusual" or irregular shape could corroborate what had already been defined for Balance, i.e. those regular shapes are either rectangular or similar to the shape of the displays and/or the holes.

### 7. Economy

*Original definition* [15]: Economy means providing few styles, techniques, and colors in order to deliver the message as simply as possible (e.g. no ornamentation).

*Generalization:* This principle is straightforward and true for all UIs. We have no reasons to believe this should change. Designers' productions did not show any major removal or addition of elements.

### 8. Unity/Grouping

*Original definition* [15]: Unity (coherence) and Grouping means using similar sizes, shapes, or colors for related information as well as leaving less space between elements than the space left at the margins. With unity, the elements seem to belong together.

*Generalization:* We have no reason to believe this should change. The productions did not show any new tendencies in term of groupings apart from using different shapes (triangles or circles) for elements. D3 did not identify a particular strategy to place content. D1's strategy was to keep the content associations: images with text, text with icons. D2 and D4 agreed to put the most important content

first (pictures): “Picture were given priority. I wanted to make sure the main focus of the picture wasn’t cut out or disturbed.” (D4); “I prioritize the photos to catch the eyes” (D1). D4 added, “I gave the icon a hierarchy of which would be most important to the audience; and then made sure they were placed in prominent positions or made larger”.

**CHOICE OF EXPERIMENTS**

To evaluate the differences between the composition principles we opted for paired comparison experiments that are typical to gather Quality of Experience (QoE) feedback. They consist in asking participants to choose between two conditions, here two layout visualizations. The experiment is designed so that each participant rates each pairs of visualizations. Here, we explain why this experimental design is adapted to our research question.

**Confound variables in performance based studies**

Quantitative controlled experiments are assumed to be the best tool to demonstrate UI efficiency. But they are only useful when it is possible to test counterbalanced variables without introducing confounds. With our research question we could not find an experimental setup following this rule. E.g. in readability study (e.g. [38]), the issue is that we know text presentations affect readability [31]. Thus any effects observed could also be a result of the text presentations that change according to the conditions (confound variables). One solution is to remove the need for text and use visual search tasks. But complex images, including photos, have the same issues than text (confound variable) because different brightness, color, saturation and content affect perception [15]. The other recourse is to use symbols like in [30]. But the task is becoming abstract and ecologically distant from the initial question on page layout.

**Paired comparison vs. Mean opinion score**

A way to get participants’ input is to use subjective judgement. Estimating preferences based on subjective judgements is a critical step in psychological experiments with applications in many fields such as marketing, environmental sciences and health economics [5]. In particular pairwise experiments have been widely used. In such studies, two conditions are presented to participants who then indicate one alternative over the other. Pairwise comparison ratings have been proven to produce more realistic results than asking for individual rankings (e.g. using a Likert scale) [39]. In particular there are powerful mathematical model, such as the Bradley-Terry-Luce [23] that can deduce an “ability” metric from the data as well as perform classical statistical hypothesis testing.

**Crowdsourc vs. not**

Paired comparison experiments are also used in crowdsourced experiments. In such cases participants generally only rate a subset of all possible combination of conditions but only a subset. Although several models have been proposed to accommodate these reduced amounts of data [6], having all possible combinations increases the statistical power of the results and it is possible to compute an individual and group consistency (see results).

**PAIRWISE EXPERIMENTS**

The goal of this experiment is to validate or invalidate the hypotheses laid out in the previous sections.

**Task**

We asked participants to compare pairs of layout visualizations and say which one was nicer (i.e. visually pleasing), clearer (i.e. not confusing) and more symmetric (aesthetics term proposed in [22]). Participants could give three answers for each question: Visualization-1, Visualization-2 or Both.

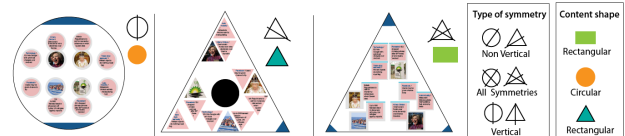
**Shapes**

We used the same shapes as in the qualitative study, i.e. a circle and triangle, with and without a hole (C, T, CH, TH).

**Design**

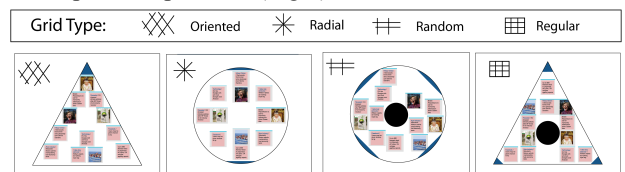
We gave one general survey to the participants but it comprises four sub-surveys matching different hypotheses. For each survey, we compared visual compositions among shapes, but not between shapes. In each survey, the order of the trials for one shape was counterbalanced. The order of the sub-surveys was always the same (H123, H4, H5, H6). Our study was composed of  $60 + 24 + 4 + 12 = 100$  pairs.

- *Sub-survey 1. Balance and symmetry (H1, H2 and H3):* we studied 3 symmetry axes (vertical, shape and all) and 2 element shapes (rectangular or matching display shape). Overall, we had 15 pair comparisons for each shape  $\times$  4 shapes = 60 pair comparisons (Fig.5).



**Figure 5. Conditions to test H1, H2 and H3 and three examples: a) vertical symmetry with circular content; b) non vertical symmetry with triangular content; and c) all axes of symmetry with rectangular content.**

- *Sub-survey 2. Regularity (H4):* we studied four grid layouts (regular, radial, oriented and random). Overall, we had 4 pair comparisons for each shape  $\times$  4 shapes = 24 pair comparisons (Fig.6).



**Figure 6. Conditions to test hypothesis H4.**

- *Sub-survey 3. Regularity (H5):* We tested whether it was better to follow the regularity but have elements cut by the display shape or to break the regularity by having the elements fit the shape. We tested 2 conditions (elements out or in). Overall, we had 1 pair comparison for each shape  $\times$  4 shapes = 4 pair comparisons (Fig.7).
- *Sub-survey 4. Predictability (H6):* We changed the position of the menu. We tested 3 positions for the menu (top, bottom and following the shape). Overall, we had



3 pair comparisons for each shape  $\times$  4 shapes = 12 pair comparisons (Fig.8).



Figure 7. Conditions to test hypothesis H5. Left: elements fit in the display by breaking their regularity. Right: elements are cut to preserve their regularity.



Figure 8. The 3 menu positions explored in our study (H6): bottom, shape and top.

### Participants

57 participants (16 female) from 8 different countries completed the study. Participants were aged 22.4 on average (SD=6). 49 of them were university students and the rest worked in academia as engineers (2), professors (2) or researchers (4). We removed results from 4 participants due to their low consistency (see step1 below).

### Collected data

We collected 100 pairs  $\times$  3 questions = 300 answers per participant so a total of 300  $\times$  57 users = 17100 answers.

## RESULTS

Our analysis consists in three steps based on [5].

### Step 1: Individual consistency checking

We computed the Transitivity Satisfaction Rate (TSR), which quantifies the consistency of a participant's judgments over multiple questions. E.g. if A is found more restrictive than B, and B more than C, then we should have A more restrictive than C. We implemented the algorithm found in [5] in Python. We removed 4 participants whose TSR was below 0.8. The mean TSR for all other users was 0.92 (SD = 0.05) and at least over 0.8 for all of them, thus denoting that they paid full attention to the study [5].

### Step 2: Overall consistency checking

To test the overall consistency across participants we checked the stochastic transitivity properties or computed Kendall's  $\mu$ -coefficient [5]. For each participant, we computed a list of rankings of visualisations and used the kendalltau Python library to produce a coefficient for each pair of participants, computed as a percentage (100% means all participants perfectly agreed, 0% they perfectly disagreed). Our results show that the mean Kendall's  $\mu$ -coefficient is above 50% for all conditions.

### Step 3: Model the data

The individual and overall consistencies were confirmed, so we proceeded to model the data. We used the Bradley-Terry-Luce model [4,23], which associates an "ability" metric to each condition that have been paired-compared as well as the p-value for each pair comparisons (BradleyTerry2 R package). The results are presented

below for each comparison set. Note that the Bradley-Terry-Luce model computes a p-value that express how the visualizations compare to one specific visualization only, which serves as reference and is a parameter of the formula. We thus performed several tests to compute the significant level for each comparison. To counteract the problem of making multiple comparisons tests we used a Bonferroni correction for each result described below.

For sake of clarity, our figures only represent the results for the shape without holes. As results are similar for both types of shapes (with and without hole), we detail in the text any differences between them. We add all the figures corresponding to the conditions with a hole to the Annexe of our results (p-values) for each pair of conditions compared in our survey (adding it to our figures would make them unreadable). Each of the following figures shows the results representing the metric of each visualization computed via the Bradley-Terry-Luce model [4,23]. The metric gives a value between 0 and 1, where the lowest condition equals 0 and the highest 1. We also indicate the standard error values given by the model. We now detail the results for each hypothesis.

Display shapes are abbreviated as C circle T for triangle, CH for circle with a hole and TH for triangle with a hole.

### Results for Balance and symmetry (H1, H2, H3)

#### Type of symmetry

Visualizations with vertical symmetry were deemed nicer, more symmetric and clearer than the ones with symmetry around a non-vertical axis or around various axes (except when using all axes with triangular content on T, which was no different than a vertical symmetry with rectangular content). The conditions involving symmetry around the shape axis were always rated worse than the others (for C we used vertical/horizontal axes).

#### Content shape

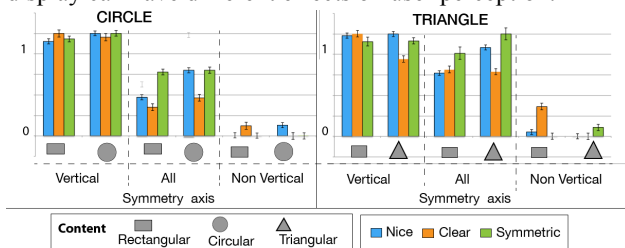
Overall the results of layout visualization with rectangular or shaped content are similar, but in cases the visualizations with shaped content was deemed nicer than the ones with rectangular content (Figure 9): on C for any type of symmetry; and on T when using all axis of symmetry. On the contrary, the layout visualizations with rectangular content were perceived as clearer than the ones with shaped content (with symmetry around the shape axis on T). Interestingly, the shape of the content sometimes influenced the perception on symmetry: for T, when the content was triangular, it was rated as more symmetric than when it was rectangular with the symmetry around the shape or all axes; and clearer with the symmetry around vertical and shape axes.

#### Display shape

The main difference between C and T was that on C, circular content was rated nicer than rectangular content (for the same type of symmetry), while on T, triangular content was rated as clearer and more symmetric than its



rectangular counterpart (for the same type of symmetry). This illustrates that shaping the content to imitate the display can have different effects on user perception.



**Figure 9. Results on H123 for the displays without hole. The Y-axis represents the mean rating for pair comparisons: blue for nice, orange for clear and green for symmetric.**

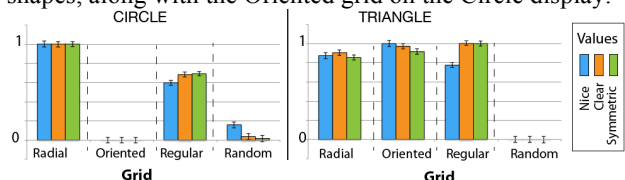
**Summary**

These results validated H1 (*balance using vertical axis is better than the Shape axes*), H2 (*balance using vertical axis is better than using all axes*) and partially H3 (*balance using all axes is better if the elements follow the shape of the displays*): results show that all symmetry axis with shaped content is nicer on C, and more symmetric on T.

**Results for Regularity and proportion (H4)**

**Grid type**

An interesting finding is that the regular grid, which we expected to be preferred, is not always rated best (Figure 10): for C, the radial grid was nicer, clearer and more symmetric; for T, the oriented grid was nicer, although there was no difference on the other measures. As expected, the random grid was the least preferred across all display shapes, along with the Oriented grid on the Circle display.



**Figure 10. Results for H4 for the displays without hole.**

**Display shape**

There was an interesting difference in how grids were rated between C and T: in C, the radial grid was rated as nicer, more symmetric and clearer than the other types of grids, which is not the case in T. This result suggests that the choice of a particular grid depends on the type of display, and that it should match the display shape (for instance, radial for C). We found an effect of the hole in displays: in TH, the oriented grid was always rated better than the radial grid (no difference on T between radial and oriented).

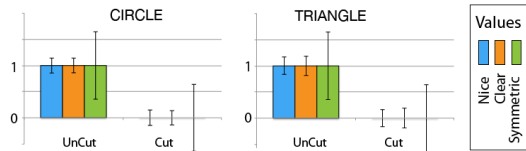
**Summary:**

Our results invalidate H4 (*better to keep regularity via a rectangular grid than different grid shape*). We found no differences between grid types on T but an effect for radial grid on C suggesting that grids need to match display shape.

**Results for Regularity (H5)**

In both C and T, with and without holes, not cutting the content (i.e. breaking the regularity of the elements to make

them fit in the display) was rated nicer and clearer than cutting it (Figure 11). We found no difference on symmetry except for CH, which interestingly was found more symmetric when the regularity was preserved (i.e. content is cut).



**Figure 11. Results for H5.**

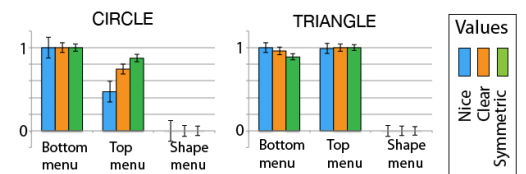
**Summary**

This result invalidates H5: it is better to break regularity, than cutting content on the edges.

**Results for Predictability (H6)**

**Menu position**

Overall, the Bottom menu and the Top menu were deemed nicer, clearer and more symmetric than the Shape menu (Figure 12).



**Figure 12. Results for H6.**

**Display shape**

For C, using the Bottom menu was deemed nicer and clearer than the Top menu. We found no difference on the symmetric measure. For T, we found no difference between Top and Bottom menus. Interestingly, when on TH, the Top menu was deemed nicer, clearer and more symmetric than the Bottom menu. This result seems to indicate that the choice of the position of the menu depends on the display's shape: bottom for C, top for TH.

It is also interesting to note that C is horizontally symmetric (as opposed to T), and thus bottom and top menus are equivalent from a geometrical point of view (on T, the bottom menu is larger and the top menu is narrower). Participants preferred the bottom menu, which goes against conventions of placing the menu at the top of the display.

**Summary:**

We cannot validate or invalidate H6 (*menus should be placed at the widest horizontal portion of the display*). We further explore this hypothesis with more shapes next.

**FOLLOW-UP PAIRWISE EXPERIMENT**

The goal of this follow-up study was to analyze how previous findings generalize to more diverse shapes. We picked the most interesting composition revealed in the first study and conducted an experiment with four new shapes.

**Shapes**

We systematically explored shapes with increasing number of edges: triangle (3), trapezoid (4), pentagon (5) and

hexagon (6). We also included an inversed triangle to see if the orientation of the shape had any effects.

**Design**

As before, we only compared visual compositions among shapes. We focused on the most interesting hypotheses (i.e. those related to the display shape) and removed the worst conditions (random grid in H4, shape menu in H6).

- *Sub-survey 1. Regularity (H4):* we studied three grid layouts (regular, radial and oriented) and 2 elements shapes (rectangular or as display), as seen in Figure 13. Overall, we had 15 pair comparisons for each shape x 4 shapes = 60 pairs.
- *Sub-survey 2. Predictability (H6):* We tested 2 positions for the menu (top and bottom). Overall, we had 1 pair comparisons for each shape x 4 shapes = 4 pairs.

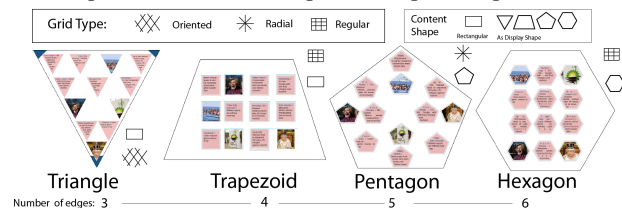


Figure 13. Shapes studied in our follow-up experiment.

**Participants**

In total 20 participants (5 female) from 7 different countries completed the four surveys. Participants were aged 29.8 on average (SD=5.1). 9 of them were university students, 6 worked in academia as engineers (1), professors (2) or researchers (3), and the rest had various professions.

**Collected data**

We collected 64 pairs × 3 questions = 192 answers per participant so a total of 192 x 20 users = 3840 answers.

**Results**

*Step 1: Individual consistency checking*

The mean TSR for all users was 0,93 (SD = .05). Individual consistency was at least over 0.8 for all users

*Step 2: Overall consistency checking*

Results show that the mean Kendall’s u-coefficient is above 50% for all conditions (H4 nicer 66.8%, clearer 61.6%, symmetric 68.2% and H7 nicer 74%, clearer 56.6%, symmetric 71.6%).

*Step 3: Model the data*

The individual and overall consistencies were confirmed, so we proceeded to model the data as in our previous survey.

*Results for Regularity:*

On the Hexagon, Pentagon and Trapezoid, the Radial Grid was rated best with some exceptions: the Regular Grid on the Trapezoid was rated as clear as the Radial Grid (Figure 14). The Oriented grid was rated worst on these three displays. This was particularly true when the Oriented grid was combined with rectangular content, which had the lower ratings. Overall, shaped content looked nicer than rectangular content, which confirms the results of the previous survey.

Results were different for the Triangle: the Regular Grid was rated as nicer and more symmetric than the Radial and Oriented Grids. When the Regular Grid was combined with rectangular content, it was also rated as clearer. Interestingly, these results on the Triangle are also different from the results of the previous survey, where we found no difference between Regular and Oriented Grids. This means that the orientation of a shape (i.e. triangle vs. inversed triangle) changes the user preference on the grid layout.

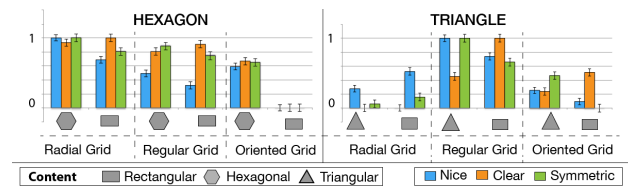


Figure 14. Results for Regularity (H4) on the Hexagon and Triangle displays.

*Results for Menu position:*

We only found a difference between Top and Bottom menus on the Hexagon and on the Triangle: on both shapes the Top menu was found nicer and clearer. This result on the inversed Triangle is different than the one from the previous survey, where no difference was found between Top and Bottom menus on the regular Triangle.

One hypothesis for this difference is that on the inversed Triangle, the Top area combines the fact of being the traditional location with a large area. Instead, on the regular Triangle, the Top was narrower, and thus participants were torn between the traditional menu position (top) and the larger area (bottom). This would also explain that no position preference emerged on the Pentagon and Trapezoid: on both shapes the top area is narrower.

*Summary:*

These results confirm the non-validity of H4 for other shapes except for the inversed Triangle (It is better to keep regularity via a rectangular grid than via different grid shape). Again, we cannot validate or invalidate H6 (“Menus should be placed at the widest horizontal portion of the display”), although results seem to indicate that the width of the menu along with the predictability of its position influence user’s preference.

**DISCUSSION**

**Guidelines for designing content for non-rect. displays**

Based on our findings we propose a set of guidelines for designers on the use of symmetry axis, content shape, grid layout, regularity and menu position for non-rectangular displays. Some of these design guidelines contradict current conventions on rectangular displays.

*Symmetry axis:* The symmetry axis should be vertical to ensure that the final design is nice, clear and symmetric.

*Content shape:* Instead of using the traditional rectangular boxes for text or images, designers can reshape the content to fit the display (circular on circle, triangular on triangles, etc.). This reshaping will have different effects depending

on the display shape: it will look nicer with circular content, or more symmetric with triangular content. However designers should be aware that sometimes reshaping content might make it appear less clear (such as in our triangle condition).

*Grid layout:* While using the traditional regular grid works well for certain shapes (regular and inversed triangles), using a grid with the same shape as the display shape can make the overall design look pleasing, clear and symmetric (as with radial grids in circle, pentagon and hexagon displays). A non-regular grid can benefit from non-rectangular content, as it better fits the shape of the grid (triangular content in oriented grid for instance).

*Breaking content:* To solve the problem of content not fitting exactly on the display, designers should favor breaking the regularity of the grid and making all content fit, rather than cutting elements by trimming the edges.

*Menu position:* While placing the menu at the traditional position on top of the interface works best for triangle and hexagon displays, designers could place it at the bottom in certain cases: this is a position that is nicer, clear and symmetric for a circular display, and that is equivalent to the top position for certain shapes (pentagon and trapezoid).

#### **Generalization to other shapes**

Since this work is the first exploration on how visual composition principles apply to non-rectangular displays, we decided to adopt a context-independent approach. We chose to study the generic properties of layout design instead of focusing on a given interface for a given application. The reason is that we wanted to provide generalizable findings rather than narrow in on specific guidelines that would be only valid for a specific case.

Our choice of shapes was based on usage scenarios envisioned for non-rectangular displays [31] and on an exploration of shapes with varying number of edges, which can also be found in [31]. Some of our results seem to be consistent across shapes, such as the fact that “shape-like” content looks better than a rectangular content, suggesting that they are probably valid for other shapes. Other results seem to depend on the display shape, such as the layout grid: while a radial grid is best for most display shapes (circle, hexagon, pentagon and trapezoid displays), a regular grid is better on triangular displays. In the future we plan to explore how these principles can be applied to specific usage contexts, which will imply a precise shape and display size, such as hand-held or in-vehicle displays. The size of the display can have an influence too, particularly when considering very small non-rectangular displays. We plan to study the size factor in the future.

#### **Data collection from graphic designers**

In our work we gathered non-rectangular designs created by 5 graphic designers, which informed our revision of traditional visual composition principles. Our designers had mostly experience in print and web design, and it could be interesting to explore how results would differ with other

expertise. Also the task was to fit the content (image, text, icons) from an existing webpage to a non-rectangular display of the same area. If we had asked designers to create a visual layout composed of images, text and menus, without giving them the actual content, maybe results could differ. In our case, we wanted to ensure that all designers used the same content. One designer for instance stated that he usually works in team. Future work will further investigate the influence of designer’s background and of task instructions.

#### **Limitations and future work**

Beyond the previously cited propositions concerning the shapes and the design task, we would like to conduct a larger study with more graphic designers. However having access to professional designers is not simple, and their experience may bias their designs on unconventional displays. Instead we plan to contact design students, whose visual creativity is still being developed.

Another limitation of our studies is that we did not test all possible combinations between conditions, given the huge number of parameters influencing visual composition. In the future we plan to further study the combinations between the factors that proved to have an effect on user preference, such as between layout grid and menu position.

Last, we really want to know what we could learn by using an eye-tracking setup on our experimental conditions. We want to explore two points in this regard: 1) visual saliency of non-rectangular displays, i.e. which regions of the shapes are most salient [3]; and 2) visual path used by the eyes when searching for a specific information on non-rectangular displays. Search paths follow well-known patterns on rectangular websites, such as Z or E patterns [7], but these may differ according to the display shape.

#### **CONCLUSION**

We studied how traditional content layouts can be adapted to fit different non-rectangular displays. We first ask graphic designers to fit the content of a newspaper webpage (text, images and icons) into different non-rectangular displays. We use their output to generalize and adapt existing composition principles for rectangular displays.

We evaluated the revised principles through two paired comparison questionnaires where participants compared pairs of layouts. The first survey explored Symmetry, Regularity and Predictability. The second survey extracted other interesting cases and applied them to additional shapes. Using the Bradley-Terry-Luce model to analyze our data we discuss the differences between the different compositions and propose a set of visual design guidelines for non-rectangular displays.

Although there are many other directions to study we believe that our work is a first step toward defining new guidelines for the design of free-form displays and it also has valuable application in the field of information visualization.

**ACKNOWLEDGMENTS**

We thank all graphic designers and study participants for their valuable time. A Leverhulme Trust Early Career Fellowship funded part of this work. We acknowledge the Canada Research Chairs award to the third author for supporting this work.

**REFERENCES**

1. Daniel Ashbrook, Kent Lyons, and Thad Starner. 2008. An investigation into round touchscreen wristwatch interaction. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services* (MobileHCI '08). ACM, New York, NY, USA, 311-314. DOI=<http://dx.doi.org/10.1145/1409240.1409276>
2. Gui Bonsiepe. A Method of Quantifying Order in Typographic Design. *Journal of Typographic Research* Close. Volume II, Number 3 (Jul 1, 1968): 203-220.
3. Georg Buscher, Edward Cutrell, and Meredith Ringel Morris. 2009. What do you see when you're surfing?: using eye tracking to predict salient regions of web pages. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '09). ACM, New York, NY, USA, 21-30. DOI=<http://dx.doi.org/10.1145/1518701.1518705>
4. Ralph Allan Bradley and Milton E. Terry. 1952. Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika*, 39(3/4), 324-345. DOI=<http://www.jstor.org/stable/2334029>
5. Kuan-Ta Chen, Chen-Chi Wu, Yu-Chun Chang, and Chin-Laung Lei. 2009. A crowdsorceable QoE evaluation framework for multimedia content. In *Proceedings of the 17th ACM international conference on Multimedia* (MM '09). ACM, New York, NY, USA, 491-500. DOI=<http://dx.doi.org/10.1145/1631272.1631339>
6. Daniel Cotting and Markus Gross. 2006. Interactive environment-aware display bubbles. In *Proceedings of the 19th annual ACM symposium on User interface software and technology* (UIST '06). ACM, New York, NY, USA, 245-254. DOI=<http://dx.doi.org/10.1145/1166253.1166291>
7. Edward Cutrell and Zhiwei Guan. 2007. What are you looking for?: an eye-tracking study of information usage in web search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07). ACM, New York, NY, USA, 407-416. DOI=<http://dx.doi.org/10.1145/1240624.1240690>
8. Antonella De Angeli, Alistair Sutcliffe, and Jan Hartmann. 2006. Interaction, usability and aesthetics: what influences users' preferences?. In *Proceedings of the 6th conference on Designing Interactive systems* (DIS '06). ACM, New York, NY, USA, 271-280. DOI=<http://dx.doi.org/10.1145/1142405.1142446>
9. Digital Signage Market Analysis, Grand View Research. Nov. 2014. Retrieved September 2016 from <http://www.grandviewresearch.com/industry-analysis/digital-signage-market>
10. Dondis, D. 1973. A primer of visual literacy. MIT Press.
11. Elam, Kimberly. 2001. *Geometry of Design: Studies in Proportion and Composition*. Princeton Architectural Press, 96 pages.
12. Elam, Kimberly. 2004. *Grid systems - principles of organizing type*. Princeton Architectural Press, 112 pages.
13. Karl Flieder and Felix Mödritscher. 2006. Foundations of a pattern language based on Gestalt principles. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '06). ACM, New York, NY, USA, 773-778. DOI=<http://dx.doi.org/10.1145/1125451.1125605>
14. Fraher, R. and Boyd-Brent, J. 2010. Gestalt theory, engagement and interaction. In CHI EA '10. ACM, 3211-3216.
15. Galitz, W. 2007. *The Essential Guide to User Interface Design: An Introduction to GUI Design Principles and Techniques*. John Wiley & Sons, Inc., NY, USA.
16. Graham, L. 2008. Gestalt Theory in Interactive Media Design. *Journal of Humanities and Social Sciences*, 2,1
17. Huntley, H.E. 1970. *The Divine Proportion: A Study in Mathematical Beauty*. Dover Publications. 208 pages.
18. Renate Haeuslschmid, Bastian Pfleging, and Florian Alt. 2016. A Design Space to Support the Development of Windshield Applications for the Car. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16). ACM, New York, NY, USA, 5076-5091. DOI: <http://dx.doi.org/10.1145/2858036.2858336>
19. IHS NewsRoom. 2014. Retrieved September 2016 from <http://press.ihs.com/press-release/automotive/automotive-display-systems-grow-186-billion-2021-driven-connectivity-infota>
20. Janin Koch and Antti Oulasvirta. 2016. Computational Layout Perception using Gestalt Laws. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA '16). ACM, New York, NY, USA, 1423-1429. DOI: <http://dx.doi.org/10.1145/2851581.2892537>
21. Masaaki Kurosu and Kaori Kashimura. 1995. Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability. In *Conference Companion on Human Factors in Computing Systems* (CHI '95), I. Katz, R. Mack, and L.

- Marks (Eds.). ACM, New York, NY, USA, 292-293. DOI=<http://dx.doi.org/10.1145/223355.223680>.
22. Lavie, T., and Tractinsky, N. 2004. Assessing dimensions of perceived visual aesthetics of web sites. *Int. J. Hum.-Comput. Stud.* 60, 3 (2004), 269-298.
  23. R. Duncan Luce. Individual Choice Behavior: A Theoretical Analysis. Wiley, New York, 1959.
  24. Aaron Marcus. 1991. *Graphic Design for Electronic Documents and User Interfaces*. ACM, New York, NY, USA.
  25. Meggs, P. 1998. A History of Graphic Design. John Wiley & Sons. 624 pages.
  26. Muller-Brockmann, J. 2011. Grid Systems in Graphic Design. Niggli Verlag. 176 pages.
  27. Ngo, D.C.L, Teo, L. and Byrne, J. Formalising guidelines for the design of screen layouts, *Display* 21, Elsevier Science Publishers (2000), 3-15.
  28. Ngo, DCL, Teo, L., and Byrne, J. 2003. Modelling interface aesthetics. *Inf. Sci.* 152, 1 (June 2003), 25-46.
  29. Don Norman. 2002. Emotion & design: attractive things work better. *interactions* 9, 4 (July 2002), 36-42. DOI=<http://dx.doi.org/10.1145/543434.543435>
  30. Carolyn Salimun, Helen C. Purchase, David R. Simmons, and Stephen Brewster. 2010. The effect of aesthetically pleasing composition on visual search performance. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries* (Nordichi '10). ACM, New York, NY, USA, 422-431. DOI=<http://dx.doi.org/10.1145/1868914.1868963>
  31. Marcos Serrano, Anne Roudaut, and Pourang Irani. 2016. Investigating Text Legibility on Non-Rectangular Displays. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16). ACM, New York, NY, USA, 498-508. DOI: <http://dx.doi.org/10.1145/2858036.2858057>
  32. Sonderegger, A., and Sauer, J. The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Applied Ergonomics* 41,3 (2010), 403-410.
  33. Thomas Stuart Tullis, An evaluation of alphanumeric, graphic, and colour information displays. *The Journal of the Human Factors and Ergonomics Society*. 1981. vol. 23, no. 5, 541-550. doi: 10.1177/001872088102300504
  34. Thomas Stuart Tullis, Screen design, in: M. Helander (Ed.), *Handbook of Human-Computer Interaction*, Elsevier Science Publishers, Amsterdam, The Netherlands, 1988, pp. 377-411.
  35. Van Schaik, P., and Ling, J. Modelling user experience with web sites: Usability, hedonic value, beauty and goodness. *Interacting with Computers* 20(2008), 419-432.
  36. Cheng Xu and Kent Lyons. 2015. Shimmering Smartwatches: Exploring the Smartwatch Design Space. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction* (TEI '15). ACM, New York, NY, USA, 69-76. DOI=<http://dx.doi.org/10.1145/2677199.2680599>
  37. Yeonsoo Yang and Scott R. Klemmer. 2009. Aesthetics matter: leveraging design heuristics to synthesize visually satisfying handheld interfaces. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '09). ACM, New York, NY, USA, 4183-4188. DOI=<http://doi.acm.org/10.1145/1520340.1520637>
  38. Jankowski, J., Samp, K., Irzynska, I., Jozwicz, M., and Decker, S. 2010. Integrating Text with Video and 3D Graphics: The Effects of Text Drawing Styles on Text Readability. In CHI '10. ACM, 1321-1330.
  39. Agresti, A. (2002). *Categorical Data Analysis*, 2nd edn, New York: Wiley.

---

<sup>i</sup> Sharp manufactures free form LCDs using Indium Gallium Zinc Oxide thin-film transistor in the backplane of flat-panel displays.

# DECO: a Design Space for Device Composition

Gary Perelman<sup>1,2</sup>, Marcos Serrano<sup>1</sup>, Mathieu Raynal<sup>1</sup>,  
Celia Picard<sup>2</sup>, Mustapha Derras<sup>2</sup>, Emmanuel Dubois<sup>1</sup>

<sup>1</sup>IRIT, University of Toulouse  
Toulouse, France  
{first\_name.last\_name}@irit.fr

<sup>2</sup>LRA, Berger-Levrault  
Toulouse, France  
{first\_name.last\_name}@berger-levrault.com

## ABSTRACT

Numerous interaction devices are designed through device composition. However, there is no conceptual support for this process and designers are left out to explore the space of combinations in an ad-hoc manner. In this paper we propose a design space for device composition, DECO, which focuses on the physical aspects of the composition. This design space is built around two main axes, namely Physical arrangement, which describes how elements are physically combined, and Physical manipulation, which describes how users manipulate each element. We first classify existing devices using our design space and then compare four of them to illustrate their similarities and differences. Using DECO, we design a new compound device: RPM2. This device is based on the combination of a regular mouse with the Roly-Poly Mouse. We describe in detail the user-centered iterative design process that leads to the final prototype. Finally we propose a set of design guidelines to assist the use of DECO for device composition.

## Author Keywords

Compound device; Design space; Device composition; Mouse.

## ACM Classification Keywords

H.5.2. Information Interfaces and Presentation: User Interfaces: Theory and methods

## INTRODUCTION

In recent years, numerous research projects have designed and implemented novel interaction devices resulting from the combination or aggregation of existing ones [12, 19, 22, 25, 29, 37, 41]. Among them, we can cite a touchscreen on a mouse as in LensMouse [41], an air balloon inside a mouse as in Inflatable Mouse [22], two keyboards fixed together as in FlipKeyboard [12], or two tablets glued to form a dual-sided device as in Codex [19]. These projects

followed an empirical design approach resulting in the development of ad-hoc solutions combining existing devices, without relying on a systematic or structured process. While these combinations proved to be useful in several contexts, designers who want to produce such type of devices are left out to explore multiple design alternatives on their own.

In this paper we investigate and refine the concept of **device composition** to promote the potential of combining existing devices to create new ones, and to leverage performing this combination in a more systematic manner. Device composition consists in **physically** putting together several existing devices to create a new one, hereafter referred to as a **compound device**. The concept of device composition can be related to the concepts of *multimodal interaction* and *composite device* [3, 16]. *Multimodal interaction* is based on the combination of several interaction modalities, each being defined as a couple of components: <device, language> [10]. In this paper we more specifically focus on the physical combination of two devices, instead of two modalities. The notion of *composite device*, as introduced by Bardram [3], defines the combination of interactive services and is thus far from the physical aspects considered in our approach.

In this work we develop the concept of device composition through the definition, illustration and evaluation of a design space for the physical combination of interaction devices. Our design space, **DECO**, is structured along two dimensions: 1) the physical arrangement, i.e. how the different devices are physically composed and 2) the physical manipulation, i.e. the way a user will manipulate each element of the compound device. Using DECO, we classify and compare existing devices to illustrate how our design space helps in describing and comparing different solutions.

In addition to the illustration of DECO on existing devices, we use this design space to elaborate a novel compound device supporting multi-dimensional interaction. This compound device aims at adding degrees of freedom to a traditional mouse. We describe in detail the user-centered iterative design process, using our design space, along four complete iterations involving user evaluations. These iterations prove that DECO contributes to the design of compound devices that are usable and perform well.

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

DIS 2016, June 04-08, 2016, Brisbane, QLD, Australia  
© 2016 ACM. ISBN 978-1-4503-4031-1/16/06...\$15.00  
DOI: <http://dx.doi.org/10.1145/2901790.2901893>



Our contributions are 1) the definition of the concept of device composition, 2) a design space, DECO, dedicated to device composition, 3) the classification and comparison of existing work using DECO, 4) the design and evaluation of a new compound device for multi-dimensional interaction and 5) a set of design guidelines for applying DECO on device composition.

#### STATE OF THE ART

Our work is inspired by the large number of existing compound devices and by related work on composition toolkits and models for HCI. We synthesize these three sets of related works in this section and underline how DECO provides a novel perspective in terms of device composition.

#### Compound Devices

Device composition has largely been used in HCI. Such devices found in the literature can be classified into three main categories according to their interaction sense (input or output): 1) composition of two *input* devices to augment input interaction; 2) composition of two *output* devices to augment the output space; and 3) composition of *input* and *output* devices to provide an additional feedback on an input device (or vice versa). In this section we cite some examples without being exhaustive.

In the first category (*input-input*), a large body of work has been dedicated to the fusion of a mouse with another input device, most commonly with a tactile surface [2, 40]. A commercial example of this composition is the Apple's Magic Mouse. Some gamer's mice are based on a combination of a keyboard and a mouse [29]. Dedicated devices, such as the Lexip 3D [25] are based on the combination of a regular mouse with a joystick to offer additional degrees of freedom to support 3D manipulation tasks.

In the second category (*output-output*), most works are based on combining displays, such as Siftables, Codex, DualScreen or HoverPad [19, 23, 26, 32]. For instance, Siftables is based on the spatial combination of tiny screens [26]. Codex is a dual display mobile device [19]. Dualscreen combines two screens on a computer to share personal content to other users [23]. HoverPad is a compound device based on a self-actuated display on top of a tactile table [32].

Concerning the last category (*input-output*), one general approach is to add a display to an input device. For instance Yang et al. proposed LensMouse, based on adding a screen on the mouse [41]. The opposite approach, i.e. adding input devices to a touchscreen, has been used in back-of-device prototypes [31, 36]. For instance, Sugimoto proposes to add a touchpad on the back of a PDA [36].

This general presentation of existing compound devices underlines their diversity and the multiplicity of composition alternatives. In the past, designers have explored this large composition space without any

systematic or structured approach. In our work we propose a design space to support this exploration.

#### Composition toolkits in HCI

The idea of composition can be found in several HCI toolkits, from hardware platforms such as Arduino or Phidgets [35] to software IDEs such as OpenInterface or PureData [3, 9, 33].

Hardware toolkits help assembling and combining different sensors like pressure sensors, push buttons, light detectors, etc., resulting in a device based on sensor fusion. Phidgets [35] for example, propose multiple sensors that can be easily plugged together to create a device. Arduino [34] also proposes a set of actuators and sensors that can be combined to produce interactive devices. Hardware composition toolkits support the composition at a low level, i.e. at sensor granularity, but do not provide any help in guiding the design process or selecting design alternatives.

Other approaches developed the concept of composition at a software level. The OpenInterface platform [33] allows assembling abstract components and linking their input and output to develop a multimodal interaction. ContextToolkit [9] is another example using software components that encapsulate several types of sensors. It is then possible to combine those components to develop an interaction technique. Most of these software toolkits rely on analytical models to help identifying and structuring software components. However, they do not help determining how to design the physical combination of devices.

Rather than developing a concrete toolkit, the goal of the present work is to propose a theoretical framework. These existing software and hardware composition toolkits could actually be used to implement our theoretical approach.

#### Analytical approaches

In HCI, several works proposed to classify and describe input devices and their characteristics on the basis of the sensors used. Among them, Card's design space [8] defines three composition operators that can be used to combine sensors: Merger, Layout and Connection composition. Those three operators refer to the data composition (Merger), the spatial composition (Layout) and to the ability to plug the output of a sensor into the input of another sensor (Connection). Inspired by these three simple composition operators, our work is intended to propose a richer composition design space.

At a higher level of abstraction, multimodal properties were design to better describe the composition of interaction modalities. Allen [1] temporal properties define the temporal composition between two interaction modalities over time. The CARE [10] properties defines relationships between multiple modalities defined by a couple *<device, language>*. The fusion of modalities is performed at the language level. In our paper, we propose to refine these concepts by focusing on the composition of devices, not languages.

Closer to our approach, Hartmann et al. studied the combination of sensors with real devices [16] and enlightened the importance of the physical aspect. According to the authors, two devices or sensors that are designed to explicitly support their composition and that are aware of each other lead to a better resulting device. This type of composition is called “dovetail joint”. Even if authors warn future designers against the risks of a bad composition, there is a lack of a design model that would better describe the properties of device composition.

To conclude, existing tools and properties describe the composition process at low level, using hardware or software toolkits, and at high level, using models to describe the composition of modalities. However we did not find any approaches focusing on the device composition process from a physical perspective. A design space to rationalize the composition process at a physical level would be a relevant complement to existing approaches, given the wide set of existing devices with different shapes, sizes or materials. To this end, we introduce DECO, a design space to inform the device composition process, using a set of relevant properties, in a similar way than other design methods such as annotated portfolio [7].

#### DEVICE COMPOSITION: DEFINITION

As seen in the previous section, existing work has paid little attention to identifying and analyzing the physical properties of device composition. Such consideration is crucial as it affects the interaction capabilities of the compound device. For instance, according to their relative position, two devices can be used simultaneously (e.g. a balloon and a mouse, as in Inflatable Mouse [22]) or only sequentially (a keyboard and a touchpad glued on its back, as in FlipKeyboard [12]). Altering the relative position of the components can also highly alter the resulting device: the LensMouse [41] would offer a completely different functionality if the touchscreen was placed on the side of the mouse instead of on its top.

Our concept of device composition takes this physical aspect into consideration. We define this concept as follows: a **device composition** is the physical assembly of several existing **elements**, according to a well-identified physical setting, and resulting in a **compound device**. The possible elements involved in a device composition include physical objects and devices.

This definition leads to two main design considerations:

- **Physical arrangement:** The integration of physical elements needs to consider their relative position and the alteration of their initial shape.
- **Physical manipulation:** Depending on the previous physical arrangement, different users’ actions and body parts can be used to interact with the compound device. In addition, the device elements may be used sequentially or in a more interleaved manner, potentially allowing compound gestures.

These two design considerations are at the core of our design space, DECO.

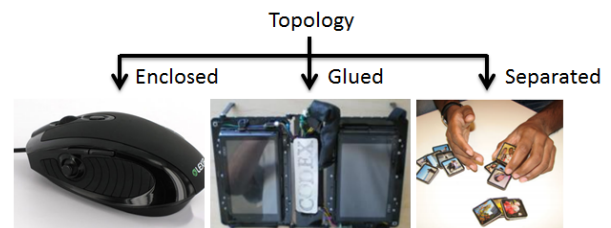
#### DECO: A DESIGN SPACE FOR DEVICE COMPOSITION

The goal of DECO is to help designing new compound devices through **device composition** as previously defined. To achieve this goal, DECO is built around two axes: the *Physical arrangement* of elements (axis 1) and the *Physical manipulation* of the compound device (axis 2). We retain these two axes as they describe the two most important aspects of any device: its shape (**A1**) and its usage (**A2**).

##### Axis 1: Physical arrangement

The Physical arrangement axis (**A1**) depicts how the different elements are physically and/or spatially composed to create a compound device. These physical or spatial aspects have a direct impact on the device’s final shape and on its physical manipulations, i.e. its usability and available DoFs. This axis describes the relative position of each element, how elements are physically linked together and how this arrangement varies over time. To address these considerations, we refine this axis into 3 properties: Topology, Fusion type and Dynamicity.

**Topology.** This property of DECO describes the spatial and physical organization of each element with respect to each other. We identified three possible categories for this axis: Enclosed, Glued and Separated (Figure 1).



**Figure 1: Illustration of the three Topology categories: Enclosed as in Lexip 3D [25], Glued as in Codex [19] and Separated as in Siftables [26].**

*Enclosed:* The composition topology is *Enclosed* when one element is confined into another one. For instance, GlobeMouse [13] results from the insertion of a SpaceMouse into a 3DoF trackball. The Inflatable Mouse results from the insertion of a balloon into a mouse [22].

*Glued:* The composition topology is ‘Glued’ when elements are physically aggregated. This is the most common form of physical arrangement in compound devices. For instance, the LensMouse [41] is a compound device resulting from the assembly of a mouse and a touchscreen (elements are glued together, see Figure 2 - center). Codex [19] is a device created by fastening two tactile screens together (Figure 1 - center).

*Separated:* The composition topology is *separated* when multiple elements are spatially composed, i.e. their spatial relationship is detected. The pair of device (mouse, keyboard) for instance can not be considered as a

compound device because they are not spatially composed. One example of *separated* compound device is Siftables [26]: these small displays are not *glued* together but can sense their relative position and orientation (Figure 1 - right). A particular instance of a *separated* topology is when the device is made of two separate elements not spatially related, but that can be *glued* if needed (see *dynamicity* later).

**Fusion type.** In the case of a *Glued* topology, a subsequent question is whether the physical shape of elements is altered to support the composition. Designers need to anticipate whether existing devices can be used or new ones must specifically be crafted. The Fusion type property describes which elements (if any) are altered during a composition. We identified three categories of *Fusion*:

(*A-B*) depicts the situation where elements are unchanged during composition: all elements are glued together without altering their original shape. For instance, a computer screen can simply be fixed to another one without further physical alteration as in the double-screen laptop Billboard [23] (Figure 2 - left). A similar approach is used in the FlipKeyboard [12], which is composed of a touchpad glued on the back of a keyboard.

(*A←B*) depicts the situation where the physical shape of element A has been modified (drilled, removed, reshaped, etc.) to accommodate or glue the element B. For instance in the case of the LensMouse [41] (Figure 2 - center), the mouse serves as a receptacle for a tactile screen: to place the touchscreen at an appropriate angle, the mouse cover has been altered and mouse buttons were also removed. In the Trackball Keyboard [37], a keyboard was carved to integrate a trackball on its top (Figure 4 - left).

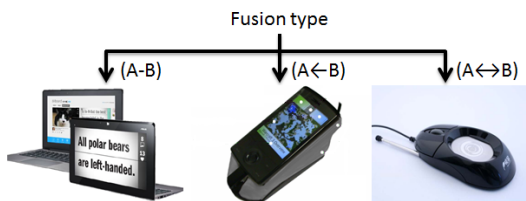


Figure 2: Illustration of the three *Fusion types*: (*A-B*) as in Billboard [23], (*A←B*) as in LensMouse [41] and (*A↔B*) as in Ink222 [20].

(*A↔B*) depicts the situation where the shape of both elements has been altered to support their composition. The Ink222 [20] mouse (Figure 2 - right) is a compound device based on a classical mouse and a touch tablet. In this case, the mouse was carved and the touch tablet was reshaped to fit a circular space. In the CapMouse [40] (Figure 6 - right), the mouse buttons have been removed and a touchpad has been adapted on top of the mouse.

**Dynamicity.** This property of DECO characterizes whether the Physical arrangement of a compound device changes over time. The physical arrangement is either *static*, i.e. defined once for all, or *dynamic*, i.e. it may change

depending on the task, the user or the context (Figure 3). Usually, this property applies to the topology values *glued* and *separated*, when a device can shift from one to the other. For instance, Rooke M. [30] proposes multiple spatial arrangements between hexagonal bezel-less screens. Using *separated* screens, a user can perform selections and rotations whereas, using *glued* screens, a user can perform swipes or tilts (Figure 3).

**Axis 2: Physical manipulation**

The Physical manipulation axis of DECO describes how users manipulate the compound device’s elements and aims to characterize its usage. From a physical perspective, focusing on the device usage means focusing on the device handling and on its expected manipulations. The Physical manipulation axis is thus described by 3 properties: Human effectors, Physical actions and Temporal usage.

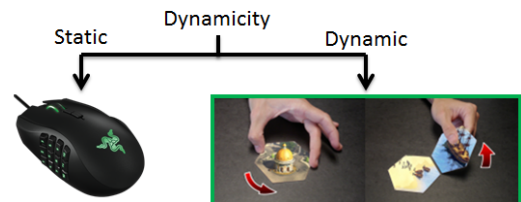


Figure 3: Illustration of the *Dynamicity* property: *static* as in the Razer Naga [29] and *dynamic* as the Hexagonal bezel-less screens [30].

**Human effectors.** This property of DECO identifies the body parts acting on each element of the compound device. Anatomical dependencies between human effectors may largely influence their use in a compound device [24]. A human effector may be for example one finger, one hand, one foot, or the combination of several of them. This property is described by the body part(s) involved in the interaction with each element. For example, the Trackball Keyboard [37] is a compound device designed for a bimanual use, with one hand on each element: it can be described by the couple <keyboard: non dom. hand; trackball: dom. hand> (Figure 4 - left). The Inflatable Mouse [22] is designed to be used with the dominant hand: it can be described by the couple <mouse: dom. hand; balloon: dom. hand> (Figure 4 - right).

This axis allows identifying whether several elements of the compound device involve the same body part, as with the Inflatable Mouse. In this case, the physical actions applied to each element need to be compatible, which may imply to adapt or modify one of the elements.

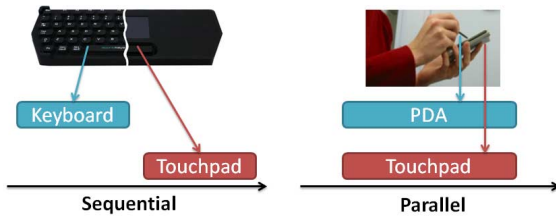


Figure 4: Illustration of the *Human effectors* property. Trackball keyboard [37]: <keyboard: non dom. hand; trackball: dom. hand> (left). The Inflatable Mouse [22]: <mouse: dom. hand; balloon: dom. hand> (right).

**Physical actions.** This property of DECO describes the expected user’s physical actions on each element. Examples of actions include translations, rotations, swipe, etc. Existing notation such as UAN [17] may also be used to describe more complex actions. The role of this property is therefore to help anticipate the ability of the user to produce the expected physical actions on each element, anticipate conflicts and consider different design alternatives.

This property is further refined by the **degrees of freedoms (DoF)** involved by each physical action. The CubicMouse [14], for example, is a compound device combining a cube and three rods (Figure 6). The cube has a 6DoF tracking sensor that allows logging the device’s position and orientation. It can be described as follow: *<Cube: translations (3DoFs), rotations (3DoFs); StickX: translation (1DoF); StickY: translation (1DoF); StickZ: translation (1DoF)>*.

**Temporal usage.** This property of DECO describes how the physical actions are expected to be performed over time. The temporal usage of an interactive system has already been described in the literature. Allen properties [1] describe the temporal composition of modalities. Simplified later with two operators in the CASE model [27]: *Parallel* and *Sequential*.



**Figure 5: Illustration of the temporal usage property: sequential as in FlipKeyboard (left) and parallel as in HybridTouch (right).**

In our design space we use these operators to define the temporal usage of elements in a compound device. Obviously, this property is influenced by the Physical arrangement and the Physical actions, which can prevent [20] or encourage [40] the concurrent usage of elements. For instance, the FlipKeyboard’s [12] and HybridTouch [36] have the same Physical arrangement (*glued* on the back). However, in the FlipKeyboard, only one of its elements can be used at a time: the temporal usage is *sequential*. HybridTouch [36], based on the composition of a PDA and a touchpad, facilitates the concurrent usage of its two elements with a bi-manual interaction: the temporal usage is *parallel* (Figure 5).

**CLASSIFICATION OF COMPOUND DEVICES**

In this section we illustrate the use of DECO to describe a set of existing compound devices. We first give an overview of how these devices fit our space. Then we describe and compare in detail four existing devices to illustrate how our design space unveils their differences.

Temporal usage	Topology Fusion type	Physical arrangement				
		Enclosed	Glued			Separated
			A-B	A<-B	A<->B	
Sequential	GlobeMouse	Hexagonal bezel-less ★	Trackball Keyboard	Ink222	Hexagonal bezel-less ★	
		FlipKeyboard	LensMouse			
Parallel	Lexip 3D	Codex	PadMouse		Siftable	
	Inflatable Mouse	Razer naga	CubicMouse	Mouse 2.0	HoverPad	
		Dualscreen	RPM2 ★		RPM2 ★	
		HybridTouch				

★ = Dynamic compound device

**Table 1. Compound devices from the literature in DECO.**

**Physical and temporal classification**

Among the properties identified for the two axes of DECO, we adopted a graphical representation based on the physical topology, the fusion type, the dynamicity and the temporal usage. The two remaining properties (human effectors and physical actions) correspond to n-uplets that are less appropriate for a graphical representation. In the table above, compound devices from the literature are positioned with regard to these selected properties of DECO. A star is used to highlight *dynamic* compound devices (Table 1).

This graphical representation highlights the diversity of existing compound devices. It also reveals that the design space is relevant to classify existing work as each possible combination of properties (cell) is illustrated with an existing compound device.

**Comparison of compound devices with DECO**

To illustrate DECO more precisely, we selected four compound devices from the literature: the CubicMouse [14], the GlobeMouse [13], the PadMouse [2] and the CapMouse [40] (Figure 6). While these devices seem similar in certain aspects, DECO allows comparing them and exposing their similarities and differences (Table 2).



**Figure 6: From left to right, the Cubic Mouse [14], the GlobeMouse [13], the PadMouse [2] and the CapMouse [40].**

**GlobeMouse.** This compound device is based on the composition of a SpaceMouse and a trackball. The SpaceMouse is *enclosed* in the trackball shell. The physical arrangement of this compound device is *static*. Only one hand of the user is required to manipulate the compound device: *<SpaceMouse: dom. hand; trackball: dom. hand >*. The expected physical actions can be formalized as follow: *<SpaceMouse: translation (3DoFs); Trackball: rotation (3DoFs)>*. These actions can only be performed *sequentially*.

**CubicMouse.** This compound device is based on four main elements: a physical cube localized in 6D (position and orientation in a 3D space) and three rods passing through the cube center. The cube and the rods are *glued* together. The cube has been altered to allow the positioning of the



rods (Fusion type:  $A \leftarrow B$ ). The physical arrangement of the CubicMouse is *static*. The two hands of the user are required to manipulate the compound device: one for holding the cube and the other to adjust the rods  $\langle \text{cube: non dom. hand; rods: dom. hand} \rangle$ . As discussed in the previous section, physical actions can be described as follow:  $\langle \text{Cube: translations (3DoFs), rotations (3DoFs); StickX: translation (1DoF); StickY: translation (1DoF); StickZ: translation (1DoF)} \rangle$ . Physical actions on the rods and on the cube can be performed in *parallel*.

We use DECO to compare CubicMouse and GlobeMouse. The Physical arrangement of those devices differs, as their respective topology is *Enclosed* and *Glued*. The Physical manipulation differs too: the CubicMouse is a bimanual device, which offers a total of 9 DoFs, whereas the GlobeMouse can be used with one hand and offers 6 DoFs. Moreover, the CubicMouse provides a *parallel* usage and the GlobeMouse a *sequential* one. Our design space accurately highlights the differences between these two compound devices that differ both in terms of *physical arrangement* and *physical manipulation*.

**PadMouse.** This compound device is based on the composition of two elements: a regular mouse and a touchpad. The two elements are *glued* together. Only the mouse has been modified to produce the PadMouse: the mouse buttons have been removed to allow the composition with the touchpad (Fusion type:  $A \leftarrow B$ ). A digital button on the touchpad replaces the mouse buttons. The physical arrangement of PadMouse is *static*. To be physically manipulated the compound device requires only one hand. It can be described as  $\langle \text{mouse: non dom. hand; touchpad: non dom. hand} \rangle$ . Physical actions include 2D translations of the compound device and 2D touch on the touchpad, represented as follow:  $\langle \text{Mouse: translation (2DoFs); Touchpad: touch input (2DoFs)} \rangle$ . Finally, it is possible to use the elements in *parallel*.

In light of DECO, the CubicMouse and the PadMouse are similar in terms of Physical arrangement (*glued*, fusion type  $A \leftarrow B$ , and *static*), but they strongly differ in terms of Physical manipulation. First, regarding the human effectors, the CubicMouse requires two hands whereas the PadMouse requires only one. In addition, the device composition has different effects in terms of DoF available: the buttons of the PadMouse were removed and their DoF replaced by the touchpad; in the CubicMouse, the DoF of each element were preserved. The comparison between these two devices highlights the ability of DECO to differentiate devices from the point of view of their physical manipulation, even if the physical arrangement is similar. This underlines the relevance of considering the physical manipulation as an axis of our design space.

**CapMouse.** This compound device is based on the composition of two elements: a regular mouse and a capacitive surface. These two elements are *glued* together but in this case, both elements have been modified to create

the CapMouse (Fusion Type  $A \leftrightarrow B$ ): the mouse buttons have been altered and the capacitive surface has been curved to fit the mouse shape. The physical arrangement of CapMouse is *static*. Under the capacitive surface, a single mouse button allows for mouse clicks. The *human effectors*, *physical actions* and *temporal usage* of this compound device are the same as in the PadMouse.

Although the CapMouse and PadMouse Physical manipulation are the same, their Physical arrangement highlights a fundamental difference. The CapMouse required physical modifications of both elements (fusion type  $A \leftrightarrow B$ ), whereas the PadMouse only of one of its elements (fusion type  $A \leftarrow B$ ). As a consequence, designers could easily prototype new versions of the PadMouse by replacing the touchpad with another rectangular tactile device, such as a smartphone or a smartwatch for example. In other words, in the  $A \leftarrow B$  fusion, designers can simply consider device variations based on the replacement of B, while in the  $A \leftrightarrow B$  fusion it would be more difficult as each element has been physically altered.

**Summary:** These four compound device descriptions and the most relevant comparisons using DECO enlighten the capabilities of our design space to accurately describe and distinguish existing solutions in terms of Physical arrangement and/or Physical manipulation (Table 2).

			GlobeMouse	CubicMouse	PadMouse	CapMouse
Physical arrangement	Topology	Enclosed	x			
		Glued		x	x	x
		Separated				
	Fusion type	A-B		x		
		A $\leftarrow$ B			x	
		A $\leftrightarrow$ B				x
Dynamicity	Static	x	x	x	x	
	Dynamic					
Physical manipulation	Human effectors	Dom. Hand	x	x	x	x
		Non Dom. Hand		x		
	Physical actions	Rotation	3 DoFs	3 DoFs		
		Translation	3 DoFs	6 DoFs	2 DoFs	2 DoFs
		Touch input			2 DoFs	2 DoFs
	Temporal usage	Sequential	x			
		Parallel		x	x	x

Table 2: Description of GlobeMouse, CubicMouse, PadMouse and CapMouse using DECO.

**APPLICATION: RPM2**

To assess the utility of DECO to generate new solutions, we report about the design and evaluation of a new compound device, RPM2, based on the combination of a regular mouse with a Roly-Poly Mouse (RPM) [28]. The RPM is a device, inspired by the roly-poly toy, with a circular basis that affords up to 5 DoF (2D translation and 3D roll and rotation).

**Initial motivation**

In this section, we detail the two main motivations that drove our design of a novel compound device combining the RPM and a mouse, namely the need to 1) overcome a technical challenge and 2) facilitate the adoption of RPM

by making it compatible with the most widespread input device, the mouse.

The initial RPM prototype relies on an expensive infrared tracking system to log the position and orientation of the device. An embedded solution using a small 6 DoF sensor was proposed but lacks the precision of a laser mouse to track its 2D position. One solution to overcome this technical limitation is to combine the RPM with a mouse to benefit from its optical sensor. This solution has the additional advantage of drastically decreasing the cost of the resulting device.

To facilitate the adoption of RPM, we decided to envisage its usage as an extension of a classical mouse: RPM could then be used in conjunction with a mouse but also independently. From a DECO perspective, we wanted to propose a *dynamic* Physical arrangement: a device in which elements can be *glued* or *separated* depending on the usage context. For instance, when *glued*, the user can manipulate both elements with the dominant hand, while using the keyboard with his other hand. When the keyboard is no longer required, both devices could be *separated* and used in a bimanual mode, benefiting from higher precision with RPM.

Therefore, we decided to use DECO to guide our design process of a novel compound device combining the mouse and RPM. In the next section we illustrate the different versions of the new RPM (simply called RPM2) through an iterative user-centered design process using DECO. Participants involved in informal tests, pre-studies or evaluations (10 males and 2 females) were aged 24.8 on average (SD=3.6). All of them were right handed.

#### First iteration

We started by designing the *glued* version of the compound device: this version could potentially lead to physically altering one of the two devices (RPM or mouse) and therefore needed to be considered before exploring the *Separated* version. Our first step was to find the fusion type. We wanted to prototype a simple version of the device and improve it using DECO's second axis, the Physical manipulation. We built a cardboard prototype by using the fusion type  $A-B$  (no device is physically altered), since it is the simplest form of fusion.

The gluing mechanism relies on the use of an intermediate cardboard support. We made several prototypes by composing a USB mouse with a cardboard support on top of which RPM was placed, not altering any of the elements in the fusion process. We then varied the position of the mouse relative to the cardboard to test different physical compositions. This process led us to 5 different versions of the compound device according to the mouse position: on the left, on the right, on the front, on the back or under RPM (Figure 7). RPM was placed on the cardboard container but not attached to it to ensure free rotation of the device.



**Figure 7: Cardboard prototypes of RPM2: the mouse is on the right side (left), on the back (center) or under (right) RPM.**

We conducted a pre-study to evaluate the impact of the spatial *topology* of elements on usability and on the physical manipulations. We asked four users to manipulate all cardboard versions of RPM2. Users had to manipulate the compound device by performing translations, rotations, rolls and compound gestures (translation + roll). At the end of the experiment, users evaluated the device usability through an informal interview.

We found that the distance between the palm of the hand and the mouse, resulting from the cardboard support, had an impact on the usability of the device. This is due to the fact that, when RPM2 is translated, the cardboard slowly derives around RPM and the mouse is no longer aligned with the hand movement.

In addition, users declared that compound gestures were complex to perform on RPM2. This was also due to the distance between the two elements. As we wanted to keep the original RPM interaction properties, this was a main issue.

#### Second iteration: changing the fusion type

To solve the problems of the first prototype, we needed to allow the user to hold both the mouse and the RPM at the same time, i.e. reduce the physical distance between them. Using DECO, we searched for another fusion type that would get the two elements closer without altering the physical actions. In our context, a physical modification on RPM was impossible as it would impact the 3D rotation DoFs of the device. The fusion type  $(A \leftrightarrow B)$  was thus not appropriate. Therefore we decided to design a  $(A \leftarrow B)$  fusion by considering how the mouse could be physically altered to incorporate RPM. One additional constraint was that both elements had to be *glued* without permanently fixing them to preserve the *dynamic* property of the device.

#### RPM2: Prototype v2

The second version of the compound device was overall smaller than the previous one as we removed the cardboard container. The cover of the mouse was replaced by a 3D printed version (Figure 8 - left). This cover was carved to hold RPM on it. A side effect was that we had to relocate the buttons on the side of the mouse. In parallel, we wanted to determine the optimal size of RPM as size could have an important impact on the physical manipulation of the device and on the user's comfort. To answer those design questions, we conducted a series of user tests.

In the first test, we varied the size of RPM to measure its impact on the *Physical actions*. Twelve users had to



perform translations, rotations, rolls and compound gestures with the device. At the end of the experiment, participants were asked to evaluate the RPM size compared to the mouse size and they had to evaluate the device usability. The radiuses of the different RPM versions were 4, 5, 6, 8 and 10cm.

According to results, the optimal size for RPM was 5cm. With this size, both elements had almost the same width and allowed for a more comfortable hand posture. Moreover, users preferred to place the buttons on the left side of the mouse rather than on the right (Figure 8 - left) because pressing the buttons with the thumb (on the left) is less tiring than with the ring or little fingers (on the right). This confirms the importance of properly choosing the *human effectors* for each element.

To assess the device efficiency in a concrete task, we conducted a second test where we asked users to perform a 2D pointing task following the Fitts law [35] experiment protocol. Four users had to select among 25 circularly arranged targets using our device and a regular mouse. We evaluated six index of difficulty (defined as a ratio of the target size on the support circle size): 3.17, 3.70, 4.08, 5.93 and 6.33. Results showed that pointing with the mouse was faster than with RPM2. Pointing errors, i.e. missing the target, were significantly higher with our prototype. Participants' feedback revealed that the printed 3D support was hard to grasp: "The support is not steady, it's complicated to hold it correctly while I click." or "The support shape is not well suited. It was uncomfortable to grasp."

**RPM2: Prototype v3**

Based on these results, we built a third version of the compound device. We removed the 3D printed support and directly carved the cover of a mouse to hold the 5cm RPM. With the best size for RPM defined, we explored the buttons location through another user test. The task was to perform translations, rotations, rolls and compound gestures with RPM2. Users had to manipulate two versions of the compound device: one on which the mouse was carved on the center and the buttons were placed on the left side (prototype v3a, Figure 8 – center); and one on which the mouse was carved on the right, leaving the original left mouse-button available (prototype v3b, Figure 8 –right). These two solutions required the use of different human effectors to press the button (thumb or index) but also to manipulate (rotate or roll) RPM.

The first result of this test is that users preferred the version with RPM on the center (prototype v3a). This is due to the fact that, when RPM is displaced on the right, the hand posture is particularly uncomfortable when performing rotations and rolls: the index finger lays on the left click while all the other fingers lay on RPM. Not only RPM manipulation is more complicated but also the posture is rather tiring. In addition, users found that compound gestures were easy to perform on the version carved on the

center. In this version, the physical manipulation of RPM is performed using the three middle fingers.



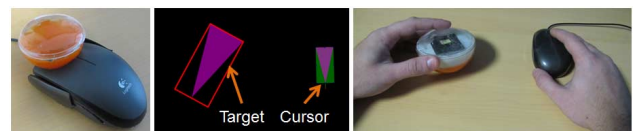
**Figure 8: Different versions of the (A←B) compound device, from left to right: based on a 3D printed support (prototype v2), carved on the center of the mouse (prototype v3a) and carved on the right side of the mouse (prototype v3b)**

Based on these results, we re-evaluated the compound device with four users in a 2D pointing task following the same protocol than before. Results showed a significant improvement in the compound device performance (10% better than the previous version). However, the prototype v3 still had a lower performance and a higher error rate than the regular mouse. Participants' feedback showed that this performance loss was due to a hard resistance of buttons, difficult and tiring to push with the thumb during a long period of time. Consequently, we improved the prototype by integrating softer buttons (Figure 9 – left).

**Third iteration: Designing parallel temporal usage**

After focusing on the Physical arrangement of the compound device, we focused on the temporal usage of its Physical manipulations. Our goal was to support compound gestures, i.e. *parallel* temporal usage of the device elements. While we included compound gestures in our previous iterations, we did not specifically focus on these gestures.

To evaluate the parallel usage of RPM2, we designed an experiment based on a multi-dimensional graphical task: rotate, scale and translate (RST, Figure 9 - center) [38]. Four users had to manipulate virtual rectangles in a 2D environment and dock them on a target rectangle. To move the rectangle, users first pressed the left click and then translated the device. To rotate and scale the rectangle, users could respectively rotate (left-right rotation) or roll (front-back rotations) the RPM. All three RST tasks could be performed in parallel with the RPM2 prototype v3a. At the end of the experiment users were asked to evaluate the device usability. We compared our device against the regular mouse, used as in PowerPoint (e.g. with anchors around the rectangle).



**Figure 9: The glued RPM2 prototype v3 (left), a random trial from the RST experiment (center) and the separated RPM2 compound device (right).**

The results reveal that our device is only 9% slower than the regular mouse, which means that our device is promising but the mapping between gestures and tasks needs to be improved. We observed that most users performed rotate, scale and translation in *parallel*. Actually, users found that RPM2 was “a good device because the manipulations are intuitive” and that “RPM2 is faster because I can scale and translate the rectangle at the same time”.

These results clearly show that we succeeded in designing a novel compound device using our design space, DECO, as a designing tool. The prototype v3a is *glued* and the experiment results established the device usability and performance, close to a regular mouse. Our main goal being to propose a *dynamic* device, we carried a last iteration to validate its usage in a *separated* topology.

**Fourth iteration: From a *Glued* to a *Separated* topology**

In our compound device, RPM and the mouse are not physically fixed. Therefore to come up with the *separated* topology, we simply take each element with a different hand and uncouple them (Figure 9 - right). The mouse cover still has a hole that impedes using the device normally: a solution can be to use a mechanism to close the hole when RPM is removed. In this iteration, we did not build this solution and used a regular mouse without hole as we focused on the performance of the bimanual interaction.

The bimanual interaction technique is based on the use of a regular mouse with the dominant hand and the RPM with the other hand (*human effectors: <mouse: dom. hand; RPM: non dom. hand>*). This compound device still offers 5 DoFs and supports the following physical actions: *<mouse: translation (2DoFs), RPM: rotation (3DoFs)>*. The two elements can be used in *parallel* (temporal usage).

We evaluated the performance of this new version of RPM2 (RPM2-Separated) through a rotate, scale and translate (RST) experiment. We compared the bimanual technique with a regular mouse under the same conditions as in the previous iteration. Twelve subjects performed the experiment. Results revealed a significant difference between the two interaction techniques: the compound device performed faster than the mouse (22% faster). These results show that the *Separated* version of RPM2 performs better than its *glued* counterpart and illustrates the interest of developing a *dynamic* compound device: according to the task to perform, the user will be able to easily adopt the *glued* or the *separated* version of the compound device.

**Summary**

We proposed a novel *dynamic* compound device, the RPM2. It can be used as a *glued* compound device or as a bimanual *separated* compound device. A series of tests show that RPM2 is an efficient device in both Physical arrangements. To design these two versions of RPM2, we used DECO in a user-centered iterative design process: we guided our design process with the two axes of DECO and

used its properties to iteratively propose several prototypes (Table 3).

		Physical arrangement			
		Enclosed	Glued		Separated
Fusion type		A-B	A<-B	A<->B	
Temporal usage	Sequential		Iteration 1 → Iteration 2		
	Parallel		Iteration 3 ★		Iteration 4 ★

★ = *Dynamic compound device*

Table 3. RPM2 design iterations in DECO.

To sum up, this illustrative example shows that DECO is a useful tool for designing novel compound devices.

**DISCUSSION**

**DECO to describe, compare and generate solutions**

In this paper we presented DECO, a design space for device composition. As illustrated on the RPM2 case study, using DECO, designers can explore the implications of a given design by changing some parameters among DECO's axes. DECO's axes also constitute leverages for suggesting new design possibilities, based on properties observed in different situations of device composition. As such, DECO contributes to research through design [15].

To validate our design space, we used the approach proposed by Beaudoin-Lafon [4] to evaluate design models. This approach is based on three properties characterizing the ability of a model to describe existing solutions (descriptive power), compare existing solutions (evaluative power), and generate novel solutions (generative power). The relevance of such a validation has already been established in previous work seeking to evaluate interaction design approaches [11, 21]. Concretely, to validate the descriptive power of DECO, i.e. its ability to describe a wide range of different solutions, we described and classified existing compound devices using the *Physical arrangement* axis and the *Temporal usage* property of our design space. To validate the evaluative power of DECO, i.e. its ability to differentiate multiple solutions, we compared four of these devices and illustrated how DECO helps identifying their similarities and differences in terms of physical arrangement and manipulation. Finally, to validate the generative power of DECO, i.e. its ability to create new design solutions, we generated a novel compound device based on the combination of the regular mouse with the Roly-Poly mouse [28]: we used DECO to explore several design alternatives in a user-centered iterative design process.

Through this 3-part exploration, we demonstrated that DECO is a useful design space that can be used to describe, compare and generate novel compound devices.

### Guidelines for generating compound devices

Generating and designing new devices often rely on an empirical approach that combines creativity and practical knowledge, as described in [39]. However, the lack of inspiration may affect the ideation phase of the design process. In the particular context of device composition, DECO is a structured analytical approach that seeks to offer designers a leverage to stimulate idea generation and to produce design alternatives. Combining such a formal approach with more informal design resources such as a brainstorming is possible and has already been proposed in HCI [6].

To help designers take advantage of DECO in a design process, we propose a set of lessons learnt during our experience in designing RPM2 with DECO.

*Physical arrangement:* One of the first steps, when designing an interactive device, is to analyze the users' requirements, tasks and context of use. These aspects may particularly impact the possibilities in terms of physical arrangement of the elements. At this design stage, using DECO should lead to choosing a specific topology. The two possible choices are an *Enclosed/Glued* topology, which leads to a single physical object, or a *Separated* topology, which leads to several objects.

In the case of a *Glued* topology, the advantage of starting with the *(A-B) fusion* is that it does not induce any modification on the initial devices. This allows evaluating early prototypes and getting feedback before considering physical alterations on any of the elements involved in the device composition.

*Physical manipulation:* The design of the Physical manipulation attributes needs to be reconsidered after each design iteration on the Physical arrangement. Even small physical changes in the prototype can induce large differences in terms of manipulation: for example, in our experience with RPM2, changing the size of the RPM considerably impacted its usability.

Finally, when the goal is to produce a *parallel* usage of the device, a good approach is to start designing and evaluating a *sequential* usage in the first iterations: the idea is to validate that each element can be properly manipulated and identify their potential in terms of manipulation (maximum range, stability, accuracy, etc.), before considering the combined manipulation.

### Limitations and future work

In this paper we presented a first design space for device composition. We identified some limitations in terms of generalizability, granularity and property definition that we detail below.

In our work we mainly considered compound devices made of only two elements. While the related work on devices combining more than two elements is limited, we still need to generalize the use of our design space and assess its

ability at describing, evaluating or helping generate compound devices based on more than 2 elements.

The definition of DECO has been largely driven by a critical analysis of compound devices presented in the literature. These solutions clearly focus on the combination of existing devices such as a mouse and a touchpad, i.e. the combination of two sets of sensors and physical objects. While we adopted the same approach to define the concept of device composition, DECO might also constitute a good support to analyze combinations at a finer grain of detail, i.e. at a sensor level. Future work will look into extending our design space to consider fusion composition.

In the future, we also plan to further analyze and refine the characteristics of the *Separated* topology. This case is particularly interesting for multi-surface systems, i.e. composing different interactive surfaces. We can for instance explore the geometric and spatial relationships between surfaces. Describing the Physical manipulations in this case will also require an extension of DECO to describe interactions such as mid-air or around-device gestures [5], which also have spatial properties.

### CONCLUSION

In this paper we introduced DECO, a design space for device composition focusing on its physical dimensions. DECO is structured along two axes: Physical arrangement and Physical manipulation. The Physical arrangement axis describes the topology of the composition (Enclosed, Glued or Separated), the fusion type (which device is physically modified to produce the composition) and the dynamicity of the physical arrangement. The Physical manipulation axis describes the human effectors, i.e. the body parts, used to manipulate the device, the physical actions and the temporal usage of the elements. We used DECO to classify existing compound devices and compare alternative existing solutions. To illustrate the generative power of DECO, we designed a novel compound device by combining a mouse with the Roly-Poly Mouse in a user-centered design process. Finally, we proposed a set of guidelines to help designers apply our design space to create novel compound devices.

### ACKNOWLEDGMENTS

This work was carried out as part of the research collaboration agreements IRIT-Berger Levraut (CNRS No.099856) and the ANR project AP2. The authors thank the members of the ELIPSE group for their useful suggestions.

### REFERENCES

1. James F. Allen. Maintaining knowledge about temporal intervals. *Commun. ACM*, 26(11):832–843, 1983.
2. Ravin Balakrishnan and Pranay Patel. 1998. The PadMouse: facilitating selection and spatial positioning for the non-dominant hand. In *Procs of CHI '98*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 9-16.

3. Jakob E. Bardram, Christina Fuglsang, and Simon C. Pedersen. 2010. CompUTE: a runtime infrastructure for device composition. In *Procs of AVI '10*. ACM, New York, NY, USA, 111-118.
4. Michel Beaudouin-Lafon. 2004. Designing interaction, not interfaces. In *Procs of AVI '04*. ACM, New York, NY, USA, 15-22.
5. Louis-Pierre Bergé, Emmanuel Dubois, and Mathieu Raynal. 2015. Design and Evaluation of an "Around the SmartPhone" Technique for 3D Manipulations on Distant Display. In *Procs of SUI '15*. ACM, New York, NY, USA, 69-78.
6. Christophe Bortolaso, Emmanuel Dubois. 2013. Model Assisted Creativity Sessions for the Design of Mixed Interactive Systems: a Protocol Analysis. In *INTERACT'13*, Springer, LNCS 8119, p. 126-143, 2013.
7. John Bowers. 2012. The logic of annotated portfolios: communicating the value of 'research through design'. In *Procs DIS '12*. ACM, New York, NY, USA, 68-77.
8. Stuart K. Card, Jock D. Mackinlay, and George G. Robertson. A morphological analysis of the design space of input devices. *ACM Transactions on Information Systems*, 9(2):99-122, April 1991.
9. ContextToolkit <http://contexttoolkit.sourceforge.net/>
10. Joëlle Coutaz and Laurence Nigay. Les propriétés CARE dans les interfaces multimodales. In *Actes de la conférence IHM'94*, Lille, pages 7-14, 1994.
11. Coutrix, C., Nigay, L.: Mixed reality: a model of mixed interaction. In: *Proc. of the Working Conf. on Advanced visual interfaces, AVI 2006*, pp. 43-50. ACM Press (2006).
12. The FlipKeyboard. <http://research.microsoft.com/en-us/um/people/bibuxton/buxtoncollection/detail.aspx?id=76>
13. Froehlich, B., Hochstrate, J., Skuk, V., and Huckauf, A. The globefish and the globemouse: two new six degree of freedom input devices for graphics applications. *CHI'06*. ACM, 191-199.
14. Bernd Fröhlich and John Plate. 2000. The cubic mouse: a new device for three-dimensional input. In *Procs of CHI '00*. ACM, New York, NY, USA, 526-531.
15. William Gaver. 2012. What should we expect from research through design? In *Procs CHI '12*. ACM, New York, NY, USA, 937-946.
16. B. Hartmann, S. Doorley and S.R. Klemmer, Hacking, Mashing, Gluing: Understanding Opportunistic Design, *IEEE Pervasive Computing*, vol. 7, no. 3, pp.46-54, 2008
17. H. Rex Hartson, Antonio C. Siochi, and D. Hix. 1990. The UAN: a user-oriented representation for direct manipulation interface designs. *ACM Trans. Inf. Syst.* 8, 3 (July 1990), 181-203.
18. Michael Haupt, Michael Perscheid, Robert Hirschfeld, Lysann Kessler, Thomas Klingbeil, Stephanie Platz, Frank Schlegel, and Philipp Tessenow. 2010. PhidgetLab: crossing the border from virtual to real-world objects. In *Procs of ITICSE '10*. ACM, New York, NY, USA, 73-77.
19. Ken Hinckley, Morgan Dixon, Raman Sarin, Francois Guimbretiere, and Ravin Balakrishnan. 2009. Codex: a dual screen tablet computer. In *Procs of CHI '09*. ACM, New York, NY, USA, 1933-1942.
20. Ink222 T&Mouse <http://research.microsoft.com/en-us/um/people/bibuxton/buxtoncollection/detail.aspx?id=118>
21. Jourde, F., Laurillau, Y., Morán, A.L., Nigay, L.: Towards specifying multimodal collaborative user interfaces: A comparison of collaboration notations. In: *Interactive Systems. Design, Specification, and Verification, 15th Int. Workshop, DSV-IS, Kingston, Canada*, pp. 281-286 (2008)
22. Seoktae Kim, Hyunjung Kim, Boram Lee, Tek-Jin Nam, and Woohun Lee. 2008. Inflatable mouse: volume-adjustable mouse with air-pressure-sensitive input and haptic feedback. In *Procs CHI '08*. ACM, New York, NY, USA, 211-224.
23. Lisa Kleinman, Tad Hirsch, and Matt Yurdana. 2015. Exploring Mobile Devices as Personal Public Displays. In *Procs MobileHCI '15*. ACM, New York, NY, USA, 233-243.
24. Lehman, S.L. and B.M. Calhoun, An identified model for human wrist movements. *Experimental Brain Research*, 1990. 81(1): p. 199-208.
25. Lexip 3D <http://www.amazon.fr/Lexip-3DM-PRO-Souris-Filaire-Noir/dp/B01IZU1NE4>
26. David Merrill, Jeevan Kalanithi, and Pattie Maes. 2007. Siftables: towards sensor network user interfaces. In *Procs of TEI '07*. ACM, New York, NY, USA, 75-78.
27. Laurence Nigay and Joëlle Coutaz. 1993. A design space for multimodal systems: concurrent processing and data fusion. In *Procs of INTERACT'93 and CHI '93*. ACM, New York, NY, USA, 172-178.
28. Gary Perelman, Marcos Serrano, Mathieu Raynal, Celia Picard, Mustapha Derras, and Emmanuel Dubois. 2015. The Roly-Poly Mouse: Designing a Rolling Input Device Unifying 2D and 3D Interaction. In *Procs of CHI '15*. ACM, New York, NY, USA, 327-336.
29. Razer naga mouse. <http://www.razerzone.com/gaming-mice/razer-naga>
30. Mike Rooke and Roel Vertegaal. 2010. Physics on display: tangible graphics on hexagonal bezel-less

- screens. In *Procs of TEI '10*. ACM, New York, NY, USA, 233-236.
31. James Scott, Shahram Izadi, Leila Sadat Rezai, Dominika Ruskowski, Xiaojun Bi, and Ravin Balakrishnan. 2010. RearType: text entry using keys on the back of a device. In *Procs of MobileHCI '10*. ACM, New York, NY, USA, 171-180.
  32. Julian Seifert, Sebastian Boring, Christian Winkler, Florian Schaub, Fabian Schwab, Steffen Herrdum, Fabian Maier, Daniel Mayer, and Enrico Rukzio. 2014. Hover Pad: interacting with autonomous and self-actuated displays in space. In *Procs of UIST '14*. ACM, New York, NY, USA, 139-147.
  33. Marcos Serrano, Laurence Nigay, Jean-Yves L. Lawson, Andrew Ramsay, Roderick Murray-Smith, and Sebastian Deneff. 2008. The openinterface framework: a tool for multimodal interaction. In *CHI EA '08*. ACM, New York, NY, USA, 3501-3506.
  34. David Sirkin and Wendy Ju. 2014. Make this!: introduction to electronics prototyping using arduino. In *CHI EA '14*.
  35. Soukoreff, R. W., & MacKenzie, I. S. (2004). Towards a standard for pointing device evaluation: Perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies*, 61, 751-789.
  36. Masanori Sugimoto and Keiichi Hiroki. 2006. HybridTouch: an intuitive manipulation technique for PDAs using their front and rear surfaces. In *Procs of MobileHCI '06*. ACM, New York, NY, USA, 137-140.
  37. The trackball keyboard. <http://www.ione-usa.com/ione-scorpis-35-trackball-keyboard.html>
  38. Jayson Turner, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2015. Gaze+RST: Integrating Gaze and Multitouch for Remote Rotate-Scale-Translate Tasks. In *Procs of CHI '15*. ACM, New York, NY, USA, 4179-4188.
  39. Anna Vallgård and Ylva Fernaeus. 2015. Interaction Design as a Bricolage Practice. In *Procs of TEI '15*. ACM, New York, NY, USA, 173-180.
  40. Nicolas Villar, Shahram Izadi, Dan Rosenfeld, Hrvoje Benko, John Helmes, Jonathan Westhues, Steve Hodges, Eyal Ofek, Alex Butler, Xiang Cao, and Billy Chen. 2009. Mouse 2.0: multi-touch meets the mouse. In *Procs of UIST '09*. ACM, New York, NY, USA, 33-42.
  41. Xing-Dong Yang, Edward Mak, David McCallum, Pourang Irani, Xiang Cao, and Shahram Izadi. 2010. LensMouse: augmenting the mouse with an interactive touch display. In *Procs CHI '10*. ACM, New York, NY, USA, 2431-2440.



# The Roly-Poly Mouse: Designing a Rolling Input Device Unifying 2D and 3D Interaction

Gary Perelman<sup>1</sup>, Marcos Serrano<sup>1</sup>, Mathieu Raynal<sup>1</sup>,  
Celia Picard<sup>2</sup>, Mustapha Derras<sup>2</sup>, Emmanuel Dubois<sup>1</sup>

<sup>1</sup>University of Toulouse, IRIT  
Toulouse, France  
{first\_name.last\_name}@irit.fr

<sup>2</sup>LRA, Berger-Levrault  
Toulouse, France  
{first\_name.last\_name}@berger-levrault.fr

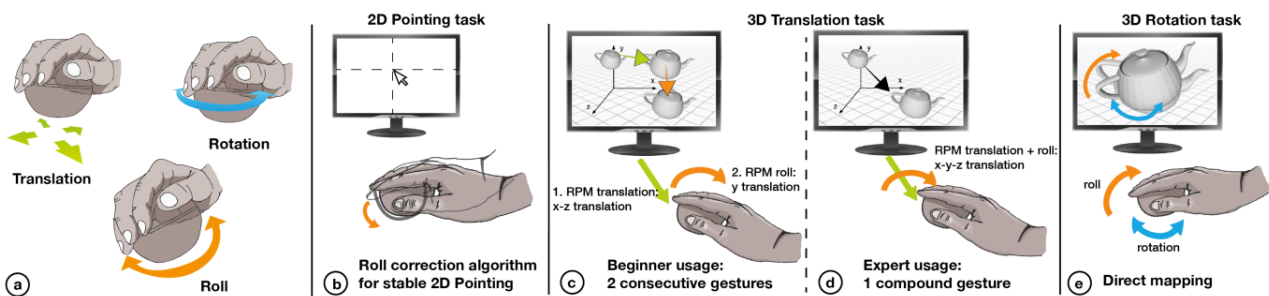


Figure 1.a) The Roly-Poly Mouse (RPM) gestures. RPM can be used for b) 2D pointing; c-d) 3D translation by combining RPM translation and roll; e) and 3D rotation by combining RPM roll and rotation.

## ABSTRACT

We present the design and evaluation of the Roly-Poly Mouse (RPM), a rolling input device that combines the advantages of the mouse (position displacement) and of 3D devices (roll and rotation) to unify 2D and 3D interaction. Our first study explores RPM gesture amplitude and stability for different upper shapes (Hemispherical, Convex) and hand postures. 8 roll directions can be performed precisely and their amplitude is larger on Hemispherical RPM. As minor rolls affect translation, we propose a roll correction algorithm to support stable 2D pointing with RPM. We propose the use of compound gestures for 3D pointing and docking, and evaluate them against a commercial 3D device, the SpaceMouse. Our studies reveal that RPM performs 31% faster than the SpaceMouse for 3D pointing and equivalently for 3D rotation. Finally, we present a proof-of-concept integrated RPM prototype along with discussion on the various technical challenges to overcome to build a final integrated version of RPM.

## Author Keywords

Input Device; 2D Pointing; 3D Interaction.

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).  
CHI 2015, April 18 - 23, 2015, Seoul, Republic of Korea  
Copyright 2015 ACM 978-1-4503-3145-6/15/04...\$15.00  
<http://dx.doi.org/10.1145/2702123.2702244>

## ACM Classification Keywords

H.5.2. Information interfaces and presentation: Interaction

## INTRODUCTION

3D applications are more popular than ever with the advent of easy-to-use 3D editors (SketchUp), 3D printing and web 3D engines (Unity Web Player, Flash 3D). Affordable multi-degrees-of-freedom (DOF) devices, such as the SpaceMouse (with over a million units sold [27]), contribute to this development. However these devices are not well suited for 2D pointing in WIMP interfaces [11] (3D editor menus, web browser GUI or when switching to 2D applications). They usually work in rate control mode, less convenient for 2D pointing [36]; DOF are difficult to operate independently [11]; and some devices are bulky, or tiresome to use such as the Phantom [24]. Therefore users need to switch between the mouse and the 3D device, which is tedious and inefficient [11]. Relying only on the mouse for 3D tasks is less effective [1,9]. An all-in-one device would therefore remove device-switching costs (Homing in KLM [6]) and improve the workflow of the users of 3D applications in 2D WIMP environments.

Previous works have augmented a regular 2D mouse with additional DOF [1,13] by adding rounded edges to enable rocking gestures. These devices are limited by their initial mouse-like form factor, with a flat surface to assure the device's stability. The different DOF, i.e. mouse translation and rocking, are physically separated and the flat spot at the base of these devices physically drives the user to control only two degrees of freedom simultaneously. Besides, the rocking amplitude is limited to a rather small curved surface, thus diminishing the range of input values. In our



approach, rather than providing a mouse with additional DOF, we choose to provide a rounded shape, intrinsically offering 3 DOF rotations, with capabilities for handling 2D translations.

In this paper we present the design of a rolling input device offering up to 6 DOF: the Roly-Poly Mouse (RPM). This device mimics the well-known roly-poly toy: the base of the device is hemispherical and can be rolled (2DOF), rotated (1DOF) and translated (up to 3DOF) in any direction (Figure 1). The key benefits of RPM make it a good candidate for an all-in-one device: RPM is free-moving as a mouse; has a large curved surface for rolling; and allows simultaneous and coordinated gestures to be performed by combining roll, rotation and translation. The design and usage of a symmetrical device with no stability constraint raises a number of challenges that we address in this paper: we study amplitude and precision along the different DOF; we propose a roll correction algorithm to support stable 2D pointing; and we design compound gestures for efficient 3D translation and rotation. We evaluate RPM against the SpaceMouse for 3D interaction, and results confirm that RPM is an efficient 3D device.

We build a proof-of-concept integrated prototype with a 'ring' button accessible on any device orientation and a 6DOF magnetic sensor. Using this prototype, we explore some usage scenarios that illustrate that RPM is not only interesting to unify 2D and 3D interaction, but also to extend several types of 2D applications by adding gestural control or commands.

Our contributions are 1) the design of a novel rolling input device through the experimental exploration of several design factors, 2) a novel roll-correction algorithm to support stable 2D pointing, 3) the design of a compound gesture for 3D interaction and a comparison with the SpaceMouse, and 4) a proof-of-concept wireless prototype along with usage scenarios and a discussion on technical challenges to build a final integrated version of RPM.

## RELATED WORK

### Mouse with multiple DOF

Prior works have augmented the regular mouse with additional DOF. Some of them added one DOF using two mouse sensors to support yawing [18,23] or a pressure sensor to control multiple levels of discrete selection modes [8, 15]. These devices are however limited to 3 DOF.

Other mouse augmentations using 2 or 3 additional DOF have been proposed to allow 3D manipulations [1,13]. Rockin'Mouse [1] is a seminal input device with the shape of a regular mouse but with rounded bottom allowing the device to be tilted and thus offering two additional DOF. VideoMouse [13] is a similar device allowing for two more DOF (z-axis rotation and translation) by using a camera. These two planar multi-DOF devices derive from the mouse's form factor: they add a rounded border to the bottom but keep the flat surface on the base to preserve

device stability. One limitation of this design principle is that the narrow rounded surface restricts tilting degrees ( $\pm 20^\circ$  for the VideoMouse). Moreover even if Rockin'Mouse and VideoMouse allowed for partial compound gestures, these were limited by their form factor: the flat spot at the base of these devices physically drives the user to control only two degrees of freedom simultaneously. Studies with the Rockin'Mouse on a 3D task revealed that movements involving both planar and tilting movements occurred during the ballistic phase of the trial, while the final closed-loop phase involved one dimension at a time. Our device design is intended to increase the use of compound gestures thanks to its large hemispherical bottom that enables extended rolling and facilitates 4D gestures such as roll + translation.

The DesktopBat [28], a 5 DOF mouse, is based on the opposite approach: attaching a dome on top of a mouse, rather than below. The dome acts as an isometric input. No evaluation of this device in comparison with others has been reported so its key benefits are difficult to assess. The SpaceMouse is a 6DOF device [27] that requires dexterity as DOF are difficult to perform separately [9]: used with the non-dominant hand it is less efficient than regular mouse for 3D interaction [3]; however with the dominant hand, it has been reported to outperform the mouse [9].

### Other techniques for 2D and 3D interaction

With the advent of multitouch surfaces, touch-based solutions for interacting with 3D environments have been proposed [10,12,20]. StickyTools [12] is a full 6DOF technique for rotation and translation of 3D objects. DS3 [20] is another technique based on the separation of translation and rotation that proved to be faster than StickyTools [20]. However, touch-based solutions are slower than the mouse or dedicated devices for 6D docking tasks [10]. Mid-air gestures have also been used to interact with 3D environments [24,34]. AirMouse [24] is a technique for 3D pointing in mid-air with a similar performance to that of 3D devices. 6D Hands [33] is a hand tracking system that supports bi-manual gestures for 3D manipulation. However, mid-air gestures induce fatigue and are not well suited for long interactions [24].

### Round input devices

Our work is inspired by previous research on round devices [9,31,36,37]. The FingerBall (or FBall) [36] is a free-moving (isotonic) spherical device that can be held and moved in mid-air. The spherical form factor was intended to improve "precision grasp", i.e. using one's dexterity. The device showed better performance than a glove for a 6DOF docking task [37]. PALLA is a similar wireless device used to explore interaction with video games [31]. Globefish and GlobeMouse [9] are two 6DOF devices based on combining isotonic rotation with isometric translation. These devices perform well in rotation but, as with isometric devices, not in translation.

**THE ROLY-POLY MOUSE (RPM)**

The RPM is based on the form factor of the well-known roly-poly toy (Figure 2). The bottom is hemispherical, contains weight and its center of mass is low. When released, the device returns to its initial upright position after an inertial roll movement.

**Gestures: translation, roll and rotation**

The device offers up to 6 DOF with three types of simple gesture (Figure 1): translation (3 DOF), roll (2 DOF: pitch and roll) and rotation (1 DOF: yaw). In the context of desktop usage, we don't consider the z-axis translation due to fatigue: our studies explore the use of a 5 DOF RPM.

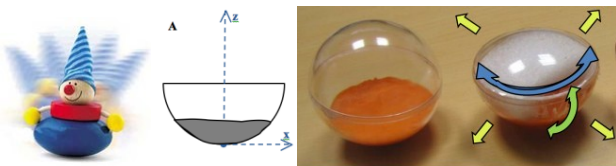


Figure 2. The roly-poly toy (a) has a low center of mass (b). Two RPM upper shapes: hemispherical and convex (c)

**Key benefits**

This form factor presents several key benefits that make it interesting for building an input device. As the device is **free-moving** in the x-y plane it can be used like an isotonic device such as the mouse [5]. The device contact with the surface is minimal, thus minimizing displacement friction. At the same time the device rolls and auto-repositions itself when released, **similar to an isometric device** [5]. The hemispherical bottom offers a large curved surface allowing for a wide range of possible input values. The device is symmetrical along its rotation axis (z) and thus can be used in any orientation. The three simple gestures (translation, rolling and rotation) can be combined to create **compound gestures**. For instance, effective 3D pointing with RPM is based on a compound gesture: translation + roll (Figure 1).

To sum up, this form factor combines characteristics from isotonic and isometric devices, making it a good candidate for an **all-in-one device** (Table 1 extends table in [13]). A 5 DOF device with such properties can be "very appealing to users who are reluctant to change their ways" [18].

Table 1. Comparison of sensed DOF for several multi-DOF mice (\*Tz can be sensed by RPM but is not used).

Input DOF	Regular Mouse		Rockin' Mouse [1]		Video Mouse[13]		Space Mouse [27]		Roly-poly Mouse	
	Sensed	Rate/Pos	Sensed	Rate/Position	Sensed	Rate/Position	sensed	Rate/Pos	Sensed	Rate/Position
Tx	✓	p	✓	P	✓	p	✓	r	✓	p
Ty	✓	p	✓	P	✓	p	✓	r	✓	p
Tz	×		×		✓	p	✓	r	✓	p*
Rx	×		✓	r,p	✓	r,p	✓	r	✓	r,p
Ry	×		✓	r,p	✓	r,p	✓	r	✓	r,p
Rz	×		×		✓	r	✓	r	✓	r,p
T+R	×		×		×		✓	r	✓	r,p

**DEVICE DESIGN FACTORS**

As a first step in the design process of RPM, we identify and analyze the various factors that affect its usage.

**Radius dimension**

The size of RPM is based on the average size of a regular mouse (approx. 12x6cm). We carried out informal tests with three different diameter dimensions: 6, 8 and 10 cm. The 8 cm version was the easiest to handle and translate so all our RPM prototypes are based on this dimension.

**Selection mechanism**

The RPM cannot hold regular front buttons that would be difficult to reach when the device rotates. Two always-available solutions are to use a 'ring' button around the device (as the one in our working prototype) or a capacitive surface. In our studies we focus on RPM gestures and shape and don't evaluate these alternatives, left for future work.

**Upper shape**

We considered different shapes for the upper half of the hemisphere; hereafter, they are referred to as "upper shapes". Focusing on a symmetrical form that a hand would comfortably hold, we opted for a hemisphere and varied its degree of curvature and direction (in and out). In our preliminary study we initially consider three upper shapes: hemispherical, convex (curved out) and concave (curved in).

**Hand grip posture**

Previous psychological and physiological studies on the grasp of tangible objects show that the type of grasp (number and position of fingers) depends on the object shape as well as on the goal [22]. Grasp postures have also been previously used as an input modality [29]. Consequently, in our preliminary study we identify the most frequently adopted hand postures and explore them in subsequent studies to evaluate their impact on RPM gesture amplitude.

**Gesture amplitude**

Each one of the three gestures (translation, roll and rotation) has different amplitudes. The RPM is a wireless device and thus its movement in the x-y plane is unlimited. Concerning the roll and the rotation, their amplitudes are theoretically +90/-90° for roll and unlimited for rotation. However the hand posture and biomechanical limitations of the joints involved, such as the wrist, will restrain this theoretical amplitude [19,26]. By considering different hand postures in our first study, we collect the possible range of values for each type of gesture.

**DOF Integration and Separation**

The different DOF of the RPM are physically integrated: when performing a displacement, there will be some rolling; when rolling or rotating, the device's center of mass probably translates. To allow for a proper separation of the DOF of the device [14], we need to know how each DOF affects the others. We study this question in the first study of the paper.

**Stability: Roll correction for 2D pointing**

Due to the rounded shape of the device, we noticed unintended roll when translations are done. This poses a problem: when the device rolls, the center of the device moves and thus the mouse pointer is translated (Figure 3). Note that this problem is inherent to the device and not to the tracking technique: whether we consider the center of the device or the contact point with the surface, the problem exists.

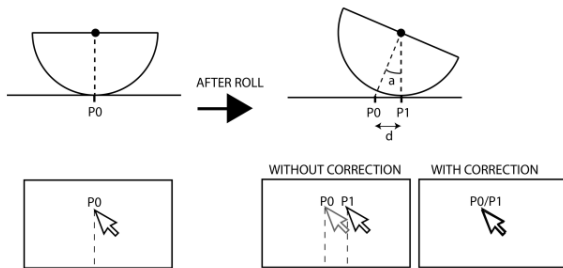


Figure 3. 2D pointing problem with a rolling device.

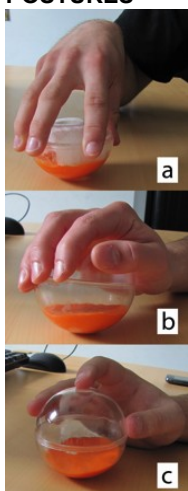
To solve this problem, we implemented a “roll correction” algorithm. This algorithm consists in calculating the position P0, which represents the device position upright, when the device is rolled into a position P1 (Figure 3). To calculate P0, we use the current position of the device P1, the roll and pitch angles  $a$  and  $p$ , and device radius  $R$  as follows:

$$P0.x = P1.x - \pi R * (\sin(a) + \cos(b))$$

$$P0.y = P1.y - \pi R * (\cos(a) + \sin(b))$$

In our implementation, when using the RPM as a mouse, we apply the algorithm for each frame to correct the roll deviation. We then calculate the displacement and apply the regular transfer function [7] to define the pointer position.

**PRELIMINARY STUDY: UPPER SHAPES AND HAND POSTURES**



The goal of this preliminary study was to collect most frequently used hand postures for each RPM gesture and user preference on the three considered RPM upper shapes.

**Study Description**

We studied 3 shapes (Concave, Convex and Hemispherical) and asked participants, for each shape, to hold the device with the dominant hand in four ways to perform four gestures: translation, roll, rotation and a compound gesture (in contact with the table). No interactive feedback was provided to participants and RPM was not tracked.

Figure 4. Most frequent hand postures.

We recruited 12 participants (2 females) aged 24.8 years on average (SD=1.8).

Two had previous experience with multiple DOF devices. We took 144 pictures corresponding to each hand posture (3 shapes x 4 gestures x 12 participants). We finally asked participants to order device shapes by preference. Each session lasted about 30 min.

**Results**

Over the 12 participants, 7 preferred Convex upper-shape, 4 Hemispherical and only 1 Concave. The Concave version of RPM was perceived to be difficult to hold and displace. In consequence we chose to remove it from the rest of our studies. Concerning the hand postures, we identified different recurrent patterns among which three were the most frequently adopted: Squeeze (Figure 4-a), Lay (Figure 4-b), and Touch (Figure 4-c). The following studies involve these three postures.

**EXPERIMENTAL RPM TRACKING SETUP**

In our following studies, we used the experimental tracking setup detailed below.

**Tracking system**

To track the translation, rotation and roll of the device we used infrared optical markers tracked by 12 OptiTrack cameras (1mm precision). The system senses the position (x, y, z) and orientation (yaw, pitch, roll) of RPM at 100Hz.

**Prototype**

The two RPM were weighted to be stable at rest (80gr each). Markers were placed on a support (Figure 5) to allow the cameras detect the device without impeding the user’s ability to grab the device with the three selected hand postures. Informal tests had also confirmed that the marker did not limit the amplitude of comfortable rolls: the maximum possible roll of RPM given these physical markers was 70° in the marker support directions. Our tracking setup did not register contact with the underneath surface, thus we did not use clutching in our experiments.



Figure 5. Experimental tracking setup.

**STUDY 1: GESTURES STABILITY AND AMPLITUDE**

We explore RPM gestures’ stability and maximum comfortable amplitude for each upper shape, considering various hand postures

**Study Description**

*Task and instructions*

Based on the preliminary study, we decided to study gestures stability and amplitude with two upper shapes (Convex and Hemispherical) and three hand postures (Squeeze, Lay, and Touch). We chose to control hand postures to avoid any confounding effect on the results as the user’s grip can have an important impact on the gesture amplitude. The task consisted in performing translations and rolls in 8 different directions and rotations in the two

possible directions. We included two different distances (12 and 24 cm) for the translation task to study its effect on RPM stability. All distances and directions were drawn on the experimentation surface with circles and lines respectively (Figure 6). We asked participants to perform translations in a comfortable way from the center to the edge of the circle and back again to the center. Concerning rolls and rotation, we asked them to perform the maximum comfortable gesture. We used the tracking setup detailed in the previous section.

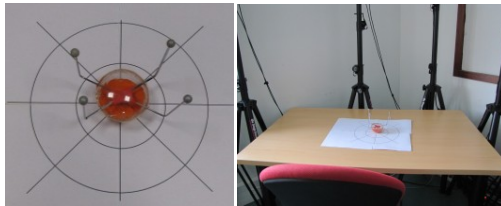


Figure 6. Study setup: distances and directions were drawn on the surface.

**Participants**

We recruited 12 participants (1 female) aged 25.7 years on average (SD=3.8). One had previous experience with multi-DOF devices and eleven took part in the pre-study session.

**Design and procedure**

The study follows a 2x3x4 within-subject design with Upper shape (Convex or Hemispherical), Hand posture (Squeeze, Touch or Lay) and Gesture (Short Translation, Long Translation, Rotation and Roll) as factors. We counterbalanced Upper shape and Hand posture. The study is made up of 6 blocks (each block is a combination of one Upper shape and one Hand Posture). Each block consists of 8 short translations, 8 long translations, 8 rolls and 2 rotations ordered randomly. After each trial, the device is placed in its initial position on the table.

**Collected data and statistical analysis**

We collected 2 devices x 3 hand postures x (16 translations + 8 rolls + 2 rotations) x 12 subjects = 1872 gestures. After each block we measured fatigue with a 6-20 Borg scale [4]. We used a Shapiro-Wilk test to determine the normality of the collected data. If the data was normal or could be normalized, we used a Univariate ANOVA test. If not, we used a Friedman test to compare more than 2 conditions and Wilcoxon tests otherwise. When we performed more than one statistical test on a particular set of data, we used the Bonferroni correction.

**Results**

**Gestures amplitude**

**Rotation:** We report the rotation amplitude for each direction (left and right), Hand posture and device Upper shape. A Wilcoxon test reveals no significant effect of the device upper shape on the rotation amplitude ( $Z=-0.98$ ,  $p=0.33$ ), thus we report mean values considering both upper shapes. A Wilcoxon test confirms a difference between left and right rotations ( $Z=-3.13$ ,  $p=0.0017$ ): left rotations are

on average larger than right rotations ( $57^\circ$  vs.  $48^\circ$ ) (Figure 7-a). Results on Hand posture indicate that Squeeze allows for a larger average rotation than Touch or Lay ( $66^\circ$  vs.  $45^\circ$  and  $48^\circ$ ;  $\chi^2(2)=24$ ,  $p<.001$ ).

**Roll:** We report results on rolling amplitude (i.e. maximum degrees rolled in a given direction). An ANOVA test establishes a significant difference between RPM upper shapes ( $F_{1,528}=34.319$ ,  $p<.001$ ). The maximum average rolling amplitude is significantly higher with the Hemispherical RPM ( $39^\circ$ ) than with the Convex RPM ( $37^\circ$ ). Rolling amplitudes range from  $27^\circ$  to almost  $60^\circ$  depending on roll direction and hand posture (Figure 7-b,c).

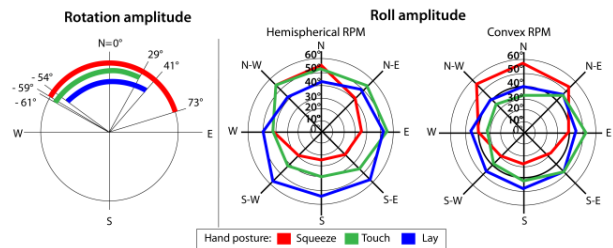


Figure 7. a) Rotation amplitude in degrees for each hand posture; b) rolling amplitude for the Hemispherical version and c) for the Convex version.

**Gesture stability**

**Roll:** Results on roll stability, i.e. deviation from the rolling direction, reveal no significant effect of the RPM Upper shape ( $F_{7,528}=2.68$ ,  $p=0.09$ ). Overall, deviation ranged from  $8^\circ$  to  $17^\circ$  for each direction (Figure 8). These results confirm that 8 directions can be reached without error, thus RPM roll can be used to interact with a Marking Menu for instance [17].

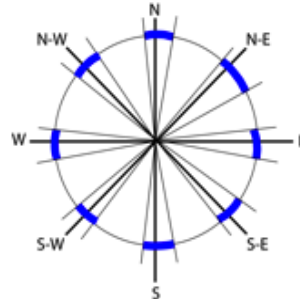


Figure 8. Rolling precision for each rolling direction

**Translation:**

Results on translation stability, i.e. the amount of unintended rotation and roll during translations, reveal that there is on average  $12^\circ$  of roll and  $15^\circ$  of rotation while displacing RPM. There is no significant difference between Hemispherical and Convex RPM ( $Z=-4.29$ ,  $p=0.08$ ). While a Wilcoxon test establishes a significant difference between long and short translations ( $Z=-2.34$ ,  $p=0.018$ ), this difference is limited to  $1^\circ$  for roll. Finally no significant differences have been observed between hand postures ( $Z=-0.78$ ,  $p=0.44$ ).

**Fatigue**

The average Borg [4] score obtained is 11.72 (on a scale of 6-20). Upper shape does not significantly affect fatigue. However, fatigue measured when using the Squeeze posture is significantly higher than with other postures (Friedman  $\chi^2(2)=8.46$ ,  $p=0.0144$ ).



### Summary

Our study on gesture amplitude reveals that the Hemispherical upper shape allows for the largest rolling amplitude; rotation amplitudes vary from left to right and is larger with the Squeeze hand posture. Concerning precision, roll allows for 8 precise directions; and translation stability is affected by unexpected rotation and roll. Given the fatigue results for Squeeze, we removed this posture from the following studies. From these results we define a roll threshold of  $12^\circ$  to avoid unintended activations and consider amplitudes up to  $42^\circ$  (to enable  $30^\circ$  rolls).

### STUDY 2: 3D MANIPULATION WITH THE RPM

Taking advantage of the key benefits of RPM, we propose a technique combining RPM roll and displacement to perform 3D translations, and RPM roll and rotation to perform 3D rotation. We evaluate our solution against a dedicated commercial device, the SpaceMouse.

#### RPM 3D Interaction Techniques

As we use 5 DOF, applying RPM to 3D manipulation relies on a modal use: translation and rotation are performed separately. As RPM offers 2 DOF for translation, 2 DOF for roll and 1 DOF for rotation, any 3DOF task will be performed using a combination of these three gestures.

##### 3D translation: roll & displacement

To perform 3D translation of an object, we use a position control technique: 2D physical displacement of the RPM is mapped to 2D translation of the cursor on the x-z plane in the 3D scene. The RPM front/back roll controls the elevation of the x-z plane, i.e. the y position of the plane. This technique offers the advantage to be easy to use for beginners and to support the transition to experts. Beginners can decompose a 3D translation task by performing first a 2D movement on the x-z plane, then a roll to define the y value. Advanced users can simultaneously combine displacement and roll in the same coordinated gesture to directly perform a 3D translation.

##### 3D rotation: roll & rotation

To perform 3D rotations, two modes are possible: position control or rate control. In position control, RPM rotation and roll are directly mapped to the object 3D orientation. This mode requires clatching, as informal tests showed that to be comfortably and efficiently operated, RPM roll can be mapped to a maximum of  $\pm 90^\circ$  rotation in the virtual scene. In rate control, the RPM roll and rotation angles are mapped to the rotation velocities of the 3D object. Informal tests revealed that position control outperformed rate control.

#### Study description: 3DOF pointing and docking

In this study we compare RPM with a 3D dedicated device, the SpaceMouse [27], in two separated 3D tasks: pointing and docking [21]. We do not compare with traditional mouse as it has been reported to be less efficient than the SpaceMouse for 3D interaction [9].

#### Tasks and instructions

We decided to conduct two sessions, each one dedicated to one 3D task: the first session consisted of performing 3D translations (3D pointing task), and the second one of 3D rotations (3D docking task). The reason we divided our experiment into two subtasks, instead of a single 6DOF docking task [21,37], is that the number of conditions for the pointing task was quite large, and combining both would make the experiment excessively long. We also wanted to study separately the performance of our device for 3D translation and for 3D rotation, and combining both would make it difficult to analyze each subtask. Previous research has proven that, even if both tasks can be carried out jointly, users usually perform them independently [21].

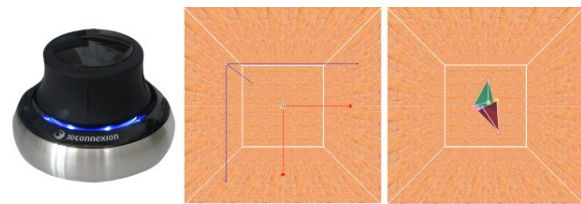


Figure 9. SpaceMouse device (left) and 3D environment for the 3D pointing (center) and 3D docking (right) tasks.

In the 3D pointing task, the participant is asked to reach a 3D spherical target in a 3D scene by translating a cursor (Figure 9). The target is displayed in the center of a cube delimiting the 3D environment. The cursor initial position is situated in 8 different directions (combining equal x, y and z translations) at two different distances (12.99 and 8.66) from the target. We implemented two different target sizes (0.5 and 1.7) to produce 4 Index of Difficulties (2.6, 3.11, 4.19 and 4.75 bits). In the 3D docking task, the user is asked to rotate a tetrahedral cursor until it fits the orientation of the tetrahedral target (Figure 9). The initial orientation is a non-trivial combination of two or three axis rotations giving 8 different orientations. In both tasks, the user starts a trial and validates the trial by pressing the space bar on the keyboard. To perform clatching with RPM the user presses the spacebar, as adding a second selection key for RPM would increase user's mental load. Therefore we did not measure selection errors for 3D docking since both clatching and validation used the same key. Participants were instructed to perform the task with their dominant hand as quickly and accurately as possible.

#### Apparatus

We use the same optical system and RPM settings as in the previous experiment. To define the gain value of the SpaceMouse rate technique, we carried out a number of tests. Some users found the regular gain too fast. We thus decided to test two gain values in our study: a gain similar to the default behavior of the device (best for most users after some training) and a smaller gain favouring precision over speed (best for most beginners). With RPM, we used a gain factor of 1:1 for translations and 1:3 for rotations and

rolls. The experiment software was developed in C++ using Irrlicht 3D engine.

### Participants

We recruited 12 participants (6 female), all right-handed, aged 28.5 years on average (SD=6). 5 were used to 3D interaction and 3 of them participated in previous studies.

### Design

The 3D pointing session follows a 4x4 within-subjects design with Interaction Technique IT (RPM Hemispherical, RPM Convex, SpaceMouse Small-Gain and SpaceMouse Default-Gain) and Index of Difficulty ID (2.60, 3.11, 4.19 and 4.75 bits) as factors. The 3D rotation session has one within-subject factor, Interaction Technique, with the same 4 values than the previous session. Each session was divided into 4 blocks, each one corresponding to an interaction technique. Order of blocks is counterbalanced across participants by means of a 4x4 Latin Square. Cursor positions and IDs (for pointing) and orientations (for rotation) were randomly ordered inside each block.

### Procedure

We performed a large training session for both tasks: users performed 30 trials/IT before the experiment, and then again 8 trials before each block. No constraints were given with respect to hand posture as we did not want to force hand posture on the SpaceMouse.

### Collected data and statistical analysis

We logged all tracking data and measured completion time from stimulus onset. We also collected user preference and usability using two 5-point Likert scales to rate each interaction technique as well as fatigue using a 6-20 Borg scale. For the 3D pointing task, we collected 384 trials per user (4 IT x 8 directions x 4 IDs x 3 repetitions) x 12 users = 4608 trials in total. The session lasted about 73 min. For the 3D rotation task, we collected 192 trials per user (4 IT x 8 orientations x 6 repetitions) x 12 users = 2304 trials in total. The session lasted about 55 min. We used the same statistical analysis approach than in previous studies.

## Results

### 3D pointing performance

Concerning completion time (Figure 10), a Friedman test reveals a significant difference between the interaction techniques ( $\chi^2(3)=25.2$   $p<.001$ ). The two RPM versions (2.28 s) are faster than the two SpaceMouse versions (3.68 s) ( $Z=-13.34$ ,  $p<.001$ ). This result is true for any ID value and the difference grows with ID difficulty (29% for easier ID vs. 33% for harder ID). There is no significant difference between ITs concerning error rate, that was overall around 5% ( $\chi^2(3)=0.28$   $p=0.96$ ).

### 3D docking performance

We found no significant difference between the two RPMs and the two SpaceMouse versions concerning completion time ( $\chi^2(3)=2.3$   $p=0.51$ ).

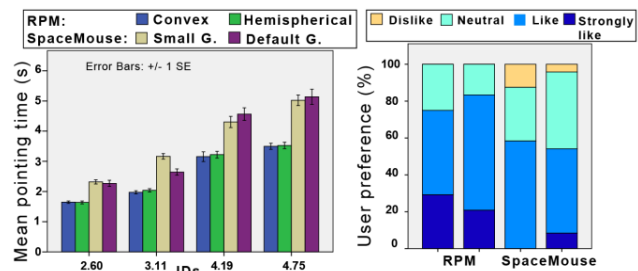


Figure 10. Mean time in s for the 3D pointing task for each IT and ID (left) and user preference (right).

### User preference

There were no significant difference in terms of user preference or fatigue between both tasks; therefore we present the overall results. In terms of preference (Figure 10), 75% of the users rated Convex RPM positively (with a score of 4 or higher); 80% Hemispherical; 54% the regular SpaceMouse and 58% the slower version. In terms of usability, 75% rated Convex positively (with a score of 4 or higher); 87% Hemispherical; 45% the regular SpaceMouse and 54% the slower version. Users found that “the SpaceMouse needs more concentration than RPM” (p4, p9) and that “RPM is more pleasant because I can perform the movement with a classical cursor position control” (p6, p8, p12).

### Fatigue

In terms of fatigue, there was no significant difference between the interaction techniques ( $\chi^2(3)=3.50$   $p=0.32$ ), with an average value of 11 (6-20 Borg scale).

## Summary

This study demonstrates that there is a significant difference in terms of 3D pointing time performance between the two RPM versions and the two SpaceMouse versions. RPM is 31% faster than the SpaceMouse. Concerning the 3D docking task, there is no significant difference between the different devices and SpaceMouse gains in terms of completion time. In terms of user’s preference, both RPM version are considered more usable and are rated more positively than the two SpaceMouse versions. Overall, this study demonstrates that RPM can be used effectively as an input device for 3D translation and rotation and that users preferred and found RPM more usable than the SpaceMouse.

## TOWARDS AN INTEGRATED VERSION OF RPM

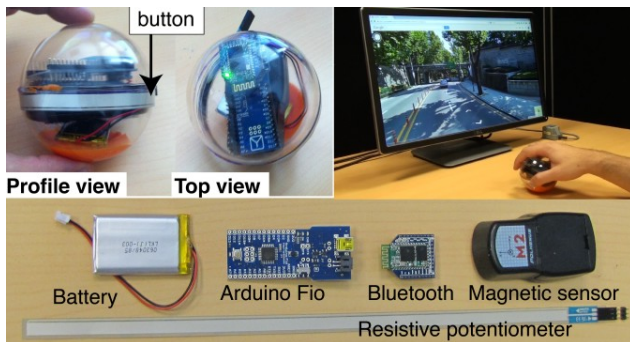
We build on our experience in prototyping a proof-of-concept version of RPM to analyze the various technical challenges that need to be overcome to build an integrated final version of RPM.

### Proof-of-concept wireless prototype

Based on our studies results, we created a working wireless Hemispherical RPM prototype with a ‘ring’ button all around the device. To detect the 3D displacement and rotation of RPM, we used a Polhemus Patriot Wireless tracker (7x3x2.5cm, 79.4gr.). We filtered the tracker data



using a 1€ filter. To build the button we used a similar approach to the WatchIt bracelet [25]: the button consists of one resistive potentiometer (81x7.5x0.5mm) that provides up to 1024 values depending on the touch position. It is thus possible to combine the clicked position on the ring with the current RPM orientation to support multiple buttons. To interface it, we used an Arduino Fio board with a Bluetooth shield and an external battery (Figure 11). Using this prototype, we implemented two usage scenarios that we describe in the discussion section: Google StreetView and a drawing application.



**Figure 11. Proof-of-concept wireless prototype with a ‘ring’ button used in StreetView (top). Components (bottom).**

#### Technical challenges

Building a final working version of RPM presents several technical challenges.

*Position:* to track RPM position several embedded and non-embedded solutions are possible. Embedded solutions include using magnetic sensors (although not precise enough to allow mouse-like pointing) or an embedded mini board camera similar to VideoMouse [13] (however requiring the use of a grid surface). Not-embedded solutions include using IR cameras or an underlying sensitive surface, similar to Wacom’s tablets. All of these solutions would allow detecting the z-dimension at a certain level.

*Orientation:* to track the orientation of the device, we could use an embedded inertial measurement unit (IMU).

*Contact with the underneath surface:* to enable mouse-like clutching, embedded solutions include using a resistive bottom, a distance sensor or a camera as in VideoMouse [13]. This contact can be easily detected if using an underlying sensitive surface.

*Hand contact:* To detect when the device is held, a capacitive surface could be used. This could prevent unwanted movements of the cursor when RPM is released.

*Selection mechanism:* There are two main options to integrate an always-available selection mechanism invariant to rotation: an all-around ‘ring’ button and a capacitive surface. The ‘ring’ button, situated all around the device as in our working prototype, permits multi-touch input with several fingers. A more elegant solution would be to use a

multi-touch surface. However, this solution would lack of haptic feedback, like on a regular button. This solution has proved to be useful for extending a traditional mouse [2,32,35] such as Apple’s Magic Mouse in which it is combined with a mechanical switch. The main challenge with this solution would be to distinguish a finger touch from a palm contact or with a grasping/squeezing gesture.

*Selection stability:* Pressing a button could cause RPM to move. To ensure RPM stability, we should favor a selection mechanism positioned all around the device: a finger pressure would then exert perpendicularly to roll direction.

*Device repositioning:* even if the low center of mass of the device allows auto-repositioning, the device wobbles when released. To limit this oscillation, in particular when switching between the mouse and the keyboard, a solution based on a self-balancing mechanism coupled with the inertial measurement unit could be integrated. The main challenge would be to fit the hardware within the RPM volume.

## DISCUSSION

### All-in-one device

Our paper illustrates that RPM constitutes an all-in-one device capable of handling 2D and 3D tasks. Our studies demonstrate that it performs 31% faster than the SpaceMouse for 3D translation. Most users preferred the Hemispherical upper shape for 3D interaction. Our work initially explores the many other possibilities for such a novel device based on the combination of translation, roll and rotations as illustrated in the usage scenarios.

### Other usage scenarios

While our main motivation to build RPM was to create an all-in-one device for 2D and 3D interaction, we explored other usage scenarios that could benefit from RPM key features. We carried out a one-hour design session with 15 participants to collect application scenarios. Many participants came up with use cases related to 3D interaction, ranging from 3D games to 3D editing. They also gave numerous ideas on a large variety of applications: in video editing; music composition; graphic drawing; web browsing; map navigation; marking menus; or exploring large datasets. For each idea, participants mapped different functions to the RPM degrees of freedom. For instance, in a map application, translation is used to translate the view and roll to adjust (rotate left/right, zoom in/out). In a drawing application, translations are mapped to drawings, rotations to the selection of the drawing colors and rolls to thickness and type of drawn lines. This design session illustrates the possible range of usage applications for RPM.

### Design guidelines for rolling devices

Our studies allow us to sum up a set of design guidelines for the future adoption of rolling devices. The hemispherical upper shape has proved to be the best in terms of roll amplitude. When mapping control to roll and rotation gestures, designers should take into account the

orientation as it enables different amplitudes. Roll gestures can be performed precisely in at least 8 directions and a roll threshold of  $12^\circ$  should be used to avoid false activations. For 3D pointing, using compound gesture is the most efficient. For 3D rotation, position control could be coupled with rate control as in [30].

#### Device instability

Unlike the traditional mouse or previous tilting devices such as the Rockin'Mouse, which have a flat bottom, the RPM bottom is hemispherical: the device is thus not stable. This is particularly true in two cases: when the device is released, as it wobbles, and when the device is moved, as it rolls accidentally. In the first case, a simple solution is to detect when the user is holding the device. The device can then be "turned off" when no touch is detected. In the second case, when the user moves the device on a flat surface like a mouse, an appropriate threshold must be used to avoid false positives due to unexpected roll.

#### Z-dimension

We did not explore this dimension in our paper as we believe it introduces the issue of fatigue. However a sporadic use is possible, for instance holding RPM on the palm during a presentation. In this context RPM could be used to perform mid-air gestures to control slides. We plan to further explore this perspective in the future.

#### LIMITATIONS AND FUTURE WORK

Our exploration was limited by the use of markers on RPM, by the tracking technology, by the lack of selection and clutching mechanism and by the separation of 3D tasks. The use of IR tracking allowed us to precisely measure all the gestures. As a counterpart we had to carefully place IR markers on the Roly-Poly Mouse in an unobtrusive way. Even if these markers may have had a minor effect on user's gestures, we plan to validate our findings with a wireless marker-free prototype. In addition, our experiments did not evaluate RPM in a 2D pointing task nor the use of an embedded selection and clutching mechanism. In the future we plan to compare RPM with the regular mouse for 2D pointing and evaluate different solutions as discussed above, such as the 'ring' button used in our proof-of-concept prototype. Concerning RPM size, currently based on the regular mouse size, it could be interesting to explore a smaller version of RPM where the user can manipulate it while resting the hand on the table. Finally, we plan to carry further studies on the performance of RPM for an integrated 6D-docking task to evaluate the impact of mode change.

In the future we plan to evaluate 2D pointing with RPM against a classical mouse as said earlier. A long-term study on comfort is pertinent to further explore fatigue issues. We will further explore the design space of RPM compound gestures by proposing and evaluating novel techniques for some of the aforementioned applications, for instance to manipulate graphical items orientation and position as in [16]. Finally, we plan to explore mid-air and free-roll

gestures and propose novel GUIs for RPM, such as circular menus or widgets.

#### CONCLUSION

In this paper we present the design of a novel input device, the Roly-Poly Mouse (RPM) combining the advantages of the mouse and of 3D devices by allowing three types of gestures: translation, roll and rotation. To enable stable 2D pointing with RPM, we propose a roll correction algorithm. We identified preferred RPM upper-shapes and most frequently adopted hand postures through a preliminary study. Based on these results, a first study on gestures amplitude and stability reveals that our device allows large and precise rolls to be performed. In a second study we compare the two versions of RPM with a popular 3D device, the SpaceMouse, in two 3D tasks: pointing and docking. We propose a novel compound gesture for the 3D translation based on rolling and displacing the device. Results reveal that RPM performs 31% faster than the SpaceMouse in terms of translation time and equivalently in terms of rotation time. Users preferred Hemispherical RPM for 3D interaction. Finally, we used our experience in implementing a proof-of-concept prototype to identify the various challenges to overcome to build a final integrated version of RPM. To sum up, RPM is an all-in-one device that removes device-switching costs, improving the workflow of users.

#### ACKNOWLEDGEMENTS

This work was carried out as part of the research collaboration agreements IRIT-Berger Levraut (CNRS No. 099856) and LMIAMDC. The authors thank Philip D. Gray, Gilles Bailly and the members of the ELIPSE group for their useful suggestions.

#### REFERENCES

1. Balakrishnan, R., Baudel, T., Kurtenbach, G., and Fitzmaurice, G. 1997. The Rockin'Mouse: integral 3D manipulation on a plane. CHI '97. ACM, 311-318.
2. Benko, H., Izadi, S., Wilson, A., Cao, X., Rosenfeld, D., and Hinckley, K. 2010. Design and evaluation of interaction models for multi-touch mice. GI '10. CIPS, 253-260.
3. Bérard, F., et al. 2009. Did "Minority Report" Get It Wrong? Superiority of the Mouse over 3D Input Devices in a 3D Placement Task. INTERACT '09. Springer-Verlag, 400-414.
4. Borg, G. Borg's Perceived Exertion and Pain Scales. Human Kinetics (1998), viii 104pp.
5. Buxton, W. Lexical and pragmatic considerations of input structures. Comput. Graph., 17(1) :31-37, 1983.
6. Card, S., Moran, T., and Newell, A. 1980. The keystroke-level model for user performance time with interactive systems. Commun. ACM 23, 7, 396-410.
7. Casiez, G. and Roussel, N. 2011. No more bricolage!: methods and tools to characterize, replicate and compare pointing transfer functions. UIST '11. ACM, 603-614.

8. Cechanowicz, J., Irani, P., and Subramanian, S. 2007. Augmenting the mouse with pressure sensitive input. CHI '07. ACM, 1385-1394.
9. Froehlich, B., Hochstrate, J., Skuk, V., and Huckauf, A. The globefish and the globemouse: two new six degree of freedom input devices for graphics applications. CHI'06. ACM,191-199.
10. Glesser, D., Bérard, F., and Cooperstock, J. 2013. Overcoming limitations of the trackpad for 3d docking operations. CHI EA '13. ACM, 1239-1244.
11. Hachet, M., Guitton, P., and Reuter, P. 2003. The CAT for efficient 2D and 3D interaction as an alternative to mouse adaptations. VRST '03. ACM, 225-112.
12. Hancock, M., Cate, T. T., and Carpendale, S. Sticky tools: full 6dof force-based interaction for multi-touch tables. In Proc. of ITS'09, ACM, 133-140.
13. Hinckley, K., Sinclair, M., Hanson, E., Szeliski, R., and Conway, M. 1999. The VideoMouse: a camera-based multi-degree-of-freedom input device. UIST '99. ACM, 103-12.
14. Jacob, R., Sibert, L., McFarlane, D. and Preston Mullen, M. Integrality and separability of input devices. ACM TOCHI, 1, 1 (1994), 3-26.
15. Kim, S., Kim, H., Lee, B., Nam, T., and Lee, W. 2008. Inflatable mouse: volume-adjustable mouse with air-pressure-sensitive input and haptic feedback. CHI'08. ACM, 211-24.
16. Kruger, R., Carpendale, S., Scott, S., and Tang, A. 2005. Fluid integration of rotation and translation. CHI '05. ACM, 601-610.
17. Kurtenbach, G. and Buxton, W. 1993. The limits of expert performance using hierarchic marking menus. CHI '93 ACM, 482-487.
18. MacKenzie, S., Soukoreff, R., Pal C. 1997. A two-ball mouse affords three degrees of freedom. CHI-EA '97. ACM, 303-304.
19. Marteniuk, R. G., MacKenzie, C., Jeannerod, M., Athenes, S., and Dugas, C. Constraints on human arm movement trajectories. Canadian Journal of Psychology 41, 3 (1987), 365--378.
20. Martinet, A., Casiez, G., and Grisoni, L. Integrality and separability of multitouch interaction techniques in 3D manipulation tasks. IEEE Transactions on Visualization and Computer Graphics 18, 3 (2012), 369-380.
21. Masliah, M. R., and Milgram, P. Measuring the allocation of control in a 6 degree-of-freedom docking experiment. CHI'00, ACM, 25-32.
22. Olafsdottir, H., Tsandilas, T. and Appert, C. 2014. Prospective motor control on tabletops: planning grasp for multitouch interaction. CHI '14. ACM, 2893-2902.
23. Olwal, A, Feiner, S. 2004. Unit: modular development of distributed interaction techniques for highly interactive user interfaces. GRAPHITE '04. ACM, 131-138.
24. Ortega, M. and Nigay, L. 2009. AirMouse: Finger Gesture for 2D and 3D Interaction. INTERACT '09. Springer-Verlag, 214-227.
25. Perrault, S., Lecolinet, E., Eagan, J., and Guiard, Y. 2013. Watchit: simple gestures and eyes-free interaction for wristwatches and bracelets. CHI '13. ACM, 1451-60.
26. Rahman, M., Gustafson, S., Irani, P., and Subramanian, S. 2009. Tilt techniques: investigating the dexterity of wrist-based input. CHI '09. ACM, 1943-52.
27. SpaceMouse. Press Note March 31, 2011. <http://www.3dconnexion.fr/nc/company/press-room/>
28. Steed, A. and Slater, M. 1995. 3D Interaction with the Desktop Bat. In Computer Graphics Forum. Blackwell Publishers, vol 14, 97-104.
29. Taylor, B. and Bove, M. 2009. Graspables: grasp-recognition as a user interface. CHI '09. ACM, 917-926.
30. Tsandilas, T., Dubois, E., Raynal, M. 2013. Modeless Pointing with Low-Precision Wrist Movements. INTERACT 2013. Springer, p. 494-511.
31. Varesano, F. and Vernerio, F. 2012. Introducing PALLA, a novel input device for leisure activities: a case study on a tangible video game for seniors. FnG'12. ACM, 35-44.
32. Villar, N., Izadi, S., Rosenfeld, D., Benko, H., Helmes, J., Westhues, J., Hodges, S., Ofek, E., Butler, A., Cao, X., and Chen, B. 2009. Mouse 2.0: multi-touch meets the mouse. UIST '09. ACM, 33-42.
33. Wang, Y., MacKenzie, C., Summers, V. and Booth, K. 1998. The structure of object transportation and orientation in human-computer interaction. CHI '98. ACM, 312-319.
34. Wang, R., Paris, S., and Popović, J. 2011. 6D hands: markerless hand-tracking for computer aided design. UIST '11. ACM, 549-558.
35. Yang, X-D., Mak, E., McCallum, D., Irani, P., Cao, X., and Izadi, S. 2010. LensMouse: augmenting the mouse with an interactive touch display. CHI '10. ACM, 2431-40.
36. Zhai, S. 1998. User performance in relation to 3D input device design. SIGGRAPH Comput. Graph. 32,4,50-54.
37. Zhai, S., Human Performance in Six Degree of Freedom Input Control, 1995, University of Toronto

# Rolling-Menu: Rapid Command Selection in Toolbars Using Roll Gestures with a Multi-DoF Mouse

Emmanuel Dubois, Marcos Serrano, Mathieu Raynal

University of Toulouse - IRIT

Toulouse, France

{first\_name.last\_name}@irit.fr

## ABSTRACT

This paper presents Rolling-Menu, a technique for selecting toolbar items, based on the use of roll gestures with a multidimensional device, the Roly-Poly Mouse (RPM). Rolling-Menu reduces object-command transition, resulting in a better integration between command selection and direct manipulation of application objects. Selecting a toolbar item with Rolling-Menu requires rolling RPM in a predefined direction corresponding to the item. We propose a design space of Rolling-Menu that includes different roll mapping and validation modes. A first user's study, with a simple toolbar containing up to 14 items, establishes that the best version of Rolling-Menu takes, on average, up to 29% less time than the Mouse to select a toolbar item. Moreover accuracy of the selection with Rolling-Menu is above 90%. Both the validation mode and the mapping between roll direction and toolbar items influence the performance of Rolling-Menus. A second study compares the three best versions of Rolling-Menu with the Mouse to select an item in two types of multidimensional toolbars: a toolbar containing dropdown lists, and a grid toolbar. Results confirm the advantage of Rolling-Menu over a Mouse.

## Author Keywords

Toolbar; Command selection; Multidimensional device; Roly-Poly Mouse; hemispherical mouse.

## ACM Classification Keywords

H.5.2. Information interfaces and presentation: Interaction

## INTRODUCTION

Most desktop applications include some type of toolbars. The simplest version of toolbar is an array of icons, usually arranged in a horizontal way on top of the application window. More advanced forms of toolbars can arrange multiple items in a grid that structures a set of items into different subsets (like the Microsoft Ribbon); they can also include dropdown lists (such as the list to select a font size in Word). The usual interaction with toolbars is carried with

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

CHI 2018, April 21–26, 2018, Montreal, QC, Canada  
© 2018 Association for Computing Machinery.  
ACM ISBN 978-1-4503-5620-6/18/04...\$15.00  
<https://doi.org/10.1145/3173574.3173941>

the mouse through a classical pointing gesture. However, it requires the user to move the mouse from the object of interest (text in a text editor, drawing area in a graphic design application) to the toolbar, and then come back once the command is selected to pursue the main task. This object-command transition breaks the interaction flow [3].

To reduce such disruption, multiple solutions have been proposed in the literature. Keyboard shortcuts [8] enable immediate access to commands but need to be memorized, and therefore are usually applied only to a few frequent commands [9]. Contextual menus, such as marking menus, allow rapid access to a relatively large set of commands, but their use is still limited in real applications because they are hidden, change with the context, may occlude some of the underlying application and do not cover the complete set of commands. As a result improvements are still required to better take advantage of the benefits of toolbars, i.e. offering a constant and always available set of commands, and advantages of contextual menus or keyboard shortcuts, i.e. rapid access to commands with minimal interruption of the interaction flow.

In this work, we explore a novel way of interacting with toolbars by using a mouse with multiple degrees of freedom, the Roly-Poly Mouse (RPM) [17] illustrated in Figure 1. The RPM has a hemispherical bottom, allowing roll and rotation gestures in addition to the regular translation. We propose to exploit roll gestures to select toolbar commands: selecting a toolbar item with Rolling-Menu requires rolling RPM in a predefined direction corresponding to the item, while translation of the RPM is used to control the application pointer. The resulting Rolling-Menu thus presents the advantage of keeping the application pointer in the working area, and therefore contributes to limit work flow interruption.

We first identify and describe the design dimensions of Rolling-Menu. These design dimensions include different mappings between roll gestures and command selection, as well as different activation and validation mechanisms. We implemented eight versions of Rolling-Menu based on different combinations of our design dimensions.

We then explore the performance of these techniques. We first experimentally establish that using Rolling-Menu require less time than using the Mouse to select items in a toolbar containing up to 14 items. We then study how Rolling-Menus can be used to support a more efficient

interaction with two different types of complex toolbars: "dropdown toolbars", containing dropdown lists, and "grid toolbar", i.e. a grid of items. We discuss the advantages and limitations of Rolling-Menus in such complex toolbars.

Our contributions are: 1) an exploration of the design space of the use of roll gestures to select toolbar items; 2) a study showing that Rolling-Menus take less time than the regular Mouse for selecting up to 14 items in a simple toolbar; and 3) a second study showing that Rolling-Menus also perform better than the Mouse for more complex types of toolbars.

### STATE OF THE ART

The goal of Rolling-Menu is to reduce the object-command transition. We summarize the process of command selection and review previous work on rapid access to commands.

#### From object of interest to command selection

In most interactive situations, command selection is not the main task, and the user focuses on other objects of interest before and after selecting a command (i.e. text on a text editor, a graphical element on a graphic design application, etc.). Dillon et al. [6] identified two main stages in the process of command selection: object-to-command transition (or command selection), and command-to-object transition (back to object of interest). According to authors, the total cost of command selection should include both stages. Bailly et al. [3] described two additional stages: command-to-command transition (i.e. when a user applies several consecutive commands) and command-to-value transition (i.e. specifying parameters). In our work, we focus on the two main stages identified by Dillon. Our goal is to minimize the total time of object-to-command and back-to-object transitions.

#### Contextual menus

A major approach to minimize object-to-command transition is to bring the menu closer to the object of interest by using contextual menus. Contextual menus are most of the time invoked through a mouse right-click and can take different forms, from linear menus to circular or pie menus [11,12,20,22,29]. All these different forms of contextual menus share the same fundamental limitations: they only include a subset of the menus, they are context dependent (i.e. menu items vary for each invocation), are difficult to use when the mouse is close to the window border, and they are hidden while performing the main task. These limitations contrast with the properties that ensure the success of toolbars, used on most desktop applications: toolbars are always visible, and their content and position are stable over time. Our goal is to combine both approaches, i.e. minimizing the selection time as ensured by contextual menus and interacting with the popular toolbars.

#### Multimodal command selection

Another approach to reduce command selection time is to use additional input modalities in parallel with pointing. For instance, Hover Widgets [10], TiltMenu [24] and PushMenu [14] use stylus input dimensions to select commands: hovering, orientation and pressure respectively.

Multitouch input has also been used to offer fast command selection. Most of multitouch work has been carried in the context of tabletop or mobile touchscreens, such as Multi-Touch Menu [2] or Microroll gestures [21]. Recent work has also explored using multitouch on a laptop touchpad: MarkPad [7] consists of several gestural touchpad shortcuts to select commands. However our goal is to offer rapid access to toolbars on regular desktop computers, i.e. with a keyboard and a mouse.

#### Multi-DOF mice

Adding degrees of freedom (DoF) to the traditional mouse extends its command selection capabilities, while preserving the normal pointing interaction. Two main approaches have been used to add DoFs to the regular mouse. The first approach consists in combining the regular mouse with other input modalities through device composition [18]. For instance, LensMouse [27] consists of a mouse augmented with a touchscreen, and Inflatable Mouse [15] includes a pressure sensor.

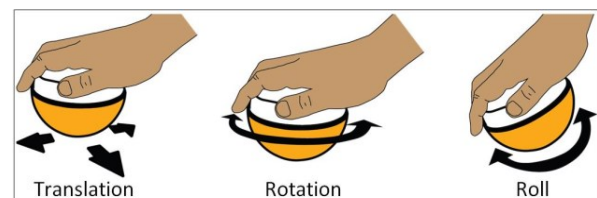
The second approach consists in modifying the shape of the mouse to allow supplemental physical manipulations. Rockin'Mouse [4] and VideoMouse [13] are similar to a regular mouse, but they have a rounded bottom allowing tilt gestures. The Roly-Poly Mouse (RPM) [17] is a hemispherical mouse offering a larger amplitude of rolls gestures. For this reason, we chose to study the use of roll gestures with RPM to select toolbar commands. To our knowledge, no previous work on multi-DoF mice has evaluated a solution for shortening object-command transitions on toolbars.

#### ROLLING-MENU

We present Rolling-Menu, a new technique for selecting a toolbar item with a roll gesture using the Roly-Poly Mouse (RPM) [17]. After describing the technique, we detail four design dimensions that lead to different versions of the technique: roll range, size of roll sectors, activation and validation mechanisms.

#### Roly-Poly Mouse: roll direction and amplitude

The Roly-Poly Mouse (RPM) is a multidimensional hemispherical mouse can be translated, rotated and rolled.



**Figure 1. Physical manipulations that can be applied to the Roly-Poly Mouse independently or in a combined way.**

Based on previous studies with this device, we decided to explore its rolling capabilities: translations are already used to control the application pointer, while rotations have a limited range and are not very comfortable [17]. A roll gesture is composed of two independent parameters:

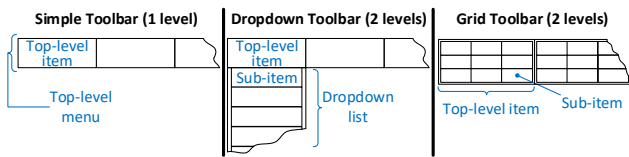


- *Roll direction*: From a top-down point of view on RPM, it indicates the direction in which the user tilts RPM (Figure 3, left). Direction can range from  $-90^\circ$  (tilt left) to  $+90^\circ$  (tilt right),  $0^\circ$  corresponding to a tilt forward.
- *Roll amplitude*: From a side view, it corresponds to the inclination angle of RPM, once it has been rolled (Figure 3, centre). The amplitude is  $0^\circ$  when RPM is in its initial upright position,  $90^\circ$  when RPM is perpendicular to the surface on which RPM is used.

**Selecting a toolbar item with Rolling-Menu**

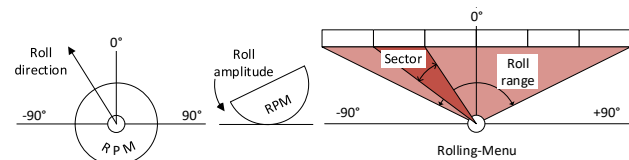
In this work we limit our explorations to horizontal toolbars, positioned on top of the window, similar to most applications toolbars. More specifically, we explore three variants of horizontal toolbars (see Figure 2):

- Simple toolbar: a single set of top-level items, where each item triggers a command;
- Dropdown toolbar: a toolbar where each top-level item is associated with a dropdown vertical list. Each sub-item from this list triggers a command;
- Grid toolbar: a toolbar structured into several blocks (hereafter referred to as the top-level items) that contain a grid of sub-items, each one associated with a command.



**Figure 2. Three toolbar variants: a simple toolbar, a toolbar with dropdown lists and a grid-structured toolbar.**

Using a Rolling-Menu to select a toolbar item relies on the activation of a quasimode [19], a specific mode of the application which ends automatically once the selection is performed, i.e. that does not require any dedicated action to exit. This quasimode is activated when the roll amplitude reaches a predefined threshold. In this quasimode the application pointer is frozen and the top-level menu is associated to a *range* of possible rolls (Figure 3, right). This range is divided into *sectors*, so that each top-level item is associated to one distinct sector. To select an item, the user rolls RPM in the direction of the corresponding sector.



**Figure 3. RPM roll direction (left: top-down view) and amplitude (centre: side view); Roll range and one sector of a Rolling-Menu (right: top-down view).**

When the user validates the selection, it will execute the command corresponding to the top-level item (simple toolbar), open the dropdown list (dropdown toolbar) or select

a grid block (grid toolbar). We explore in subsequent user studies different RPM manipulations to interact with these three variants of toolbars.

The benefits of a Rolling-Menu are twofold. First, it strongly decouples the menu navigation from the application pointer, reducing workflow interruptions due to object-command transitions (see first section of State of the art). Second, each item corresponds to a specific gesture and can therefore be encoded through muscle memory. This could potentially lead to a more effective expert mode [3].

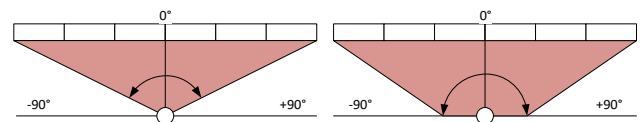
**Design dimensions**

Four design dimensions result from the previous general principle. Two of them are related to the roll mapping (type of roll range and size of the roll sectors). The two other dimensions are related to the use of the Rolling-Menu quasimode, which needs to be activated, before the item selection is validated. Therefore we consider different activation and validation methods in our design space.

*Roll Range*

We envisioned two types of roll range:

- The *Direct-range* ensures a coherent mapping between the physical roll directions of RPM and the width of the toolbar (Figure 4 – left). In this case, the roll range is a triangle, the summits being the left and right extremities of the toolbar, and the origin of the Rolling-Menu.
- The *180-range* maximizes the roll range by mapping  $180^\circ$  of RPM physical roll directions (from  $-90^\circ$  to  $+90^\circ$ ) to the toolbar width (Figure 4 – right). The roll range is here trapezoidal.



**Figure 4. Direct-range (left) vs. 180-range (right).**

We exclude roll directions above the  $180^\circ$  range because informal pre-tests revealed that the resulting mapping is hard to understand (as a user performs a roll back to select an item on the opposite direction), and previous studies showed that performing such rolls is less efficient [17].

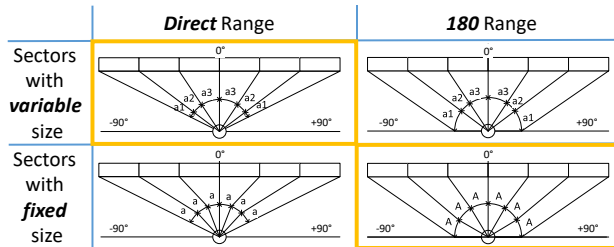
*Size of roll sectors*

Independently of the roll range, we identified two different ways for establishing a mapping between a roll sector and a toolbar item, based on the size of the sectors:

- *Variable size*: Each sector is defined through a direct mapping between a toolbar item and the origin of the Rolling-Menu (Figure 5 – line 1). As a consequence, the size of each sector is different from each other and depends on the position of the toolbar item: sectors on the extremities of the toolbar are smaller than those near the centre. This ensures a straight mapping between the RPM roll direction and the position of the toolbar item, but can make some items more difficult to select.



- **Fixed size:** The roll range is divided into the same number of sectors than the number of toolbar items. As a consequence, the size of each sector is identical (Figure 5 – line 2), which offers the same selection difficulty for every item, but may lead to a mismatch between the roll direction and the item direction.



**Figure 5.** Sectors with variable size (line 1) or fixed size (line 2) used with a direct roll range (left column) or a range of 180° (right column).

#### Activation method of the Rolling-Menu

To activate the quasimode without affecting the user's interaction flow, we avoid using a modifier key. Instead, we exploit the roll amplitude and defined an activation threshold. We designed two alternatives:

- **Bottom activation:** The user activates the quasimode as soon as the roll amplitude is greater than a predefined small threshold. We know that unintentional rolls occur when translating the RPM, preventing the use of any roll below 12° [17]. Therefore, rolls over 12° activate the quasimode (see Figure 6-left).
- **Top activation:** To activate the quasimode, the user needs to perform a roll with an amplitude greater than a predefined big threshold. Since rolls are considered comfortable up to 37° [17], we adopt this value as the big threshold. The advantage of this method is that the sector arcs are wider as roll amplitude increases, offering more precision to select an item (see Figure 6-right).

#### Validation mechanism

We considered two mechanisms to validate the selection of a toolbar item and trigger the execution of a command or the opening of the sub-menu.

- **Tap:** The user can perform a finger's tap (on any tactile surface underneath RPM or on the keyboard), to select a sector when the roll amplitude of RPM is over the activation threshold.
- **Roll:** it is based on a validation threshold depending of the activation method. For a Bottom activation, validation occurs when roll amplitude is greater than a validation threshold of 37°. A complete activation and validation gesture in this case consists on a straight roll. For a Top activation, validation occurs when roll amplitude becomes smaller than a validation threshold of 12°. A complete gesture in this case consists on a roll forward to activate, then a roll backward to validate. Validation thresholds (37° with Bottom activation, and 12° for the Top activation)

were chosen so that the same angle exists between the activation and validation thresholds, whatever the activation method.

#### IMPLEMENTING ROLLING-MENUS

Our four design dimensions produce a large number of design combinations (16 possible Rolling-Menus). We decided to implement and study a subset of 8 Rolling-Menus after an analytical and empirical exclusion of the others. We carried an iterative process to design the visual feedback for these versions and implement the input apparatus.

#### Choice of Rolling-Menu versions

Among the four possible combinations of Roll range and Size of sectors, we selected two of them, hereafter referred to as RMDirect and RM180:

- **RMDirect** (Figure 5, top left) combines a **Direct range** with a **variable** sector size. It offers the most direct mapping between roll actions on RPM and the position of toolbar items: the roll range corresponds to the width of the toolbar and the roll direction corresponds to the direction in which the toolbar item is.
- **RM180** (Figure 5, bottom right) combines a **180 range** with a **fixed** sector size. It minimizes the accuracy required by maximizing the size of sectors: the roll range is extended to 180° and each sector has the same size, which depends on the number of toolbar items.

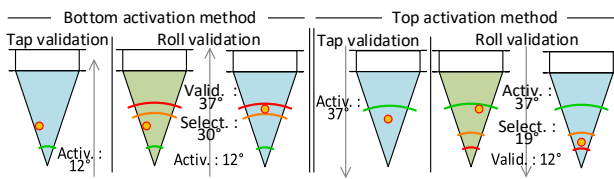
We implemented these two designs with the two methods of validation (tap or roll) and activation (bottom or top), resulting in 8 different interaction techniques, hereafter referred to as the Rolling-Menus (RM).

#### Implementing the validation mechanisms

We tried to insert various forms of button on top of RPM. However our pre-tests showed that using a physical button on RPM altered the device handling gesture and brought a number of technical issues (button position, etc.). Instead, we considered the use of a tactile surface underneath RPM to detect a user's finger tap: the user can employ any finger of the same hand that manipulates RPM to tap on the surface, although participants seemed to prefer the thumb. An algorithm associates the first touch on the tablet to the RPM position, and triggers a tap event only when detecting a second touch. Alternatively, the user can press a key on the keyboard with the non-dominant hand: as the user's main task is probably involving keyboard input, this bimanual setting offers a fluid interaction compatible with regular keyboard input (the keyboard is only used as a validation once the Rolling-Menu quasimode is activated).

Regarding the roll validation, we also decided to test an additional threshold to increase the robustness of the validation: the selection threshold. Between this selection threshold and the validation threshold, a modification of the roll direction does not change the sector selected (i.e. the sector is locked). We empirically established that an angle of 7° between the selection and validation thresholds was the most appropriate. Selection thresholds are therefore 30° with

Bottom activation method and 19° with Top activation method (see Figure 6).



**Figure 6. Graphical feedback provided in the 4 possible combinations of validation and activation methods.**

### Visual feedback

After the activation of the Rolling-Menu quasimode (through a roll), a visual feedback shows the roll sectors, from the centre of the screen to the toolbar items. In addition a coloured circle (10px diameter) is displayed: its position conveys the RPM roll direction and amplitude (see Figure 6). Finally the selected sector is coloured: it becomes blue as soon as the selection can be validated.

### RPM Input apparatus

To interact with the Rolling-Menu, we enhanced the original Roly-Poly Mouse [17]. As the original one, our version consists of a sphere with a diameter of 8 cm, which includes a Bluetooth enable Inertial Measurement Unit (xIMU by xIO Tech – sensor rate: 512 Hz, angular accuracy: 1°). In comparison to the original RPM, our version was placed on a Wacom Intuos 3D tablet (216x135 mm, resolution: 2540 lpi). As the tablet is multitouch, it can detect the translation of RPM and finger taps (Figure 7-right). We therefore covered the RPM surface with a graphite lacquer (Graphit 33 – Kontakt Chemie) to give the device a conductive coating (Figure 7-left and centre).



**Figure 7. Overview of the RPM input apparatus (left): rolling RPM (centre) and tapping on the tablet to validate (right).**

### PRELIMINARY STUDY: ADJUSTING ROLLING-MENU SETTINGS

We performed a preliminary user study to fine tune our design dimensions (in particular the roll thresholds). The goal was also to assess the ability of Rolling-Menu to tackle a large set of items and to study the impact of the activation point (i.e. the centre of the roll sectors).

### Design and procedure

The task consisted in selecting items from a simple toolbar positioned on the top of the screen and containing 4 to 10 items. The activation point was placed vertically at the middle of the screen, and horizontally in one of three different positions: left, centre or right of the screen. Toolbar items were 25px height and their width varied from 90 to 225px according to the number of items.

The eight Rolling-Menus we implemented (see section "Choice of Rolling-Menu versions") were compared to the Mouse. With the Mouse, the task consisted in clicking on the starting point (a circle with a diameter of 100px), selecting the highlighted top-level item and returning to the starting point. Completion time was measured between the two mouse-press events. This task simulates the object-to-command and command-to-object transitions, which are all part of the total command selection time, as explained in the related work section. With Rolling-Menus, the application pointer is separated from the roll interaction, and selecting a toolbar item consists simply in rolling the RPM in the direction of the toolbar item to select.

The toolbar and feedback were displayed on a 17" screen (1280px by 1024px). Participants were sitting in front of it. We used the keyboard spacebar as validation method. Twelve participants (aged 29.4 on average; SD=9.5) took part in this experiment. All of them were University students or researchers. After a training period, each participant performed 9 techniques x (4+6+8+10) top-level items x 3 starting points = 756 trials. They filled a SUS questionnaire for each technique.

### Data analysis

Regarding the data analysis, we chose to rely on estimation techniques with 95% confidence intervals and ratio analysis as recommended by the APA [25]. Ratio is an intra-subject measurement that expresses the effect size (pair-wise comparison) and is computed between each of the geometric means. All CIs are 95% BCa bootstrap confidence interval. For the reader more used to interpret the p-values, a parallel might be drawn with results obtained through the estimation technique and CIs reports (see Figure 3 in [16]). Scripts used to compute the geometric mean and confidence intervals were used in [26] and are available online [1].

### Results

Results established that on average Rolling-Menus using the tap validation always takes less time (1322ms CI[1155;1563]) than the Mouse (1719ms CI[1562,1918]) to select a top-level item, even when removing the time of the final mouse press (74ms on average). They also established that all Rolling-Menus with 4 or 6 items always require less time than the Mouse (from 14.2%-CI [7.9%, 22.8%] to 34.0%-CI [24.3%, 43.4%]) to select an item. The accuracy ranges from 94% to 97.5% for the 9 techniques. We thus decided to consider larger toolbars, with up to 14 items.

Results also highlighted that with the Mouse, the time taken from the menu selection and back to the starting point represent on average 47% of the total time, thus justifying the need to avoid the object-to-command transition. In addition it appeared that selecting an item requires on average 17.7% less time when the starting point is on the centre than on the left or right of the screen. Using always the same origin for the Rolling-Menu would in addition offer a more stable interface and could facilitate command memorization since each toolbar item would be associated to

a unique roll direction. For the following experiments we therefore considered the central starting point only.

We finally observed that, for the tap validation method, the average roll amplitude when the spacebar is pressed is  $30.9^\circ$ ,  $CI[27.0^\circ; 34.7^\circ]$ . Informal comments also stated that the roll amplitude of the validation threshold was too large. In addition, the SUS scores of Rolling-Menus using roll validation mechanism were the lowest (72 on average, vs. 85 for the other RM and 90 for the Mouse). From these results it appears necessary to reduce the roll amplitude required when validating (bottom activation) or activating (top activation). This modification also affects the selection threshold. We thus empirically defined new thresholds for a Bottom activation (activation at  $15^\circ$ , selection at  $26^\circ$  and validation at  $31^\circ$ ) and for a Top activation (activation at  $31^\circ$ , selection at  $29^\circ$  and validation at  $24^\circ$ ).

### STUDY 1: ITEM SELECTION IN A SIMPLE TOOLBAR

The goal of this study is to explore the impact of the four design dimensions (roll range, sector size, activation method and validation method) on the selection of a simple toolbar item using the eight Rolling-Menus, compared to a baseline (Mouse). We applied the settings derived from the preliminary study. The goal is also to study the impact of the number of toolbar items on completion time, accuracy and user preference.

#### Study description

##### Task and instruction

The task consisted in selecting items in simple toolbar displayed on top of the screen. The toolbar included 8, 10, 12 or 14 items. To limit the length of the experiment, we preliminary identified, in each toolbar, 8 items uniformly spread over the width of the toolbar: these items were then the only items targeted during the experiment. The height of the toolbar items was 25px and their width ranged from 102 to 180px depending on the number of items in the toolbar. We asked participants to perform the selection task as quickly as possible and with accuracy.

##### Apparatus

We used the same screen and mouse than in the preliminary experiment. We used the RPM apparatus described in the Implementation section. Tap validation was performed on the underlying touch surface with the hand holding RPM, in order to validate the use of a more integrated RPM version.

##### Participants

We recruited 16 participants (8 female, 2 left-handed), aged 27.9 years on average ( $SD=9.9$ ). All of them were University students or researchers. Six of them took part in the preliminary study. The experiment lasted 75min on average.

##### Design

This experiment followed a 4x9 within-subjects design with Number of top-level items (8, 10, 12 and 14) and Interaction techniques (Mouse and eight Rolling-Menus) as factors. Each session was divided into 9 blocks, each block corresponding to one Interaction technique (the Mouse or

one of the eight Rolling-Menus). Half of the participants used the Mouse prior to the eight Rolling-Menus, while the other half used the Mouse after them. The eight Rolling-Menu blocks were counterbalanced across participants by means of a 4x4 Latin Square.

##### Procedure

Each participant completed nine blocks. Each block contains a training and an experimental session. The training consists in selecting 8 top-level items in four successive toolbars containing 8, 10, 12 or 14 items. The training session contains 8 items x 4 toolbar sizes = 32 trials. The experimental session is similar to the training session but with two repetition for each toolbar size. It contains 8 items x 2 repetitions x 4 toolbar sizes = 64 trials. Participants had the possibility to take a break between each trial. Overall, each participant performed 9 blocks x 64 trials = 576 trials (without training). In total we collected 16 participants \* 576 trials = 9216 trials.

##### Collected data and data analysis

We logged all tracking data (RPM rolls, translations and taps) and measured completion time from stimulus onset. Participants also had to fill in a SUS questionnaire after each block (i.e. for each technique). Data analysis is performed with the same approach than in the preliminary study. As underlined in [28], many studies on menu techniques have focused on selection time, accuracy and learnability. In our case we did not address learnability and focused on time and accuracy, since our goal is to reduce object to command transitions [3].

### Results

We first report quantitative results and then discuss qualitative results. We use the following naming convention to refer to the eight implemented Rolling-Menus: RM\_"Activation"\_"Range"\_"Validation" where:

- "Activation" is T for Top or B for Bottom;
- "Range" is 180 for large or Dir for direct roll range;
- "Validation" is Roll or Tap.

#### Quantitative results

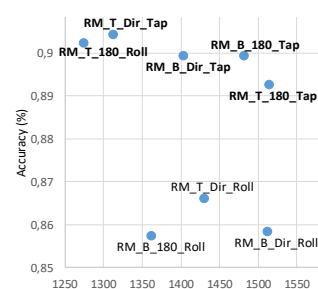


Figure 8. Cross-analysis of selection time (ms) and accuracy (%).

Among the eight techniques compared to the Mouse, a group of five Rolling-Menus offered the best accuracy, with selection times similar to the three remaining ones (five Rolling-Menus bolded in Figure 8). We focus on these five techniques and compare them in detail with the Mouse in terms of selection time and accuracy.

**Selection time analysis.** Computing averages and 95% confidence intervals for the selection time of the Rolling-

Menus establishes that selecting a toolbar item with any Rolling-Menu (average of all Rolling-Menus: 1410ms, CI[1222; 1658]) is faster than with the Mouse (1790ms, CI[1705; 1886]), as detailed in Figure 9-left.

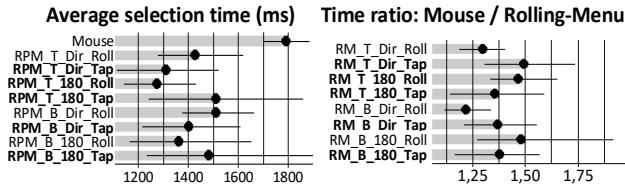


Figure 9. Average selection time in ms (with 95% CIs) (left), and time ratio (with 95% CIs) of the Mouse / Rolling-Menu.

When focusing on the five Rolling-Menus identified in the cross analysis (selection time X accuracy), three of them are definitely faster (RM\_T\_Dir\_Tap: 1312ms, CI[1116;1523]; RM\_T\_180\_Roll: 1274ms, CI[1146;1432], RM\_B\_Dir\_Tap : 1403ms, CI[1217; 1610]) than the Mouse to select a top-level item. Due to the large confidence intervals of the 2 other Rolling-Menus (RM\_T\_180\_Tap and RM\_B\_180\_Tap), which intersect with the Mouse confidence interval, the difference is less obvious. The intra subject analysis based on the time ratio (Mouse/Rolling-Menu) confirms the differences already highlighted between these different techniques (Figure 9-right): the ratio is above 1 (meaning that from an intra-subject point of view, the selection time with the mouse is greater than the selection time with the Rolling-Menu) and the corresponding CIs do not intersect with the value 1.

**Accuracy analysis.** Now, regarding the average accuracy of each interaction technique, the same analysis establishes that the Mouse is more accurate (95.1%, CI [93.6; 96.3]) than any of the Rolling-Menus (average of all Rolling-Menus: 89.0%, CI[85.0%; 92.0%], see details in Figure 10-left); the analysis of the accuracy ratio (Mouse/Rolling-Menu) confirm these conclusions (see Figure 10-left). The five Rolling-Menus identified in the cross-analysis are very similar in terms of accuracy (average of the five techniques: 90.0%, CI [86.9; 92.3], see details in Figure 10-left) and they are clearly but only slightly more accurate than the three other Rolling-Menus (average of the three techniques: 86.0%, CI[79.2%; 90.1%], see details in Figure 10-left). The ratio analysis also confirms the validity of this result (Ratio>1 and CIs not intersecting the value 1, see Figure 10-right).

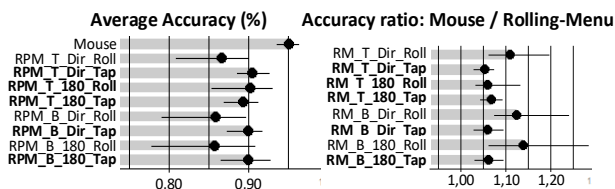


Figure 10. Average accuracy percentage (with 95% CIs) for each technique (left), and accuracy ratio (with 95% CIs) of the Mouse / Rolling-Menu.

These results in terms of selection time and accuracy still hold for any toolbar size (8, 10, 12 or 14 items): selecting an item in a simple toolbar with any Rolling-Menu is always faster than with the Mouse (Figure 11 – left), but with less accuracy (Figure 11 – right). As expected, we also notice that selection time increases and accuracy decreases when the toolbar size augments (Figure 11 – left).

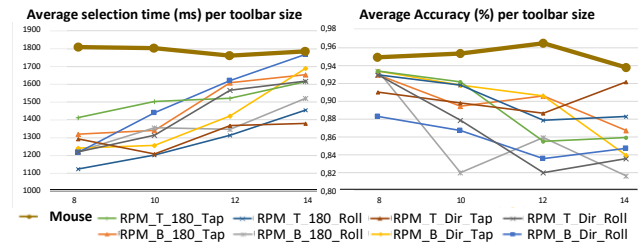


Figure 11. Average time (ms) and accuracy (%) for each technique with the four toolbar sizes (Mouse line is thicker).

**Complementary analysis.** Further data analysis, focusing on the design dimensions independently, did not reveal any major and clear distinction in terms of **activation method** (Top: 1382ms, CI[1219;1546]; Bottom: 1439ms, CI[1293;1636]), in terms of **validation method** (Roll: 1394ms, CI[1287;1551]; Tap: 1427ms, CI[1238;1643] or in terms of **roll range** (180: 1407ms, CI[1234;1626]; Direct:1414ms, CI[1291;1549]). We can therefore conclude that it is the combination of these design dimensions that affects the overall performance of the technique.

Finally, results establish that the average selection time of the first and second repetitions are very similar (repetition1: 1891ms, CI[1756; 2047]; repetition2: 1858ms, CI[1716; 2014]). We conclude that the training session was sufficient and that it is easy to learn how to use the Rolling-Menu.

**Qualitative results**

We computed SUS scores to assess the usability of the techniques. The Mouse reaches an average score of 80.8 while the average SUS score of the 8 Rolling-Menus is 72.0, which corresponds to a “good” usability level [5]. SUS scores obtained by the three best Rolling-Menus are above this average (RM\_T\_Dir\_Tap: 76.0; RM\_T\_180\_Roll: 73.3; RM\_B\_Dir\_Tap: 71.3).

**Summary**

Results of the first experiment strongly identify three of the eight Rolling-Menus as the best techniques for selecting an item in a simple toolbar:

- RM\_T\_Dir\_Tap: Top activation, direct roll range, validation with tap,
- RM\_T\_180\_Roll: Top activation, 180° roll range, validation via RPM roll,
- RM\_B\_Dir\_Tap: Bottom activation, direct roll range, validation with tap.

They take on average 25% less time than the Mouse, with an accuracy above 90%, to select a simple toolbar item. These



three Rolling-Menus therefore represent the best combinations of our design dimensions for selecting an item in a simple toolbar with up to 14 items.

## STUDY 2: DROPDOWN AND GRID TOOLBARS

The goal of this experiment is to compare the three best Rolling-Menus from the first study with the Mouse, for selecting a sub-item in two different toolbar variants: the dropdown and grid toolbars (cf. Figure 2).

### Study description

#### Task and instruction

The task consisted in selecting one top-level item, and then one sub-item in a Dropdown toolbar and a Grid toolbar. Since experiment 1 revealed that accuracy tends to decrease when the size of the toolbar increases, we limit this experiment to toolbars containing 8 top-level items. As illustrated in Figure 2, each top-level item is associated to 1) a dropdown list containing 8 sub-items for the Dropdown toolbar, or 2) a 3x3 grid of sub-items for the Grid toolbar. We asked participants to perform each selection as quickly and accurately as possible.

Selecting all items from both levels would make the overall experiment too long, so we decided to predefine target positions, balancing their difficulty. For the top-level selection, we predefined 4 positions (2, 4, 5 and 7, from left to right) to cover the left, middle and right of the toolbar. In the dropdown lists, we predefined 3 positions (2, 4 and 7, from top to bottom) to cover the top, middle and bottom of the list. In the Grid toolbar, we randomly selected 3 positions among the 3x3 possible sub-item positions to cover all possible directions.

#### Apparatus

We used the same apparatus than in the first experiment. The same feedback was provided during the first phase of the task, i.e. the selection of the top-level item. For Rolling-Menus using the Tap validation method, we asked participants to use a bimanual approach (pressing a keyboard key with the non-dominant hand): our pre-tests showed that while taping on the underlying tablet with the hand holding RPM worked well to select a top-level item, it disrupted the fluidity of a two-level selection. Indeed, a sub-item selection requires a rapid sequence of rolls / translations of RPM and the pre-tests revealed that moving the thumb to tap on the tablet is difficult to perform during this rapid combination of RPM gestures. As explained earlier, this bimanual validation does not specifically requires to move the non-dominant hand, which is usually already on the keyboard. Further, it does not interfere with the regular use of the keyboard since the key validation only works when the user activates the menu quasimode.

With the dropdown toolbar, rolling RPM towards the targeted top-level item and validating it opens the dropdown list. To navigate through the list of sub-items, the user rolls RPM forward or backward. A final tap validates the sub-item selection.

With the grid toolbar, rolling RPM towards the target top-level item and validating it highlights the central sub-item of the 3x3 grid. To select one of the 8 surrounding sub-items on the grid, the user translates RPM in the corresponding direction (similar to a Marking Menu). Validating the central item selection was based on a tap input. While using two different validation methods according to the sub-item position (tap for the central item, translation for the others) could seem to be cognitively complex, our pre-tests showed that participants did not mix up both types of validation.

#### Participants

We recruited 12 participants (5 female, 2 left-handed), aged 28.8 years on average (SD=8.8). All of them were University students or researchers. Eight of them took part in the preliminary study and/or the first experiment. The experiment lasted one hour on average.

#### Design and procedure

This experiment followed a 4x2 within-subjects design with Interaction techniques (Mouse and three Rolling-Menus) and Toolbar variant (Dropdown toolbar and Grid toolbar) as factors. Each participant completed 2 phases, each one corresponding to one toolbar variant. We used the same order for all participants (Dropdown then Grid), as we did not want to compare the toolbars between them. Each phase was composed of 4 blocks corresponding to each interaction technique. Blocks were counterbalanced across participants by means of a 4x4 Latin Square.

For each technique, participants first had a training session composed of 2 series: each series consisted of selecting 3 sub-items in the 4 predefined top-level items (i.e.  $2 \times 3 \times 4 = 24$  training trials). The experimental session then followed the same procedure and was composed of 6 series, corresponding to  $6 \times 3 \times 4 = 72$  trials. Participants had the possibility to take a break between each trial.

Overall each participant performed 2 phases (toolbar variant) x 4 blocks (techniques) x 72 trials = 576 trials (without training). Over the 12 participants we collected 6912 trials in total.

#### Collected data and data analysis

We logged all tracking data and measured completion time from stimulus onset. We also asked participants to fill in a SUS questionnaire after each block (i.e. for each combination of technique and toolbar variant). At the end of each phase (i.e. toolbar variant), participants were requested to rank the four techniques according to their preference and were invited to comment about them. Data analysis is performed according to the same approach than in study 1.

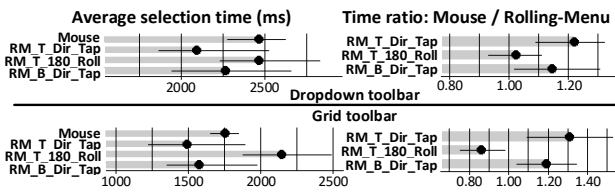
### Results

We first discuss selection time and accuracy for both types of toolbars before reporting qualitative feedback.

#### Quantitative results

**Selection time.** The average selection time per technique tends to establish that on a Dropdown toolbar (Figure 12-Top), Rolling-Menus with Tap validation require less time

than the Mouse and the other Rolling-Menu for selecting a sub-item (RM\_T\_Dir\_Tap: 2099ms, CI[1862;2527]; RM\_B\_Dir\_Tap: 2263ms, CI[1940;2658]); Mouse: 2462ms, CI[2273;2628]; RM\_T\_180\_Roll: 2465ms, CI[2230;2830]). The intra-subject analysis based on the time ratio Mouse/Rolling-Menu (Figure 12-Top right) strongly confirms that Rolling-Menus with Tap validation took up to 18% less time than the Mouse (ratio=1.22 and 95% CIs not intersecting the value 1.0). This ratio analysis also shows that no obvious difference can be established between RM\_T\_180\_Roll and the Mouse.

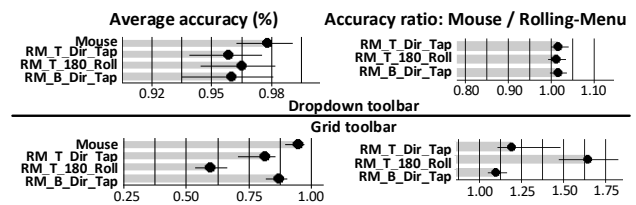


**Figure 12.** Sub-item selection time in a Dropdown toolbar (top) and Grid toolbar (bottom). On the left, the average selection time in ms and, on the right, time ratio of the Mouse / Rolling-Menu (all bars represent 95% CIs).

Regarding the selection of a sub-item in a Grid toolbar (Figure 12-bottom), Rolling-Menus with Tap validation also require less time than the Mouse and the other Rolling-Menu (RM\_T\_Dir\_Tap: 1493ms, CI[1222;1894]; RM\_B\_Dir\_Tap: 1577ms, CI[1348;1980]); Mouse: 1753ms, CI[1653;1853]; RM\_T\_180\_Roll: 2146ms, CI[1872; 2491]). From the ratio analysis, we can strongly conclude that these two Rolling-Menus takes up to 24% less time than the Mouse (ratio=1.31 and 95% CIs not intersecting the value 1.0). The ratio analysis also firmly establishes that RM\_T\_180\_Roll takes 14% more time than the Mouse.

**Accuracy.** When selecting a sub-item in a Dropdown toolbar, the accuracy of the three Rolling-Menus is on average 95.6%, CI[94.1%; 97.5%], very similar to the accuracy of the Mouse (97.3%, CI[96.1%; 98.4%]). Results are detailed in Figure 13-Top.

When using a Grid toolbar the accuracy obtained with Rolling-Menus using the Tap validation reaches on average 84.2%, CI[76.5%; 88.0%]. Accuracy with the Rolling-Menu using a Roll validation technique (RM\_T\_180\_Roll) is below 60% while the Mouse shows an accuracy of 94.7%, CI[89.4%; 97.1%]). These results are further detailed in Figure 13-Bottom. The ratio analysis strongly confirm that the Roll version is less accurate than the Mouse. The intra-subject difference between the Mouse and the two others is also confirmed: the Mouse is on average 8.2% more accurate than RM\_T\_Dir\_Tap and 15.9% than RM\_B\_Dir\_Tap. Two reasons might explain this drop of accuracy: the gestures with the mouse (translation) and the RPM (roll) are different, and the user's expertise is optimized for the mouse but only at novice level with RPM.



**Figure 13.** Accuracy when selecting a sub-item in a Dropdown toolbar (top) and Grid toolbar (bottom). On the left, average accuracy in % and, on the right, accuracy ratio of the Mouse / Rolling-Menu (all bars represent 95% CIs).

**Complementary analysis.** We observed for both types of toolbars that selection time during the training phase was on average 12% longer than during the experiment. We found no evolution of selection time during the six series of the experimental phase. The training phase was thus sufficient for participants to familiarize with the Rolling-Menus.

We also observed that the results discussed above still apply when considering every predefined sub-item independently. Roll direction and amplitude do not affect the user's interaction efficiency.

#### Qualitative results

We computed the SUS score for each technique with every toolbar. Overall the usability of the Mouse is rated "excellent" (average SUS: 91.6). With the Dropdown toolbar, the usability of Rolling-Menus using a Tap validation method is "good" (75.5 on average). However the usability of Rolling-Menus using a Roll validation method (RM\_T\_180\_Roll) is "highly marginal" (average SUS: 61.0). With the grid toolbar, Rolling-Menu using a Tap validation method obtains a score of 67.8 on average and almost corresponds to a "good" acceptability, while the one with roll validation obtains 37.3, which is below an acceptable usability. We propose some perspectives to address this issue in the discussion section. As already mentioned, the roll validation requires a high precision, which might be annoying and thus justify the low score of the Rolling-Menu with Roll validation.

#### Summary

The results of this second experiment establish that, with Dropdown and Grid toolbars, two versions of Rolling-Menu are faster than the Mouse while offering a very good (Dropdown toolbar: 95.6%) or good accuracy (Grid toolbar: 84.2%). Rolling-Menus are thus an efficient solution for roll-based command selection in toolbars.

## DISCUSSIONS AND PERSPECTIVES

### Factors influencing Rolling-Menu performance

In this work, we established that the use of rolling gestures with a multi-DOF Mouse is an efficient solution for interacting with different types of toolbars. Through a comparison of eight Rolling-Menus, we determined that their efficiency is influenced by the combination of various design dimensions (activation, roll range and validation method), rather than by each design dimension on its own. In



particular three configurations emerge as being more efficient than the others for interaction with a simple toolbar (RM\_T\_Dir\_Tap, RM\_T\_180\_Roll, RM\_B\_Dir\_Tap).

Among them, those based on the Tap validation method outperform the one based on Roll validation in complex toolbars (dropdown and grid toolbars) containing 64 to 72 sub-items. The Tap validation method appears to be particularly well suited for use with Direct roll mapping, whatever the activation method (Top or Bottom) and the type of toolbar (simple, Dropdown or Grid). Roll validation method seems more appropriate for 180 roll range as the sector size is larger.

While Top activation requires longer gestures to activate the quasimode, we did not observe clear differences in selection time and accuracy with the Bottom activation method in our second study. Intuitively, the Top activation with Tap validation should be a good solution for newcomers, as it uses the widest part of the sectors, while experts could benefit from the shorter gestures using a Bottom activation. Further experiments focusing on the benefits of this version for experts are planned.

#### Rolling-Menus in existing applications

We confronted Rolling-Menu to different types of toolbars that correspond to concrete usages in common applications: simple toolbars with up to 14 items are for example used in Thunderbird (get mail, write, address, etc.); dropdown toolbars with up to 8x8 items are for instance used in Gimp; a grid-toolbar with up to 9x9 items corresponds to the tool palette in Photoshop. Rolling-Menu can therefore easily be introduced in everyday applications.

Toolbars in existing applications usually result from a combination of the three types of toolbar explored in our work: they include lists as in our Dropdown toolbar, buttons in a single row as in our Simple toolbar, or in multiple rows as in our Grid-toolbars. Since Rolling-Menu performs better than the Mouse in these different cases individually, it will theoretically still perform better on a more complex toolbar. Another difference with the toolbars of our studies is that toolbar items do not have necessarily the same width. In this case, the activation of our quasimode could momentarily spread the items over the width of the window to facilitate their selection by maximizing their size.

Further studies will focus on the use of Rolling-Menu in different contexts: vertical toolbars induce wrist biomechanical constraints that will probably affect left-right rolls differently than up-down rolls; Microsoft menu combine Tab-panes with Ribbons and thus requires to adapt and assess the use of Rolling-Menu; and finally a valuable follow-up will involve domain experts to compare memorized keyboard shortcuts to muscular memory with Rolling-Menu.

#### Technical limitations

The current implementations of Rolling-Menu are affected by the lack of resolution of the Inertial Measurement Unit. In

addition, the Wacom tablet detects residual translations during rolls, which disrupts the correct detection of the direction and length of translation gestures. These aspects induced limitations in the ability to detect accurate or compound gestures (i.e. roll + translate). For example, when selecting a sub-item in a grid toolbar (study 2), a RPM roll followed by a RPM translation are used to select and validate the appropriate sub-item: in this situation we observed that the Rolling-Menu accuracy decreased to 84%. We believe that adopting a more accurate technology would lead to better results and richer usages of Rolling-Menu: the mouse laser would be the ideal solution, but its adaptation to a spherical device remains challenging; the detection of RPM micro-rolls might be another promising solution.

Using a tap uni-manually on the underlying tablet, or bimanually on the keyboard, ensures a consistent interaction: the user is moving the mouse with one hand, and most often leaving the other hand on the keyboard to continue typing afterwards. However, the best solution would be a more integrated validation mechanism activated by the interactive hand. To this end we envision integrating a touch or tactile sensor on the RPM itself, instead of a button in our first tests. Such an approach has already been explored in TDome [23] but only in combination with physical gestures. The challenge in this case is how to place the sensor on the device without changing the grip, losing comfort or triggering accidental touch events.

Addressing these technical limitations will allow to carry longitudinal studies to evaluate the ergonomic aspects of RPM and Rolling-Menu.

#### CONCLUSION

In this work we presented different roll-based techniques for command selection in toolbars. The aim of these techniques is to minimize disruptive transitions between the working area and the toolbar. To do so, Rolling-Menus rely on the detection of roll gestures performed in the direction of the toolbar items to select. Based on different design dimensions, we proposed 8 versions of Rolling-Menu that we compared to the Mouse for selecting a top-level item in a Simple toolbar, or sub-items in more complex toolbars, i.e. a Dropdown and a Grid toolbar. Our user studies demonstrated that two Rolling-Menus reduce the selection time for a top-level item or sub-item, while keeping a good or very good accuracy.

#### ACKNOWLEDGMENTS

This work is partly funded by the AP2 project (research grant AP2 ANR-15-CE23-0001). Authors wish to thank Mathias Faure for his early design contributions to Rolling-Menu and the first implementations of the techniques.

#### REFERENCES

1. AVIZ Group. R Macros for data analysis. [www.aviz.fr/reliefshearing](http://www.aviz.fr/reliefshearing).
2. Bailly, G., Demeure, A., Lecolinet, E., and Nigay, L. MultiTouch menu (MTM). Proceedings of the 20th

- International Conference of the Association Francophone d'Interaction Homme-Machine on - IHM '08, ACM Press (2008), 165.  
DOI=<http://dx.doi.org/10.1145/1512714.1512746>
3. Bailly, G., Lecolinet, E., and Nigay, L. Visual Menu Techniques. *ACM Computing Surveys* 49, 4 (2016), 1–41.  
DOI: <https://doi.org/10.1145/3002171>
  4. Balakrishnan, R., Baude, T., Kurtenbach, G., and Fitzmaurice, G. The Rockin' Mouse: Integral 3D Manipulation on a Plane. *Proceedings ACM SIGCHI Conference on Human Factors in Computing Systems*, (1997), 311–318.  
DOI: <https://doi.org/10.1145/258549.258778>
  5. Bangor, A., Kortum, P.T., and Miller, J.T. An empirical evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction* 24, 6 (2008), 574–594.
  6. Dillon, R.F., Edey, J.D., and Tombaugh, J.W. Measuring the true cost of command selection: techniques and results. *Proceedings of the SIGCHI conference on Human factors in computing systems Empowering people - CHI '90*, April (1990), 19–26.  
DOI: <https://doi.org/10.1145/97243.97247>
  7. Fruchard, B., Lecolinet, E., and Chapuis, O. MarkPad: Augmenting Touchpads for Command Selection. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, ACM Press (2017), 5630–5642.  
DOI: <https://doi.org/10.1145/3025453.3025486>
  8. Giannakis, E., Bailly, G., Malacria, S., and Chevalier, F. IconHK: Using Toolbar button Icons to Communicate Keyboard Shortcuts. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, ACM Press (2017), 4715–4726.  
DOI: <https://doi.org/10.1145/3025453.3025595>
  9. Grossman, T., Dragicevic, P., and Balakrishnan, R. Strategies for accelerating on-line learning of hotkeys. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07*, ACM Press (2007), 1591.  
DOI: <https://doi.org/10.1145/1240624.1240865>
  10. Grossman, T., Hinckley, K., Baudisch, P., Agrawala, M., and Balakrishnan, R. Hover Widgets: Using the Tracking State to Extend the Capabilities of Pen-Operated Devices. *CHI 2006*, ACM Press (2006), 861–870.  
DOI: <https://doi.org/10.1145/1124772.1124898>
  11. Guimbretière, F. and Winograd, T. FlowMenu - Combining Command, Text, and Data Entry. *Proceedings of the 13th International Symposium on User Interface Software and Technology (UIST'00)* 2, (2000), 213–216.  
DOI: <https://doi.org/10.1145/354401.354778>
  12. Hascoët, M. Throwing models for large displays. *Proceedings of HCI2003, British HCI Group (2003)*, 73–77.
  13. Hinckley, K., Sinclair, M., Hanson, E., Szeliski, R., and Conway, M. The VideoMouse: : a camera-based multi-degree-of-freedom input device. *Proceedings of the 12th annual ACM symposium on User interface software and technology - UIST '99*, ACM Press (1999), 103–112.  
DOI: <https://doi.org/10.1145/320719.322591>
  14. Huot, S., Nancel, M., and Beaudouin-Lafon, M. PushMenu: Extending Marking Menus for Pressure-Enabled Input Devices, *Research Report LRI-RR1502*. Paris, 2008.
  15. Kim, S., Kim, H., Lee, B., Nam, T.-J., and Lee, W. Inflatable mouse: volume-adjustable mouse with air-pressure-sensitive input and haptic feedback. *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*, ACM Press (2008), 211–224.  
DOI: <https://doi.org/10.1145/1357054.1357090>
  16. Krzywinski, M. and Altman, N. Points of Significance: Error bars. *Nature Methods* 10, 10 (2013), 921–922.
  17. Perelman, G., Serrano, M., Raynal, M., Picard, C., Derras, M., and Dubois, E. The Roly-Poly Mouse: Designing a Rolling Input Device Unifying 2D and 3D Interaction. *Proceedings of the ACM CHI'15 Conference on Human Factors in Computing Systems 1*, (2015), 327–336.  
DOI: <https://doi.org/10.1145/2702123.2702244>
  18. Perelman, G., Serrano, M., Raynal, M., Picard, C., Derras, M., and Dubois, E. DECO: A Design Space for Device Composition. *Proceedings of the 2016 ACM Conference on Designing Interactive Systems - DIS '16*, ACM Press (2016), 435–446.  
DOI: <https://doi.org/10.1145/2901790.2901893>
  19. Raskin, J. and Jef. *The humane interface : new directions for designing interactive systems*. Addison-Wesley, 2000.
  20. Roudaut, A., Bailly, G., Lecolinet, E., and Nigay, L. Leaf Menus: Linear Menus with Stroke Shortcuts for Small Handheld Devices. *IFIP TC 13 International Conference on Human-Computer Interaction - Interact'99*, Springer, Berlin, Heidelberg (2009), 616–619.  
DOI: [https://doi.org/10.1007/978-3-642-03655-2\\_69](https://doi.org/10.1007/978-3-642-03655-2_69)
  21. Roudaut, A., Lecolinet, E., and Guiard, Y. MicroRolls: Expanding Touch-Screen Input Vocabulary by Distinguishing Rolls vs. Slides of the Thumb. *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, ACM Press (2009), 927.  
DOI: <https://doi.org/10.1145/1518701.1518843>

22. Roy, Q., Malacria, S., Guiard, Y., Lecolinet, E., and Eagan, J. Augmented letters. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13, ACM Press (2013), 2325. DOI: <https://doi.org/10.1145/2470654.2481321>
23. Houssein Saidi, Marcos Serrano, Pourang Irani, and Emmanuel Dubois. 2017. TDome: A Touch-Enabled 6DOF Interactive Device for Multi-Display Environments. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17). ACM, New York, NY, USA, 5892-5904. DOI: <https://doi.org/10.1145/3025453.3025661>.
24. Tian, F., Xu, L., Wang, H., et al. Tilt Menu : Using the 3D Orientation Information of Pen Devices to Extend the Selection Capability of Pen-based User Interfaces. Proceedings of ACM CHI 2008 Conference on Human Factors in Computing Systems, (2008), 1371–1380. DOI: <https://doi.org/10.1145/1357054.1357269>
25. VandenBos, G.R., ed. Publication Manual of the American Psychological Association (6th ed.). American Psychological Association, 2009.
26. Willett, W., Jenny, B., Isenberg, T., et al. Lightweight Relief Shearing for Enhanced Terrain Perception on Interactive Maps. Proceedings of the 33rd ACM Conference on Human Factors in Computing Systems (CHI 2015), (2015). DOI: <https://doi.org/10.1145/2702123.2702172>
27. Yang, X.-D., Mak, E., McCallum, D., Irani, P., Cao, X., and Izadi, S. LensMouse: augmenting the mouse with an interactive touch display. Proceedings of the 28th international conference on Human factors in computing systems - CHI '10, ACM Press (2010), 2431. DOI: <https://doi.org/10.1145/1753326.1753695>
28. Zhao, S., Agrawala, M., and Hinckley, K. Zone and polygon menus: Using Relative Position to Increase the Breadth of Multi-Stroke Marking Menus. Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI '06, ACM Press (2006), 1077. DOI: <https://doi.org/10.1145/1124772.1124933>
29. Zhao, S. and Balakrishnan, R. Simple vs. compound mark hierarchical marking menus. Proceedings of the 17th annual ACM symposium on User interface software and technology - UIST '04, ACM Press (2004), 33. DOI: <https://doi.org/10.1145/1029632.1029639>

# Identifying how Visually Impaired People Explore Raised-line Diagrams to Improve the Design of Touch Interfaces

Sandra Bardot<sup>1,2</sup>, Marcos Serrano<sup>1,2</sup>, Bernard Oriola<sup>2,1</sup>, Christophe Jouffrais<sup>2,1</sup>

<sup>1</sup> University of Toulouse – IRIT  
Toulouse, France  
{firstname.lastname}@irit.fr

<sup>2</sup> CNRS - IRIT  
Toulouse, France  
{firstname.lastname}@irit.fr

## ABSTRACT

Raised-line diagrams are widely used by visually impaired (VI) people to read maps, drawings or graphs. While previous work has identified general exploration strategies for raised-line drawings, we have limited knowledge on how this exploration is performed in detail and how it extends to other types of diagrams such as maps or graphs, frequently used in specialized schools. Such information can be crucial for the design of accessible interfaces on touchscreens. We conducted a study in which participants were asked to explore five types of raised-line diagrams (common drawings, perspective drawings, mathematical graphs, neighborhood maps, and geographical maps) while tracking both hands fingers. Relying on a first set of results, we proposed a set of design guidelines for touch interfaces.

## Author Keywords

Raised-line diagram; blind; tactile exploration; tactile maps; tactile drawings; bimanual exploration; finger tracking.

## ACM Classification Keywords

H.5.2. User Interfaces: User-centered design. Input devices and strategies,

## INTRODUCTION

Raised-line diagrams provide visually impaired (VI) people with access to graphs, drawings or maps [16]. Usually, they are hand-made because the content must be adapted and simplified. Moreover, they must be printed beforehand, which can be tedious and expensive in a teaching context. Then, it is pertinent to use digital versions of the diagrams that are accessible through adapted non-visual interactions.

Many approaches have been proposed, and can be summarized into three large categories. The first one consists in using refreshable displays that can dynamically raise up and down small pins providing a perceptible relief for manual exploration [20, 24]. A second category relies on tangible objects that represent important elements of a

drawing, and can be linked to each other to render lines and areas [18, 4]. The last category consists in using finger tracking devices (touchscreens, cameras, etc.) to follow finger movement over a digital map, and render the content with auditory or vibrational feedback [2, 6, 5, 10, 25]. However, the process of converting physical content to digital cues is performed empirically. Understanding how VI people explore physical raised-line diagrams, and which elements are of importance, could help to design non-visual tactile interfaces.

In psychology, several studies aimed to assess the capacity to identify raised-line diagrams. In general, these studies compared tactile exploration of sighted, early and late blind people who had to recognize drawings of common objects (car, fruit, tool, etc.). For instance, Heller et al. [1] compared the blind and sighted children exploring raised-line drawings, and showed that they reach the same performance when sighted are guided during exploration. Lebaz et al. [13] showed that the performance depended on the type of common drawing being used in the study (“flat” 2D drawings, or with 3D cues). However, these studies mainly relied on identification rates, and did not inform about the hand movements that were used.

Our goal was to understand the main role: 1) of hands (and fingers) during raised-line diagrams exploration, according to 2) diagram type, and 3) user expertise. We used different raised-line diagrams including drawings, mathematical graphs, and neighborhood or geographical maps. We developed an experimental setup to track the fingers of both hands during tactile exploration. We recruited 6 visually impaired and 6 sighted blindfolded subjects who explored the diagrams. The results were based on accurate tracking of the exploration movements, and highlight different exploration patterns concerning the movement of both hands and the covering of the diagrams. They showed that performance was significantly different according to the diagram types and user profiles.

## RELATED WORK

It has been shown that it is difficult to name tactile pictures for naive subjects [7]. It has also been shown that tactile recognition of drawings depends on previous visual experience. However, studies comparing tactile recognition rate of sighted, early and late blind people showed that blind people can either perform better (see e.g. [7]) or worse (see e.g. [14, 17]) than sighted people. In fact, it

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).  
CHI 2017, May 06–11, 2017, Denver, CO, USA  
© 2017 ACM. ISBN 978-1-4503-4655-9/17/05...\$15.00  
DOI: <http://dx.doi.org/10.1145/3025453.3025582>

seems that the difficulty of picture recognition varies with complexity, familiarity, and categorical information [8, 9].

The procedures used to explore real 3D objects have an impact on the type and quality of information available [15], and affect the performance at haptic recognition tasks. Hence exploration procedures, but also the more general strategies combining these procedures, probably depend on the intended goal of the task. For instance, one might want to learn quickly as much as possible about an object, or alternatively test some hypothesis about that object. These general principles probably apply to the tactile exploration of 2D raised-line diagrams too. Recent studies depicted specific hand movements during the exploration of raised-lines diagrams with blindfolded participants [22]. They showed that, most of the time, subjects used their index finger(s), either alone or in combination with other fingers. Although subjects were usually unaware of how they moved their hands, the movements were both purposive and systematic [22]. Another study showed that both hands were used more than 83% of the time, which significantly increased identification of raised line drawings [23]. In fact these patterns of hand movement have been called exploratory procedures [19]. They include lateral motion (moving the fingers back and forth across a texture or feature), contour following (tracing an edge within the image), and whole-hand exploration of global shape [19]. However these different studies did not systematically depict exploratory procedures according to the drawing being explored or the task being performed [12].

Tactile exploration of a digital drawing on a touch-screen display is even more difficult [11]. Guerreiro et al. [6] focused on the observation of bimanual exploration (one finger of each hand) on large touchscreens during four specific tasks (locate, relocate, count, relate). They identified seven features and four strategies that appeared when visually impaired subjects were exploring drawings that resemble geographical maps with a few landmarks only. This work clearly showed that specific strategies appear in relation to the tasks being performed. However, the study relied on one map-like diagram only and did not consider other types of diagrams (e.g. common drawing or mathematical graph).

In the current study, we used a setup made of a touchscreen and a camera which allowed for accurate tracking of hands and fingers movements. Our method extends previous work [3] based on a depth-camera that could not accurately separate the fingers from the surface during tactile exploration. Our setup leverages quantifying the hand movements involved during the tactile exploration of different types of raised-line diagrams.

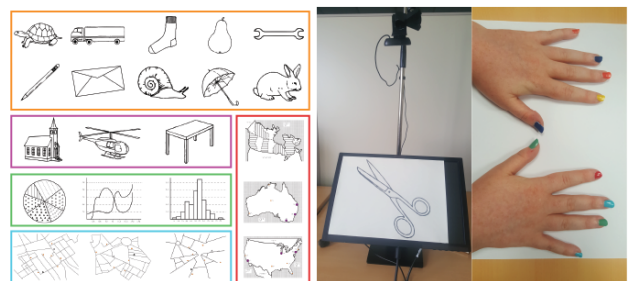
#### STUDY: EXPLORATION STRATEGIES

The goal of our study was to understand the role of each hand during the tactile exploration of different types of diagrams (drawings, graphs or maps) by users with

different expertise (sighted blind-folded or visually impaired).

#### Raised-line diagrams

In psychology research, most of the studies concerning tactile exploration relied on common drawings [13, 21, 23]. But common drawings are not the most used diagrams by VI people. Indeed, in specialized schools, students have geography, math, and locomotion lessons, which rely on raised-line charts and maps. Hence, we selected five different types of diagrams (Figure 1). All the diagrams were made with the assistance of a professional tactile document maker. We also performed several iterations with VI users to refine height and width of the raised lines, as well as the legibility of Braille text. We designed three diagrams of each type, except for the Common Drawings (we used 10 of them, see task below). All the diagrams were printed in A3 landscape format on Zytech swell paper.



**Figure 1. Diagrams used (left) and experimental setup (right): a raised-line drawing is placed over the touchscreen. A camera located above the drawing tracks fingers movements.**

*Common Drawings (C-Drawings) and Perspective Drawings (P-Drawings):* they were issued from the Snodgrass and Vanderwart set of images [21]. Selected C-Drawings were: scissors, envelope, sock, open-end wrench, pencil, umbrella, truck, snail, turtle, rabbit, and pear. P-Drawings represented a couch, church, table, and helicopter. They included perspective cues.

*Mathematical graphs (Graphs):* We used graphs frequently used in specialized schools, i.e. histograms and plots.

*Neighborhood maps (N-Maps):* Delimited zone of a city. We added two itinerary points (starting and ending) as well as shops, represented by an empty triangle, a solid triangle, and filled circles, respectively.

*Geographical maps (G-Maps):* Two types of maps were included, representing a country with either a few main cities or the regions within the country. The cities were represented by solid points, with the first two letters written in Braille. Borders between regions were represented by dotted lines. A few regions included a different texture that emphasized a specific element. Seas and oceans were represented with a specific texture.

#### Tasks and instructions

During the study, each participant explored 22 diagrams: 10 Common Drawings, and 3 drawings of the four other types.

The study was divided in two steps. The first step consisted in identifying 10 C-Drawings as fast as possible (max time allowed: 90s). The drawing category was mentioned beforehand (e.g. object or animal). We measured the time needed to identify the drawing. The goal was to assess the expertise level of each participant in terms of raised-line diagram exploration. The second step consisted in the exploration of the four other types of diagrams by blocks of 3 trials (total of 12 trials). First, the participants explored a diagram for 30s (*free exploration*). Then, they had another 60s to explore the same diagram in order to answer a question (*driven exploration*). The questions varied according to the type of diagram: for P-Drawings, they had to identify the object among 4 choices. For Graphs, they had to find specific min and max values. For N-Maps, they had to find the number of stores between the starting and ending points. For G-Maps, they had to compare different regions. No instructions were given concerning the use of one or two hands.

### Participants

We recruited two groups of participants: 6 sighted subjects who were blind-folded (BF, 3 females), and 6 visually impaired subjects (VI, 5 females). These two groups should show contrasted results because BFs are non-experts of tactile exploration but can rely on previous visual knowledge to identify drawings. In contrast, VIs are experts of tactile exploration but with very few, if any, previous visual knowledge. BFs were 2 university students and 4 staff members aged 27 on average (SD=2). VIs were 5 teachers and 1 radio-program presenter aged 46 on average (SD=14). Among them, 5 were early blind and 1 had very limited residual vision (light perception). She was blindfolded during the study.

Before the experiment, we conducted an interview to assess proficiency in braille reading and raised-line diagrams exploration (on a 5-points Likert scale). All VI participants rated their expertise between 2 and 5 (M=4.5) for Braille reading and between 1 and 4 (M=3) for tactile exploration. They all used their left hand as the main reading hand but differently: 1 read with left hand only, 5 with both hands. The 5 VI teachers explore raised-line diagrams several times per week. BFs had no prior experience with raised-lines diagrams.

### Design and procedure

Our study followed a within-participants design, with one factor: Diagram type (C-Drawings, P-Drawings, Graphs, N-Maps, and G-Maps). The order of the last 4 blocks (1 block for each diagram type) was counterbalanced across participants. Within each block, the order of the 3 trials was random. Users were free to take a break between blocks.

### Experimental setup

The subject was comfortably sitting in front of the tactile drawing placed over a 22-inch (1680x1050px) multi-touch screen (Fig. 1). We used a Logitech C270 webcam (1280x720 px) located above the touch screen in order to

track the ten fingers according to colored markers placed on each nail (Fig. 1). The acquisition rates were 50 Hz for the camera and 100 Hz for the touchscreen.

### Collected data

We collected the coordinates of the 10 fingers, as well as the touch status. We also measured the exploration time needed to answer the questions. In addition, subjects had to rate the difficulty for each type of diagram. At the end of the session, we asked them whether they were aware of using any specific exploration strategy. We also collected their subjective feeling about the number of hands and fingers they used, and why for.

### RESULTS

We computed a Univariate ANOVA with a Bonferroni Pairwise post-hoc test to compare the results.

#### Exploration times and accuracy

We found main effects of diagram type ( $F_{4,44}=6.9$ ,  $p<.001$ ) and user group ( $F_{1,11}=8.1$ ,  $p=.005$ ) on exploration times. The averaged exploration times were 45.9s for BF participants and 39.4s for VI participants. Post-hoc comparison revealed a significant difference ( $p<.001$ ) between C-Drawings (M=44.7 s) and P-Drawings (M=27.5 s) exploration times.

Concerning accuracy, we found a main effect of diagram type ( $F_{4,44}=8.7$ ,  $p<.001$ ) and an interaction between diagram type and user group ( $F_{1,11}=8.7$ ,  $p=.006$ ). Post-hoc comparison showed a significant difference between C-Drawings and N-Maps ( $p=.002$ ) and C-Drawings and P-Drawings ( $p<.001$ ). The accuracy per diagram type was (BF vs. VI participants): C-Drawings (58.6 vs. 41.3%); G-Maps (27.7 vs. 70.5%); Graphs (62.5 vs. 75%); N-Maps (11 vs. 31%); and P-Drawings (93.7 vs. 5%).

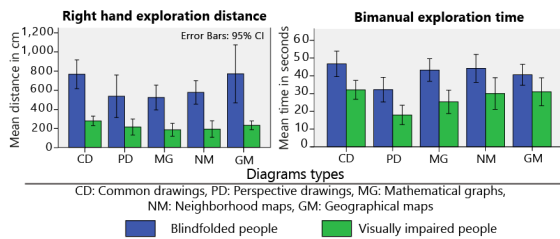
#### Diagram covering

We measured the covering for each type of diagram, i.e. the percentage of diagram that was explored. We found a main effect of diagram type ( $F_{4,44}=4.5$ ,  $p=.003$ ) and user group ( $F_{1,11}=9.1$ ,  $p=.003$ ) on covering. Post-hoc comparison showed a significant difference between C-Drawings on one side, and N-Maps ( $p=.002$ ), G-Maps ( $p=.002$ ), and Graphs ( $p=.01$ ) on the other side. The averaged diagram covering was 65.7% for BF and 52.2% for VI participants. The covering per diagram type (BF vs. VI participants) were: C-Drawings (74.2 vs. 59.7%); G-Maps (51.6 vs. 44.1%); Graphs (52.3 vs. 51.1%); N-Maps (55.4 vs. 40.6%); and P-Drawings (75.7 vs. 46.1%).

#### Exploration distance per diagram and per hand

We measured the total exploration distance for each hand and diagram (Fig. 2 Left). We found an effect of the user group on the distance covered by the right ( $F_{1,11}=61$ ,  $p<.001$ ) and left hands ( $F_{1,11}=17.5$ ,  $p<.001$ ). For BF participants, the exploration distances were 460 and 727 cm for the left and right hand respectively. For VI participants, the exploration distances were 769 and 257 cm for the left and right hand respectively.





**Figure 2. Left: Mean exploration distance for the right hand. Right: Time when both hands moved simultaneously (bimanual exploration).**

### Bimanual exploration

For each trial, we also computed the time during which both hands were moving simultaneously (bimanual exploration time; see Fig 2). We found an effect of diagram type ( $F_{4,44}=3.8$ ,  $p=.004$ ) and user group ( $F_{1,11}=27.5$ ,  $p<.001$ ). The bimanual exploration times were 43s (83.6%) and 28.7s (74.6%) for BF and VI participants respectively.

### DISCUSSION

These results showed that VI subjects were faster than BF subjects when considering all the diagram types. They are similar to the results from [7], and probably reflect the greater expertise of VI subjects. However, another result is striking. Although BF and VI subjects reached the same overall identification performance (52%), the percentage of correct responses was significantly different according to the type of drawing. For instance, VI subjects reached 31% of correct responses on N-Maps whereas BF subjects only reached 11% correctness. In fact, VI subjects were better at exploring G-Maps and N-Maps. On the contrary, BF subjects were better at identifying C-Drawings and P-Drawings. This result confirms that exploration of tactile drawing depends on complexity, familiarity, and categorical information [8,9], but also shows that recognition depends on the type of drawing being explored, as well as the expertise of the user. It is probable that VI adults more frequently explore maps and mathematical diagrams than drawings of objects. In addition, drawings rely on visual conventions (occlusions and perspectives) that are less significant to VI people [7].

Our results also highlighted that VI participants covered a smaller exploration distance than BF for all diagram types (52% vs. 66%), although they got equivalent success rates. We observed that VI subjects focused on salient areas of the diagrams (such as the ears on the Rabbit drawing) to identify the diagram type, which made their exploration more efficient. Overall, Drawings required a more extensive exploration than Maps and Graphs.

Exploration distance per hand highlighted that VI and BL subjects had opposite hand behaviors: VI mainly used left hand (769 cm per trial), which is their braille reading dominant hand, while BL subjects used mostly right hand (727 cm). Interestingly, VI subjects covered a limited distance with right hand (257 cm per trial), and BF subjects spent more time performing bimanual exploration. This

observation probably reflects that VI subjects appropriately use the second hand as an anchor that helps to understand the drawing, which is a valuable exploratory procedure.

### DESIGNING INTERACTIVE ACCESSIBLE DRAWINGS

Our results confirm that VI people are able to explore tactile displays effectively but, depending on their own expertise or the type of diagram being explored, they may benefit from instructions or guidance. According to the format (A3) of the diagrams used in the current study, this preliminary work can provide general design guidelines for tactile displays larger than smartphone or tablet screens.

*Touch robustness:* Because participants laid their hand on the surface, we frequently observed more than ten simultaneous touch events (16 for BF and 17 for VI), corresponding to additional contacts with the palm. Then, although touch interfaces should enable more than two fingers for tactile exploration, they must prevent unexpected touch events, for instance by combining touch and camera tracking.

*Content Simplification:* A substantial amount of the raised-lines are not used for completing the different tasks. Then tactile drawings could be further simplified, but in a way that is specific to each type of drawing.

*Multimodal Information Sharing:* Common drawings require thorough exploration for identification. Then, interfaces may provide both contextual and local feedback (e.g. “rabbit” and “ears”). N-Maps raised exploration issues related to locating specific points. Then, interfaces may provide hand guidance cues (e.g. vibrotactile cues). In general, additional cues (e.g. sound or vibratory pattern) may highlight salient regions of the diagrams to facilitate exploration.

*Interaction Menus:* VI subjects mainly use their dominant hand for exploration. The non-dominant hand often stays steady, as an anchor. Hence, validation or selection commands could be assigned to the non-dominant hand. These commands should be contextual.

### CONCLUSION AND FUTURE WORK

This preliminary study, as well as the general design guidelines that we provide, should be extended. The method can be used in the field of experimental psychology to better understand the role of each finger during tactile exploration. The method can also be used to address specific design questions according to the types of diagrams, tasks, and targeted users.

### ACKNOWLEDGMENTS

We thank all the participants, “Cherchons pour Voir” lab, and Lestrade specialized institute (Toulouse, FR). This work was granted by AccessiMap ANR-14-CE17-0018.

## REFERENCES

1. D'Angiulli, A., Kennedy, J. M., & Heller, M. A. 1998. Blind children recognizing tactile pictures respond like sighted children given guidance in exploration. *Scand.J.Psychol.*, 39(3), 187–190.
2. Sandra Bardot, Marcos Serrano, and Christophe Jouffrais. 2016. From tactile to virtual: using a smartwatch to improve spatial map exploration for visually impaired users. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '16)*. ACM, New York, NY, USA, 100-111. DOI: <http://dx.doi.org/10.1145/2935334.2935342>
3. Anke Brock, Samuel Lebaz, Bernard Oriola, Delphine Picard, Christophe Jouffrais, and Philippe Truillet. 2012. Kin'touch: understanding how visually impaired people explore tactile maps. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12)*. ACM, New York, NY, USA, 2471-2476. DOI: <http://dx.doi.org/10.1145/2212776.2223821>
4. Julie Ducasse, Marc J-M Macé, Marcos Serrano, and Christophe Jouffrais. 2016. Tangible Reels: Construction and Exploration of Tangible Maps by Visually Impaired Users. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2186-2197. DOI: <http://dx.doi.org/10.1145/2858036.2858058>
5. Nicholas A. Giudice, Hari Prasath Palani, Eric Brenner, and Kevin M. Kramer. 2012. Learning non-visual graphical information using a touch-based vibro-audio interface. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility (ASSETS '12)*. ACM, New York, NY, USA, 103-110. DOI=<http://dx.doi.org/10.1145/2384916.2384935>
6. Tiago Guerreiro, Kyle Montague, João Guerreiro, Rafael Nunes, Hugo Nicolau, and Daniel J.V. Gonçalves. 2015. Blind People Interacting with Large Touch Surfaces: Strategies for One-handed and Two-handed Exploration. In *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces (ITS '15)*. ACM, New York, NY, USA, 25-34. DOI: <http://dx.doi.org/10.1145/2817721.2817743>
7. Heller MA. 1989. Picture and pattern perception in the sighted and blind: The advantage of the late blind. *Perception*;18:379.
8. Heller MA, Calcaterra JA, Burson LL, Tyler LA. 1996. Tactual picture identification by blind and sighted people: Effects of providing categorical information. *Percept Psychophys*;58.
9. Heller, M. A. 2002. Tactile picture perception in sighted and blind people. *Behavioural Brain Research*, 135(1–2), 65–68. Journal Article. <http://doi.org/Pm:12356435>
10. Shaun K. Kane, Meredith Ringel Morris, Annuska Z. Perkins, Daniel Wigdor, Richard E. Ladner, and Jacob O. Wobbrock. 2011. Access overlays: improving non-visual access to large touch screens for blind users. In *Proceedings of the 24th annual ACM symposium on User interface software and technology (UIST '11)*. ACM, New York, NY, USA, 273-282. DOI=<http://dx.doi.org/10.1145/2047196.2047232>
11. Klatzky, R.L., Giudice, N.A., Bennett, C.R., & Loomis, J.M. 2014. Touch-Screen Technology for the Dynamic Display of 2D Spatial Information without Vision: Promise and progress. *Multisensory Research*. 27(5-6), 359-378.
12. Klatzky R L, Lederman S J. 1987. The Intelligent Hand", chapter in *The Psychology of Learning and Motivation* Ed. G Bower (New York: Academic Press) pp. 121-151
13. Lebaz, S., Jouffrais, C., & Picard, D. 2012. Haptic identification of raised-line drawings: high visuospatial imagers outperform low visuospatial imagers. *Psychological Research*, 76(5), 667–675. <http://doi.org/10.1007/s00426-011-0351-6>
14. Lederman SJ, Klatzky RL, Chataway C, Summers CD. 1990. Visual mediation and the haptic recognition of two-dimensional pictures of common objects. *Percept Psychophys*. Jan;47(1):54-64.
15. Lederman, S. J., & Klatzky, R. L. 1993. Extracting object properties through haptic exploration. *Acta Psychol (Amst)*, 84(1), 29–40.
16. AK Lobben. 2005. Identifying the needs of tactile map makers. *Proceedings of XXII International Cartographic Conference*.
17. Jack M Loomis, Roberta L Klatzky and Susan J Lederman. 1991. Similarity of Tactual and Visual Picture Recognition with Limited Field of View. In *Perception* vol. 20 no. 2 167-177. DOI: 10.1068/p200167
18. David McGoekin, Euan Robertson, and Stephen Brewster. 2010. Clutching at straws: using tangible interaction to provide non-visual access to graphs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 1715-1724. DOI=<http://dx.doi.org/10.1145/1753326.1753583>
19. O'Modhrain, S., Giudice, N. A., Gardner, J. A., & Legge, G. E. 2015. Designing media for visually-impaired users of refreshable touch displays: Possibilities and pitfalls. *Transactions on Haptics*. 8(3), 248-257.
20. Bernhard Schmitz and Thomas Ertl. 2012. Interactively Displaying Maps on a Tactile Graphics Display.

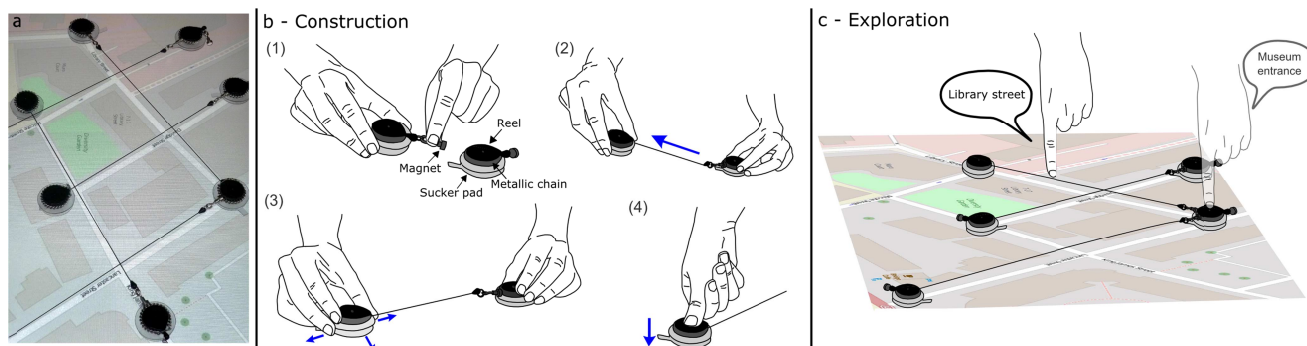
- Spatial Knowledge Acquisition with Limited Information Displays* (SKALID 2012), 13–18.
21. Snodgrass J, Vanderwart M, 1980. A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6: 174-215.
  22. Symmons M, Richardson B. 2000. Raised line drawings are spontaneously explored with a single finger. *Perception* 29(5):621-626.
  23. Wijntjes MW, van Lienen T, Verstijnen IM, Kappers AM. 2008. The influence of picture size on recognition and exploratory behaviour in raised-line drawings. *Perception*;37(4):602-14.
  24. Limin Zeng and Gerhard Weber. 2012. ATMap: Annotated Tactile Maps for the Visually Impaired. *International Training School, Cognitive Behavioural Systems (COST 2102)*, LNCS Volume 7403, Springer Berlin Heidelberg, 29.
  25. Zhao, H., Plaisant, C., Shneiderman, B., and Lazar, J. 2008. Data Sonification for Users with Visual Impairment. *ACM Transactions on Computer-Human Interaction* 15, 1: 1–28.

# Tangible Reels: Construction and Exploration of Tangible Maps by Visually Impaired Users

Julie Ducasse<sup>1</sup>, Marc Macé<sup>2</sup>, Marcos Serrano<sup>1</sup>, Christophe Jouffrais<sup>2</sup>

<sup>1</sup> University of Toulouse – IRIT  
Toulouse, France  
{first\_name.last\_name}@irit.fr

<sup>2</sup> CNRS - IRIT  
Toulouse, France  
{first\_name.last\_name}@irit.fr



**Figure 1.** a) The interface enables a visually impaired user to render digital maps tangible by using Tangible Reels. b) The user is guided step by step to correctly place each Tangible Reel. To construct a line, the user attaches two Tangible Reels (b1), follows rough (b2) and fine (b3) audio guidance instructions, and presses the sucker pad (b4). c) The user can explore the tangible map, receiving auditory feedback when pointing at lines and pads.

## ABSTRACT

Maps are essential in everyday life, but inherently inaccessible to visually impaired users. They must be transcribed to non-editable tactile graphics, or rendered on very expensive shape changing displays. To tackle these issues, we developed a tangible tabletop interface that enables visually impaired users to build tangible maps on their own, using a new type of physical icon called Tangible Reels. Tangible Reels are composed of a sucker pad that ensures stability, with a retractable reel that renders digital lines tangible. In order to construct a map, audio instructions guide the user to precisely place Tangible Reels onto the table and create links between them. During subsequent exploration, the device provides the names of the points and lines that the user touches. A pre-study confirmed that Tangible Reels are stable and easy to manipulate, and that visually impaired users can understand maps that are built with them. A follow-up experiment validated that the designed system, including non-visual interactions, enables visually impaired participants to quickly build and explore maps of various complexities.

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. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).  
CHI'16, May 07 - 12, 2016, San Jose, CA, USA  
Copyright is held by the owner/author(s). Publication rights licensed to ACM.  
ACM978-1-4503-3362-7/16/05...\$15.00  
DOI: <http://dx.doi.org/10.1145/2858036.2858058>

## Author Keywords

Non-visual tangible interfaces; interactive graphics;

## ACM Classification Keywords

H.5.2. User interfaces: interaction styles.

## INTRODUCTION

In educational centers for blind and visually impaired people, graphical representations such as charts, schemas and maps are widely used in STEM fields (Science, Technology, Engineering and Mathematics), but also during orientation and mobility lessons [16]. As visual graphics are inherently inaccessible to visually impaired users, they must be transcribed into tactile graphics. In this study we observed usages in a specialized Institute: visual graphics are transcribed into raised-line, hand-crafted or magnetic graphics. The production of the first two types relies on a time-consuming process that needs to be done by a tactile graphics specialist [30], and results in non-editable maps [10]. Magnetic maps are easier to produce and more adjustable, but they cannot be constructed or accessed by visually impaired users without assistance.

These issues can be alleviated by automating the production of graphics or maps (see [30]) and/or making them interactive. However interactive graphics are also limited: audio-tactile graphics cannot be edited [2], and acoustic and/or haptic graphics (see [31]) usually enable exploration based on a single point of contact only. Shape-changing displays are very promising but are still extremely expensive [29].



We aimed to overcome these limitations by providing visually impaired users with a way to construct physical representations of visual graphics on their own. Tangible tabletop interfaces are, in this sense, particularly relevant, as they provide a way to interact with digital information via the manipulation of phicons (i.e. physical icons, as defined in [9]), and therefore also provide a way to translate digital information into a physical form. A tangible interface for the exploration of graphs by visually impaired people has previously been implemented [18]. Although it provided a limited accessibility to mathematical plots with a limited number of points, it was a rousing first step towards non-visual tangible interfaces. We intended to design a low cost device that is accessible to visually impaired users without any assistance, and that allows them to render tangible any digital graphical content that is composed of points and lines. It should complement existing tools (e.g. raised-line maps) because it allows dynamic construction and exploration of graphics (with the benefit of “learning-by-doing”), without the restriction of a single point of contact exploration. In this study, we designed and constructed a novel type of physical icon called Tangible Reels. A Tangible Reel is composed of a sucker pad that ensures stability, and of a retractable reel that is used to create physical links between phicons. We also developed a non-visual tabletop interface that enables a visually impaired user to build a tangible map using Tangible Reels and to interact with it (see Figure 1). During the construction, audio guidance is provided so that the user can place the phicons correctly. As each phicon can be linked to another with the string of a retractable reel, lines can be constructed between two Tangible Reels. During tactile exploration, the user can listen to the name of the digital points and lines by performing a tap and hold gesture on their physical representations.

This paper describes three main contributions: 1) we report on observations of the use and production of tactile maps in a specialized educational Institute; 2) we describe the design process of Tangible Reels, and report on the evaluation conducted to check that Tangible Reels are stable, easy to manipulate, and that the physical representations built with them can be understood by a visually impaired user; 3) we introduce a novel interface that enables a visually impaired user to independently make digital maps tangible. We report on its evaluation, consisting in constructing and exploring four maps of increasing complexity.

#### OBSERVATION: USAGE OF TACTILE MAPS

The term “graphics” brings together a variety of materials whose spatial layout is used to provide content or data to the reader: diagram, figure, drawing, map, etc. [27] Maps are particularly interesting as they are essential in our day-to-day life, but also because they require the reader to understand and mentally integrate the spatial layout of the elements. The various types of maps (metro map, overview map, itineraries, etc.) clearly illustrate how diverse the

complexity of tactile graphics can be. In this study, we chose to focus on the construction and exploration of maps.

To better understand how and why tactile maps are used in specialized educational centers, we observed two geography lessons with four students, one orientation and mobility (O&M) lesson with one student, and interviewed a geography teacher and two O&M instructors in the CESDV-IJA Institute in Toulouse, France. This Institute provides education and training for around 120 visually impaired people, from early childhood to later adulthood. We identified three types of maps: raised-line, magnetic and hand-crafted.



**Figure 2. a) A raised-line map. b) A wooden model used for orientation and mobility lessons. c) Magnets are used to represent itineraries.**

Raised-line maps are mainly used by geography teachers. In this Institute, they are made with thermoforming or swell-paper techniques but other techniques exist [4]. Thermoforming consists in placing a sheet of plastic upon a hand-crafted mold made of different textured papers. When heated in a vacuum, the plastic sheet is shaped by the mold and permanently deformed. Swell-paper maps are printed on a special heat-sensitive paper that swells when heated and creates relief.

Each teacher owns different maps whose content may differ depending on the age, ability and visual impairment of the students. During a lesson, teachers often need to use different maps as the amount of information that can be displayed on a tactile map is limited, but also because they usually do not include more than five different textures on a given map. This number can be increased when the students are progressively introduced to new textures. However, since a tactile map cannot be readily edited, several versions of the same map are used to introduce a limited set of textures (1-3) each time. In the Institute, four transcribers produce around 1200 tactile documents each year.

Magnets are more often used by O&M teachers to help a student learn an itinerary (see Figure 2c). To prepare a new itinerary for the students, the teacher builds a simplified representation on a magnetic board: every time a new magnet is placed on the board, the teacher indicates what it stands for and lets the student touch it. After walking the route, the students may be asked to rebuild the route so that the teacher can check whether it has been understood. Magnets are also used to represent complex spatial configurations (road crossing or configuration of a building for example).

We also observed a number of hand-crafted graphics made out of rope, wood, felt or cardboard. They represent geometrical concepts (shapes, open/close features, parallel/perpendicular) and O&M elements (different types of crossings and a map of the school neighborhood for example).

These observations highlight one major issue: when a map is required, it has to be “materialized” with the assistance of a sighted person, which is a time consuming process. The students cannot access digital maps immediately and independently. Furthermore, the maps that are produced this way are not interactive and not editable, which limits the way users can interact with them and hence their autonomy.

To be fully accessible for visually impaired users, maps should be available without assistance and instantly. In addition, they should be interactive and editable, so that they could support dynamic operations such as zooming, panning, annotations, as well as other advanced functions (e.g. computing distances).

#### **RELATED WORK ON INTERACTIVE MAPS**

To alleviate the aforementioned issues, different approaches relying on new technologies have been used. Zeng and Weber [31] classified the different types of interactive maps in four categories, depending on the device and interaction used. Virtual acoustic maps use verbal and non-verbal audio output to render geographical data. For example, Zhao et al. [33] presented thematic maps explored with a keyboard or a tablet and producing string pitch and spatial sound. Virtual tactile (or haptic) maps most often rely on a force-feedback device. For instance, SeaTouch [26] enables visually impaired users to explore a maritime environment relying on haptic feedback, sonification and speech synthesis. The BATS [20] or HapticRiaMaps [11] are other examples of virtual tactile map using force-feedback devices. TouchOver map [22] provides visually impaired users with a basic overview of a map layout displayed on a mobile device through vibrational and vocal feedbacks. Kane et al. [13] described complementary non-visual interaction techniques that allow finding a target on large touch screens. Audio-tactile maps consist in a raised-line paper map placed over a touch-sensitive screen that provides audio descriptions of tactile elements when touched (see [19] and [30]). In contrast to virtual acoustic and tactile maps, these maps provide multiple points of contact (potentially all the fingers), and proved to be usable for learning spatial configurations [2]. Finally, Braille maps displayed on refreshable displays are a promising approach to create interactive maps. Zeng et al. [32] used a BrailleDis 9000 tablet device consisting in a matrix of 120x60 pins that can be moved up or down. Their prototype allowed visually impaired users to explore, annotate and zoom in or out. Similarly, Schmitz and Ertl [23] used a HyperBraille to present different types of maps representing buildings or large outdoor areas. The main drawback of the virtual maps

is that they provide a single point of contact (e.g. a phantom device), which forces the user to explore the map sequentially, and mentally integrate a large amount of information through space and time. However, they do not require a raised-line map overlay and can theoretically allow panning, zooming, and dynamic updating. Refreshable displays can provide both multiple fingers exploration as well as dynamic update, but these devices are extremely expensive, and hence relatively unusual.

Tangible maps for the visually impaired may present a number of advantages: they could be built autonomously using appropriate feedback, which may support learning-by-doing activities, provide multiple fingers exploration and allow dynamic updating while being affordable.

#### **Towards tangible maps**

A number of tangible user interfaces have been developed to enable sighted users to interact with a map. GeoSpace [9] is an interactive map onto which objects are placed. Their location modifies the digital map position and extent. Urp [28] allows urban planners to simulate wind flow and sunlight, and to observe their consequences on physical building models placed onto the tabletop. With the MouseHouse Table [8], users can model several arrangements of urban elements such as streets and buildings by placing paper rectangles on the table and visualizing the behavior of pedestrians.

Two devices have been specifically designed for visually impaired users. The Tangible Pathfinder [25] allows them to construct a map using small objects that represent pavements, sidewalks, etc. Audio instructions and feedback assist the user in placing the objects and exploring the map. Schneider et al. [24] designed a prototype for route construction by telling the user the length of building blocks and where to place them on a magnetic board. These devices are devoted to route or neighborhood exploration, and can hardly be adapted to other types of graphical content. In addition, to our knowledge, they have not been formally evaluated, and the construction of a tangible map by a visually impaired user on its own has not yet been demonstrated. In this study, we designed and evaluated a tabletop tangible interface that enables a visually impaired user to construct and explore different types of maps, with different levels of complexity.

The current work is in line with two other research projects [17][18]. In [17] the device provides visually impaired users with multimodal feedback to accurately place objects (called TIMMs) in order to create and modify graphs and diagrams. The authors suggest that a tactile line could be added between two TIMMs with a piece of yarn for example, but did not indicate how the user would select the correct length and could interact with the line. In [18], the device allows the exploration of line graphs and bar charts. Phicons are placed in a restricted physical grid (9x7 cells) in order to represent the top of a bar or the turning point of a linear function. Relying on an evaluation with four users,



the authors observed that the objects were regularly knocked over during the exploration. Hence they provide a few recommendations concerning the design of phicons for visually impaired users.

This review of the literature highlights three important points. First, although tangible user interfaces are promising to address some limitations of interactive tactile maps, research projects about autonomous construction of maps by visually impaired users are seldom. Second, the last two prototypes described above have only focused on rendering punctual symbols tangible (i.e. physical and associated to a digital content). Obviously, it is essential to make lines tangible as well because they are mandatory in maps and other graphical representations. Third, the question of how to design phicons that are stable and reliable has not been fully addressed yet. The magnetic board used in [24] is stable but may be unreliable. The objects are tracked by a camera placed above the tabletop, whose view may be occluded by users' hands. McGookin et al. [18] used a physical grid that holds the objects, which limits their potential location, but still observed unintentional knocks. In addition, such a solution is not relevant for the construction of most graphics that include lines and not only points.

### MAKING DIGITAL MAPS TANGIBLE

In this section, we describe the characteristics of graphics that should be made with the system, the design of the Tangible Reels (they are tracked by the tabletop, and physically linked to each other to represent digital points and lines), and the interaction techniques designed to construct and explore maps.

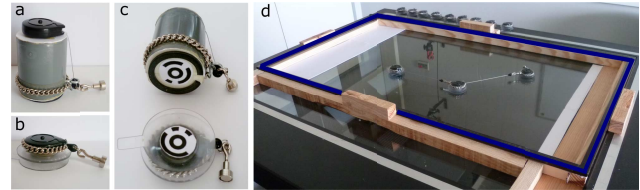
#### Digital and physical maps

First of all, it is important to know that simplification of graphics is mandatory for tactile exploration [4]. Hence, fine details of an outline must be removed during this adaptation process and curves are most often straightened. In our system, maps are defined by a set of points and lines using a simple syntax: each point is associated with an ID, a name and two coordinates; each line is associated with an ID, a name and the IDs of starting and ending points. The system does not aim at the construction of detailed maps (such as mobility maps), but rather at the construction of sparse spatial configurations such as overview or metro maps.

#### Design of the Tangible Reels

To build the physical representation of a digital map, it was essential to provide the user with a way to draw lines of different lengths. Retractable reels appeared to be an appropriate tool as their strings can be pulled out to different lengths. This is more usable than using material of different lengths (e.g. pre-cut wool yarn such as in [17]) and limits the number of required steps and therefore the likelihood of errors. Retractable reels have already been used in the field of HCI research as input devices (see e.g. [14][1][21]).

According to [18], it was also essential to firmly set the objects to the table so that a visually impaired user could explore the map without moving them or knocking them over. This was even more important in our case as a tension was applied to the objects by the retractable reels. We identified three requirements for the design of the Tangible Reels. They had to be: easy to move during construction but stable during exploration; identifiable and trackable by the system; as small as possible to maximize the number of objects that can be placed onto the tabletop.



**Figure 3: Weights (a) and Sucker pads (b) are two designs of Tangible Reels. c) A tag is positioned underneath so that they can be tracked by a camera placed under a plate glass (d). An infrared frame (highlighted in blue) detects the users' fingers during exploration.**

We tried a number of adhesive materials such as Blu-Tack, glue dots, adhesive tape, Wikki Sticks, anti-slip gel pads or electrostatic screen protectors but none of them fulfilled the first two requirements. Finally we end up with two types of objects fulfilling the requirements: small plastic cylinders filled with lead called Weights, and flat Sucker pads (see Figure 3).

#### Weights

Weights were inspired by [18] who suggested "varying the weight" of the phicons to ensure stability. After several tests with a visually impaired user, we found that filling a 6 cm high and 4 cm wide cylinder with 180 g of lead was adequate. To further improve adherence of the Weights, a silicone O-ring was added under the base. The tag used to track the objects was placed inside this ring (Figure 3c). The base and the top of the cylinder were made out of a thick cardboard strongly glued to the cylinder. The reel was then glued to the top of the cylinder, and its string was passed through a hook fixed at the bottom of the cylinder (1cm. high) to keep it close to the tabletop.

#### Sucker pads

Sucker pads can easily slide along a smooth surface such as the tabletop screen, and strongly stick to the screen when pressed. We used professional flat sucker pads (4 cm wide and 2 cm high once compressed) that present a large surface under which a tag can be attached (Figure 3c). The reel was glued onto the top. The sucker pads can easily be detached by pulling on a small strip that extends from its base.

#### Add-ons

Reels can be harmful when the string retracts. They can also apply a strong tension to the objects they are attached to and make them move. Therefore we used retractable reels with a lock/unlock button. Besides, to make it easier to

link two objects, we fixed a strong neodymium magnet at the extremity of the reels' strings, and added a metallic bracelet to the objects. The bracelet was wrapped around the bottom part of the cylinder for the Weights and around the reels for the Sucker pads.

### CONSTRUCTION AND EXPLORATION TECHNIQUES

Tangible Reels are placed next to the user, on the bottom side of the table. Audio instructions and feedback are provided so that the user can gradually construct a simple physical representation of the map by placing the Tangible Reels (see Figure 3d). During exploration, the user can retrieve the name of the points and lines using finger interactions. All the values mentioned afterward (distances and timers) were based on observations made during preliminary tests.

#### Constructing the map

##### Construction instructions

Each line is constructed using two Tangible Reels attached to each other. Three instructions indicate to the user what is the next action to perform (see Figure 1):

- *“New object”*: at the very beginning of the construction, and each time a new line has to be built, the user has to place a Tangible Reel on the table. As soon as the Tangible Reel is detected, guidance instructions are provided.
- *“Attach an object to the right/to the left/below/above”*: To construct a line the user has to pull out the string of a new Tangible Reel and attach it to the metallic bracelet of the last one that was placed.
- *“Attach an object to <name of the object>, to the right/to the left/below/above”*: The start point of the line to be built is not always the last Tangible Reel that the user has placed. In this case the system gives the name of the object to which the new Tangible Reel must be attached to.

##### Feedback

- *“Attached”*: This instruction is played when the system detects that the new Tangible Reel is close enough to the one that it must be attached to, and is immediately followed by guidance instructions.
- *“<Name of the point> found”*: The system informs the user when the Tangible Reel is at the right location by giving the name of the point represented by the Tangible Reel. If the Tangible Reel is the end point of a line, the instruction *“<name of the line> built”* is played.
- *“Object lost”*: The user is informed when the Tangible Reel that is being moved has not been detected by the system for more than 2500 ms.

The last instruction is repeated every 7000 ms until the appropriate action is done by the user. When the *“attach an object”* instruction is repeated, the name of the object to which the user must attach a new Tangible Reel is also given.

##### Guidance instructions

Depending on the distance between the Tangible Reel that the user is currently moving and the position of the target

point, two types of guidance instructions are provided: rough guidance (every 3500 ms) and fine guidance instructions (every 1500 ms).

- *Rough guidance instructions* (Figure 1b, step 2). When distance is superior to 15 cm, the system indicates the direction of the target (up / down / left / right / up and right / down and right / up and left / down and left) as well as the distance in centimeters. This enables the user to either quickly slide or lift the object towards the target.
- *Fine guidance instructions* (Figure 1b, step 3). When the distance to the target is inferior to 15 cm, the system provides more frequent feedback to indicate the direction to follow (up / right / down / left). As long as the target has not been reached, the system repeats the procedure.

#### Exploring the map

When exploring the map, the user can listen to the name of a point or a line by performing a tap and hold gesture above it. To avoid unintentional selections, the user must select one point or line at a time for at least 700 ms (see Figure 1).

### IMPLEMENTATION

#### Hardware

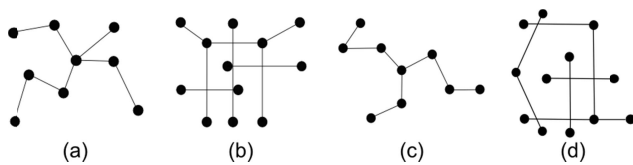
Our tabletop was a 100 x 100 cm plate glass. The setup also included a projector to illuminate the surface and a webcam to detect tagged objects. Both were placed beneath the plate glass. A multitouch IR frame was placed two centimeters above the plate glass (Figure 3d) in order to detect the fingers. To achieve a high quality of tag detection, we restricted the area of work to 80 x 57 cm. The projector, webcam and IR frame were connected to a laptop.

#### Software

The Tangible Reels were tracked using the TopCodes library [7], which allowed using small circular tags that fit under the objects. The IR frame sent messages containing the finger input state (pressed, updated or ended) and position using the TUIO protocol [12]. We used the MultiTouch4Java library (MT4J, [15]) to receive TUIO messages, and to display the image of the map when needed (e.g. for debug) as well as the position of the physical objects and lines. Audio instructions were provided with a SAPI4 compliant Text-To-Speech engine distributed as part of the CloudGarden TalkingJava SDK 1.7.0.

#### PRE-STUDY: TANGIBLE REELS USABILITY

The aim of this pre-study was to investigate whether the two types of Tangible Reels were stable and easy to manipulate, but also to verify that built tangible maps were understandable by visually impaired users. It was done for testing the object design only and was performed without any interactive instruction or feedback. We used two types of maps that are frequently used by visually impaired users: metro maps and overview maps (Figure 2a). The Braille Authority of North Canada defined overview maps as maps that “may not have specific detail that would allow some readers to plan a walking route, but instead are designed to familiarize and orient the reader with the area encompassed”.



**Figure 4: A set of adapted maps used for the exploration task (a, b) and for the construction task (c, d). (a) and (c) are two metro maps while (b) and (d) are two overview maps.**

#### Participants and tasks

We recruited four legally blind persons (two females, two males) aged between 31 and 65 years ( $M = 48.2$ ,  $SD = 14.9$ ). The study consisted in one training phase, one exploration task and one construction task using twelve Sucker pads and twelve Weights. During the training phase, participants were told how to construct a line by attaching two objects together. They could practice until they felt comfortable. Task 1 consisted in exploring one metro map and one overview map that had been previously made by the experimenter. Participants had respectively three and four minutes to explore those maps, and immediately after the exploration, they had to draw the map. They were asked to do it as accurately as possible, focusing on the topology rather than the distances. To draw the maps, three subjects used a Dycem sheet of paper, and one subject, who was not used to Dycem paper, used magnets on a board. Task 2 consisted in reconstructing two maps with the Tangible Reels, as quickly as possible. Participants were shown a raised-line map and had to memorize it before reconstruction. No time limit was imposed for the memorization or the construction. Once they started constructing the tangible map, they could not explore the raised-line map again.

#### Experimental design

We used a within-subjects design with two independent variables for tasks 1 and 2 (Object design and Map):

- Object design (O). We evaluated the two object designs described below: the Sucker pads and the Weights.
- Map (M). For the exploration task we designed four maps (see Figure 4): two represented a part of a fictive metro map and required nine objects to be built and the other represented an overview map (inspired by Figure 2) and required twelve objects. For the construction task, we designed two other fictive metro maps (requiring nine objects) and two other overview maps (requiring twelve objects).

#### Procedure

The study was made up of two blocks corresponding to the two designs. A block consisted in training followed by the exploration task and finally the construction task. For both of these tasks a metro map was presented and then an overview map. After each task, participants answered a questionnaire. They also ranked the object designs according to their preference at the end of the session. The order of the blocks was counterbalanced among users. We also counterbalanced the two sets of maps, each containing

two maps for the exploration and two maps for the construction.

#### Measures

For the exploration task, we measured the unintentional object displacement during exploration (distance in cm). In order to evaluate subjects' spatial learning, we presented the hand-drawn maps to four independent judges who were not involved in the project, alongside pictures of the maps that have been explored. The maps that were made with magnets (Participant 4) were thoroughly reproduced on a Dycem sheet. We asked the judges to evaluate the correctness of the drawn maps as compared to the model: 0/10 means that the two maps were not similar at all; 10/10 means that the two maps were highly similar. We asked the judges to focus on the topology of the map rather than on distances. We had previously shown them three examples of drawings that should receive 0, 5 and 10.

#### Results

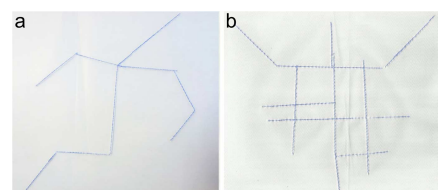
In this pre-study, we mainly focused on qualitative data. Participants are later referred as P1, P2, etc.

#### Objects Stability

The average distance in centimeters between the positions of the objects before and after the exploration of the maps was 0.28 cm ( $SD = 0.03$ ) for the Weights and 0.08 cm ( $SD = 0.01$ ) for the Sucker pads. During all the explorations, one Sucker pad got detached; none was moved. P1 almost knocked over three Weights, and P3 knocked over two Weights.

#### Map drawings and map construction

Figure 5 shows examples of drawings. The average marks given were 8.1 ( $SD = 1.3$ ) for the metro maps and 5.7 ( $SD = 2.7$ ) for the overview maps. P4 obtained the lowest marks for the overview drawings (1.5 and 2.5). If we exclude her marks, the average mark for the overview maps was 7.0 ( $SD = 1.4$ ). One metro map and three overview maps constructed by the participants were not similar to the model. The subjects indicated that they could not remember the whole raised-line map.



**Figure 5: two drawings made after exploring the metro map (a) and the overview map (b) presented in Figure 4. They were respectively given a score of 8.6 and a 6.6.**

#### Questionnaire and ranking

The participants specified on a 7-point Likert scale their level of agreement on a series of statements (1 = strongly disagree and 7 = strongly agree), for each task and each object design. The first four items were: *Building a map with these objects is: 1) pleasant; 2) difficult; 3) fast; 4) frustrating*. The last item was: *5) it is easy to*

unintentionally move or knock over these objects. For the exploration task, a sixth item was evaluated: “these objects allowed me to understand the maps”. Table 1 shows the percentage of agreement for each statement.

All the participants preferred the Sucker pads for the exploration task but results were mixed for the construction task (two preferred the Sucker pads and the other two the Weights). Overall, three participants preferred the Sucker pads as a global best choice.

Task	Design	Pleasant	Difficult	Fast	Frustrating	Easy to knockover	Helpful
Exploration	Weights	25%	25%	50%	25%	50%	75%
	Sucker pads	100%	0%	50%	0%	0%	100%
Construction	Weights	75%	0%	100%	25%	50%	
	Sucker pads	100%	0%	75%	0%	0%	

**Table 1. Percentage of subjects who answered a 5, 6 or 7 for each item and Object Design, after the exploration and the construction tasks. The darker the cell, the higher the value.**

**Qualitative feedback**

Concerning the Weights, two participants stated that when they had to replace several objects it was easier to do with the Weights than the Sucker pads (P2, P4). P3 stated that as they could be easily moved, it was handy to adjust their position when constructing the map. Participants reported concerns when exploring the map with the Weights: P2 declared that he “missed one object because [he] was paying attention not to knock them over”. The same issue was reported by P1 who said that he was “afraid of knocking them over” and that they hindered the exploration. Two participants also reported that the height of the Weights was an issue, rather than their diameter (P1, P3).

As for the Sucker pads, three participants said that their reduced height allowed them to better explore the map (P1, P3, P4), and two stated that they were fewer risk to move them during the exploration (P1, P4). P4 also declared that “the advantage is that they do not take a lot of place” while P3 described the sucker pads as “cool”, “light” and “fun”. P4 said that “attaching and detaching the Sucker pads is a little bit annoying”.

**Conclusion to the pre-study**

Both Tangible Reels proved to be easy to manipulate by a visually impaired user. Over the four Object Design \* Tasks conditions, one participant only considered that building a map with the Tangible Reels was difficult while the majority found it pleasant. However, the Sucker pads appeared to be more stable than the Weights. This is coherent with the fact that two participants found that the Weights were easy to unintentionally knock over. One drawback of the Sucker pads is that they cannot be removed as easily as the Weights. However, it should be noted that participants had to remove the Sucker pads several times during the construction task in this pre-study. This is unlikely to happen in a real scenario because the positioning of the object is precisely guided before fixation.

Several participants also reported that the height of the Weights hindered the exploration, and two participants knocked some Weights over. It is essential for the users not to be hindered by the objects when exploring or constructing a map. To sum up, it appeared that Sucker pads better meet the stability requirement, and that they were globally preferred by three participants. According to these observations, even though we do not consider that Weights were not usable, we chose the Sucker pads for the following experiment.

The marks attributed to drawings show that maps constructed with Tangible Reels can be explored and memorized by visually impaired users. Three out of four participants found that the Tangible Reels allowed them to understand the map. However, the existence of incorrect drawings showed that some participants experienced difficulties, especially on overview maps that were more complex. Indeed, three participants said that they were quite difficult to understand and memorize. This observation suggests that the maps built with Tangible Reels should not be too complex. In the follow-up study, we specifically investigated the effect of map complexity on the usability of Tangible Reels.

**STUDY: MAP CONSTRUCTION AND EXPLORATION**

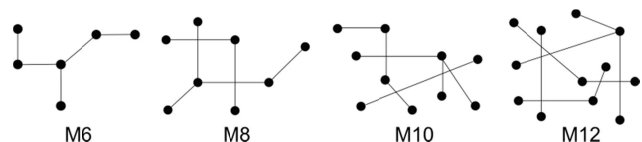
The aim of this study was to evaluate the usability of the whole interactive device with Sucker pads, and to investigate whether increasing the levels of complexity of the maps led to usability issues. We used the apparatus described in the Implementation section above.

**Participants**

We recruited 8 legally blind persons (2 females, 6 males) aged between 24 and 65 (M = 43.8, SD = 14.4). Four were born blind; one became blind before the age of 1; three between 4 and 6 years old and one at 16. Three had residual light perception but were unable to discern shapes; others could not perceive anything at all.

**Experimental design**

We designed four maps of different complexities (see Figure 6) by gradually augmenting the number of points (6, 8, 10, 12), lines (5, 6, 7, 8) including oblique lines (1, 2, 3, 4), as well as crossings between lines (0, 2, 4, 6). Besides, each map contained two horizontal lines and two vertical lines as well as one point that was the start or end point of three different lines. These maps are later referred to as M6, M8, M10 and M12. All points and lines were associated with a numerical label, ranging from 1 to 12.



**Figure 6: the four maps that the participants had to build using Tangible Reels. The complexity is increasing from left to right.**

We used a within-subjects design with the Complexity of maps (four levels of complexity) as independent variable. All the participants had to construct the maps in the same order of increasing difficulty (M6, M8, M10 and M12).

### Instructions and tasks

The study consisted in one familiarization phase followed by construction and exploration tasks with the Sucker pads. During familiarization, participants were told how to manipulate the Sucker pads, how to interpret audio instructions to construct a map, and how to explore it. They could practice as many times as they wanted to on a map made of five points.

After the familiarization phase, the participants built each of the four maps, and then answered three questions of this type: *what are the names of the two points at the extremities of <name of a line>?* They had to explore the map (see Figure 1c) and select the appropriate line and points before answering. During construction, when participants spent more than three minutes on one instruction, it was considered as a failure, and we provided help so that they could continue.

After each construction, participants had to rate the difficulty of the task on a 7-point Likert scale. At the end of the session, they answered a SUS [3] and a NASA-TLX [5] questionnaires, as well as the following question: *do you think that constructing the map helped you to understand it?* They were also invited to provide any comment on the system.

### Measures and statistical methods

Usability was evaluated by measuring efficiency, effectiveness and satisfaction. Efficiency was assessed with three measures: 1) time needed to build the entire map; 2) time to place one object, from the first construction instruction until the object was correctly placed, and 3) time to answer each question. Effectiveness was assessed by the number of maps each participant successfully built and the number of correct answers to the questions. Satisfaction was measured using the SUS questionnaire and the participants' comments. During the construction, we logged successive positions of each Sucker pad being placed, as well as the occurrence of the different instructions. During exploration, we logged for each question the ID of the selected elements as well as the time at which they were selected. We therefore collected  $(6 + 8 + 10 + 12) * 8$  participants = 288 trials for construction, and  $4$  maps \*  $3$  questions \*  $8$  participants = 96 trials for exploration.

We used the Shapiro-Wilk test to determine if the collected data followed a normal distribution. When distributions were normal, we computed a Univariate ANOVA test. Otherwise we used a Friedman test. Post-hoc tests were performed with the Wilcoxon Signed-Ranks Test for Paired Samples with a Benjamini and Hochberg correction.

## Results

### Construction

Twenty seven maps out of 32 were correct, which corresponded to 283 Sucker pads out of 288 (98.3%) correctly positioned and linked. Two maps were incorrect because one line did not start from the right Sucker pad (M6, M12), and three maps were considered incorrect because the participants required assistance from the experimenter (M8, M12, M12). Because the number of errors was very low (5) we provide further details on each one. For Participant 2 (P2) in M6, the instruction "attached" was played before the user effectively attached the Sucker pad. By the time, the user had moved away and therefore attached the new Sucker pad to an incorrect one. P1 (on M12) pulled out a string too strongly and detached the Sucker pad. He then reattached it to the wrong Sucker pad. P3 (M12) spent several minutes trying to attach a Sucker pad to the point 9, whereas the instruction "attach an object to 8" was repeated several times. P8 experienced difficulties focusing on the task: when placing one object on M10, he moved the extremity of the string instead of the Sucker pad, so that the guidance instructions remained unchanged. When constructing M12, he found the good Sucker pad to attach a new one, but did not wait enough to hear the feedback ("attached"), and then tried to attach it to other Sucker pads.

Figure 7 shows the average time to construct correct maps (not including 3 maps constructed with the help of the experimenter). A Friedman test showed a significant effect of Map Complexity on completion times ( $X^2=16.4$ ,  $p<.001$ ). Post-hoc pairwise comparison revealed significant completion time differences between M6 and M8, M10 and M12 ( $p<.05$ ), between M12 and M8, and between M12 and M10 ( $p<.05$ ).

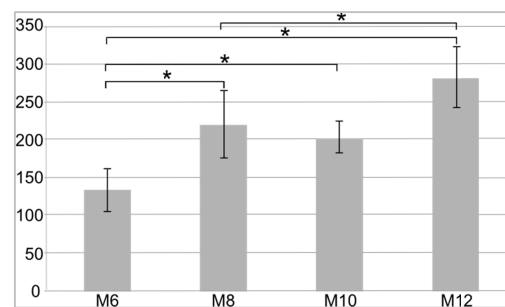
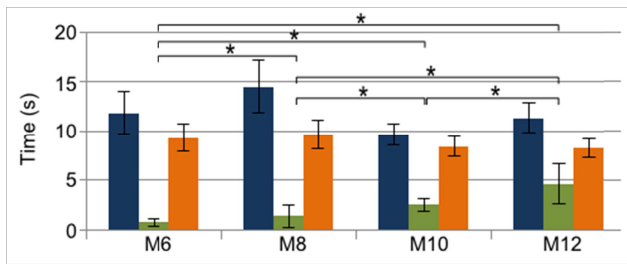


Figure 7: Average completion times (in seconds) to construct maps (error bars are IC95).

The average time to place a Tangible Reel was 23.7 on average. Post-hoc pairwise comparison revealed a significant difference in time needed to place a TR between M10 and M12 ( $p<.05$ ). There was also a significant effect of Map Complexity on rough guidance time per object ( $X^2=20.85$ ,  $p<.001$ ). Post-hoc pairwise comparison revealed significant completion time differences between all the conditions ( $p<.05$ ) as shown in Figure 8.





**Figure 8.** Average times and IC95 for the three steps required to correctly place a Tangible Reel: 1) attaching a Sucker pad to a previous one or placing it anywhere on the table (blue); 2) following rough guidance instructions (green); 3) following fine guidance instructions (orange).

*Exploration*

Table 2 indicates the average completion times, and the number of points and lines selected for each question (correct answers only). We found an interaction between Map Complexity and Number of elements selected ( $X^2= 91, p=.01$ ). Post-hoc pairwise comparison showed that participants selected less elements when exploring M6 as compared to M8, M10 and M12 ( $p<.05$ ). Completion times to answer each question followed a normal distribution. An ANOVA with Map Complexity as factor showed that exploration times differed significantly ( $F(3,28)=5.70, p<.01$ ). A post-hoc Tukey’s HSD test indicated that exploration times for M12 were significantly higher than for M8 ( $p<.05$ ) and M6 ( $p<.01$ ).

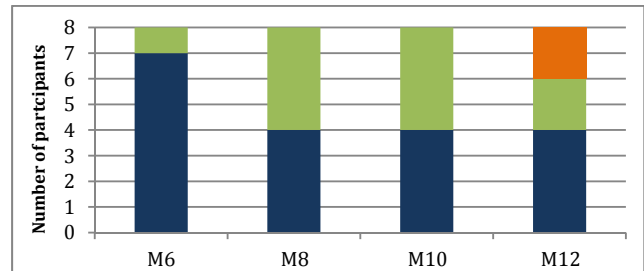
	M6	M8	M10	M12
(a)	14.6 (11.1)	22.5 (17.7)	31.4 (23.9)	41.9 (27.8)
(b)	3.9 (0.9)	6.0 (1.0)	7.3 (2.0)	8.6 (0.8)

**Table 2.** Average time to answer one exploration question (a) and average number of elements selected (b). SDs are indicated in parenthesis.

The percentages of correct answers to the exploration questions were: 91.7% for M6; 95.8% for M8; 91.7% for M10, and 79.2% for M12. Errors were due to the fact that: 1) participants pointed to the right intersection but followed the wrong line (P2 twice, P3, P6); 2) maps were incorrectly constructed (P1, P2); 3) Sucker pads got detached, creating a misalignment between the tangible map and the digital map (P8 twice); 4) participants had trouble to correctly perform the pointing gestures and therefore to select the lines (P3, P6).

*Questionnaires*

The average SUS score for the system was 83.6. Participants had to rate the difficulty of the task on a 7-point Likert scale (1 = very easy; 7 = very difficult). Figure 9 illustrates the number of participants who found the task rather easy (1, 2), normal (3, 4, 5) or difficult (6, 7). Table 3 indicates the scores of the NASA-TLX questionnaire.



**Figure 9:** for each map, number of participants who found the task rather easy (blue); normal (green); difficult (orange).

**Conclusions about map construction and exploration**

*Construction*

Results shows that most of the participants managed to construct the maps without experiencing any usability issues. 85% of the maps were correctly constructed, and the most complex map was constructed in approximately four minutes. The rough guidance technique allowed the users to quickly move a Sucker pad close to the target and hence proved to be efficient. On average it was three times longer to follow fine than rough guidance instructions, which suggests that the overall completion time could be reduced by improving the fine guidance mode. Most of the participants found the instructions easy to understand. P4 stated that “instructions are extremely clear, it is impossible to make mistakes” and similar comments were made by P1, P2, P3, P6. However, P3 and P5 declared that at the end of the experiment they were getting tired, notably because it was necessary to remain concentrated (P3 said that she failed in constructing M12 because she felt like she “couldn’t hear the instructions anymore”). Similarly, we observed that P4 failed in constructing M10 and M12 because he did not keep on focusing on the instructions.

*Exploration*

89.6% of the answers were correct, which shows that the system is usable to retrieve specific spatial information. However, we observed two issues: 1) some participants had difficulties to perform pointing gestures with one finger only, and did not always manage to quickly select the lines and points. For example, P3 kept using several fingers and spent on average 8.7 seconds to select one element, whereas P7, who perfectly understood the gesture, spent only 0.96 seconds per element. 2) Some participants pointed at the intersections of two lines and therefore did not always manage to select the line they wanted to.

	<33	<66	<100
Mental demand	62.5	25	12.5
Physical demand	75	0	25
Temporal demand	75	25	0
Effort	75	25	0
Performance	0	12.5	87.5
Frustration	75	25	0

**Table 3.** For each NASA-TLX criteria, percentage of participants who indicated a value inferior to 33, 66 and 100.



### Complexity

Four participants found that difficulty remained similar in the last three conditions (P1, P2, P4, and P5) or in the last two conditions (P4 and P7). The two participants who failed at reconstructing M12 found this task difficult. P1 and P5 stated that it became difficult to attach two Tangible Reels together when they were already close to other Tangible Reels. Two subjects (P4 and P7) declared that it will be too difficult to understand and explore more complex maps. It is interesting to note that except between M10 and M12, the completion times to place one object did not significantly increase with the map complexity, neither the number of usability issues. Three failures (in M10, M12 and M12) were likely due to a lack of concentration that could either result from the length of the whole experiment (these conditions happened at the end of the experiment) or from the complexity of the map *per se*.

Overall, we did not observe any statistical difference between the times to construct oblique, vertical or horizontal lines. However, we observed that the time to place an object anywhere on the table after hearing the “new object” instruction (i.e. an object that must not be linked to a previous one) gradually increases as the map was being constructed (a Pearson correlation coefficient revealed a positive correlation :  $r=0.64$ ,  $p<0.01$ ). This probably reflects the time needed to find room where to place a new Sucker pad, suggesting that the complexity of the maps that can be built is limited by physical constraints (number and size of Tangible Reels) rather than by the user’s ability to interpret audio instructions and perform appropriate actions.

### Satisfaction and usefulness

Overall, participants were highly satisfied with the performance (NASA-TLX score superior to 66% for 85% of the participants), and found the system usable, as indicated by the SUS score. Six participants mentioned that they were able to understand the maps after construction, and two pinpointed that it would have been easier if they were explicitly asked to understand and memorize the maps. Finally, three participants said that the system would be very helpful for educational purposes (mathematics and orientation and mobility lessons).

## DISCUSSION

Overall, we showed that Tangible Reels are efficient to materialize points and lines on tangible maps, and are easy to move and stable. A reel with strong magnets is used to link two or more phicons together and provide tangible lines. The pre-study showed that participants were able to easily manipulate Tangible Reels, and understand maps built with them. Results of the main study demonstrated that maps of different complexity are very often correct (85%) and can be explored by the users with accuracy (89.6% of correct answers). In addition, the system was efficient (23.7s only in average to place an object) and satisfying (SUS of 83.6). The exploration mode could be improved by allowing multiple fingers selection of points

and lines, and providing specific feedback for lines crossings. However, several participants reported that Weights hindered map exploration, which suggests that the height of phicons should be carefully considered when they are used by visually impaired users.

### Map complexity

The number of Tangible Reels that can be placed above the table limits the complexity of the map. In addition, although most of the participants managed to construct the more complex maps, some reported that the task was difficult when objects were too close to each other. When a larger amount of information is required, it would be interesting to materialize the most important elements using Tangible Reels, and then provide access to less important elements using gestural interactions as well as audio and haptic feedback (see [22] for instance).

### Designing advanced functions

We designed a solution to make both points and lines tangible. Areas could also be represented and conveyed through audio (see [27] for example) or illusory tactile textures [6]. Besides, we started to design small objects (called “modifiers”) that are attached to the strings and bend them to construct curves. Tangible Reels may also provide advanced functions like displaying/hiding particular points of interests, or zooming and panning. For example, for zooming, users will have to select the zoom mode, and move apart two objects. The digital map will be modified accordingly and the user will be guided to reposition Tangible Reels and place new ones.

### More than tangible maps, tangible graphics

With the materialization of points and lines, we made it possible to construct maps. Obviously, Tangible Reels can potentially materialize any type of graphical representation including points, lines, and areas (graphs, flow charts, bar charts, etc.). Figure 10 illustrates three examples of graphical representations that a visually impaired person may access with Tangible Reels.

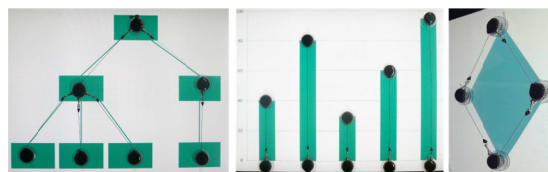


Figure 10. Other graphics with Tangible Reels

## CONCLUSION

Graphics (maps, organigrams, bar charts, etc.) are widely used in education but also in everyday life. However technologies that help visually impaired users to perceive them are still uncommon. In this paper we introduced a tangible tabletop interface that allows building and exploring tangible graphics. More precisely, we described the design of Tangible Reels, a new type of phicons that can be used to materialize points and lines. We showed that they are stable, easy to manipulate, and can be used to convey spatial representations of different complexity.

**ACKNOWLEDGMENTS**

We thank the LACII and notably Claude Griet, Laurence Boulade and Nathalie Bedouin. We are very grateful to all the participants of the study. We also thank Gilles Bailly, Anke Brock and Bernard Oriola for their meaningful comments. This work is part of the AccessiMap project (research grant AccessiMap ANR-14-CE17-0018).

**REFERENCES**

1. G Blasko, Chandra Narayanaswami, and Steven Feiner. 2006. Prototyping retractable string-based interaction techniques for dual-display mobile devices. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*: 369–372. <http://doi.org/10.1145/1124772.1124827>
2. Anke M. Brock, Philippe Truillet, Bernard Oriola, Delphine Picard, and Christophe Jouffrais. 2015. Interactivity Improves Usability of Geographic Maps for Visually Impaired People. *Human-Computer Interaction* 30: 156–194.
3. John Brooke. 1996. SUS: A “quick and dirty” usability scale. In *Usability Evaluation in Industry*, P. W. Jordan, B. Thomas, B. A. Weerdmeester and I. L. McClelland (eds.). Taylor & Francis, London, UK, 189–194.
4. Polly Edman. 1992. *Tactile graphics*. AFB press, New York, USA.
5. Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (eds.). Elsevier, 139–183. [http://doi.org/10.1016/S0166-4115\(08\)62386-9](http://doi.org/10.1016/S0166-4115(08)62386-9)
6. V. Hayward. 2008. A brief taxonomy of tactile illusions and demonstrations that can be done in a hardware store. *Brain research bulletin* 75, 6: 742–752.
7. M. T Horn. TopCode: Tangible Object Placement Codes. Retrieved from <http://hci.cs.tufts.edu/topcodes/>
8. CJ Huang, Ellen Yi-luen Do, and D Gross. 2003. MouseHaus Table: a Physical Interface for Urban Design. *16th Annual ACM Symposium on User Interface Software and Technology (UIST)*: 41–42.
9. Hiroshi Ishii and Brygg Ullmer. 1997. Tangible bits: towards seamless interfaces between people, bits and atoms. *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*: 234–241. <http://doi.org/10.1145/258549.258715>
10. R.D. Jacobson. 1998. Navigating maps with little or no sight: An audio-tactile approach. *Proceedings of Content Visualization and Intermedia Representations*, 95–102. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.21.2860>
11. Nikolaos Kaklanis, Konstantinos Votis, Panagiotis Moschonas, and Dimitrios Tzovaras. 2011. HapticRiaMaps: Towards Interactive exploration of web world maps for the visually impaired. *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility - W4A '11*, ACM Press, 20.
12. T. Kaltensbrunner, M., Bovermann, R. Bencina, and E Costanza. 2005. TUIO - A Protocol for Table-Top Tangible User Interfaces. *Proceedings of the 6th International Workshop on Gesture in Human-Computer Interaction and Simulation (GW 2005)*.
13. Shaun K. Kane, Meredith Ringel Morris, Annuska Z. Perkins, Daniel Wigdor, Richard E. Ladner, and Jacob O. Wobbrock. 2011. Access Overlays: Improving Non-Visual Access to Large Touch Screens for Blind Users. *Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST '11*, ACM Press, 273–282. <http://doi.org/10.1145/2047196.2047232>
14. Erik Koch and Hendrik Witt. 2008. Prototyping a chest-worn string-based wearable input device. *2008 International Symposium on a World of Wireless, Mobile and Multimedia Networks*: 1–6. <http://doi.org/10.1109/WOWMOM.2008.4594882>
15. Uwe Laufs, Christopher Ruff, and Jan Zibuschka. 2010. MT4j - A Cross-platform Multi-touch Development Framework. *ACM EICS 2010, Workshop: Engineering patterns for multi-touch interfaces*, ACM, 52–57.
16. AK Lobben. 2005. Identifying the needs of tactile map makers. *Proceedings of XXII International Cartographic Conference*.
17. Muhanad S Manshad, Enrico Pontelli, and Shakir J. Manshad. 2012. Trackable Interactive Multimodal Manipulatives: Towards a Tangible User Environment for the Blind. *Proceedings of ICCHP 2012*, Springer Berlin Heidelberg, 664–671. <http://doi.org/10.1007/978-3-642-31534-3>
18. David McGookin, Euan Robertson, and Stephen Brewster. 2010. Clutching at Straws: Using Tangible Interaction to Provide Non-Visual Access to Graphs. *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, ACM Press, 1715–1724. <http://doi.org/10.1145/1753326.1753583>
19. Joshua A. Miele, Steven Landau, and Deborah Gilden. 2006. Talking TMAP: Automated generation of audio-tactile maps using Smith-

- Kettlewell's TMAP software. *British Journal of Visual Impairment* 24, 2: 93–100.  
<http://doi.org/10.1177/0264619606064436>
20. Peter Parente and Gary Bishop. 2003. BATS : The Blind Audio Tactile Mapping System. *Proceedings of ACM South Eastern Conference*, ACM Press.
  21. Norman Pohl, Steve Hodges, John Helmes, Nicolas Villar, and Tim Paek. 2013. An interactive belt-worn badge with a retractable string-based input mechanism. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*: 1465.  
<http://doi.org/10.1145/2470654.2466194>
  22. Benjamin Poppinga, Charlotte Magnusson, Martin Pielot, and Kirsten Rassmus-Gröhn. 2011. TouchOver map: Audio-Tactile Exploration of Interactive Maps. *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI '11*, ACM Press, 545–550.  
<http://doi.org/10.1145/2037373.2037458>
  23. Bernhard Schmitz and Thomas Ertl. 2012. Interactively Displaying Maps on a Tactile Graphics Display. *SKALID 2012–Spatial Knowledge Acquisition with Limited Information Displays (2012)*, 13–18.
  24. Jochen Schneider and Thomas Strothotte. 2000. Constructive exploration of Spatial Information by Blind Users. *Proceedings of the fourth international ACM conference on Assistive technologies*: 188–192.  
<http://doi.org/10.1145/354324.354375>
  25. Ehud Sharlin, Benjamin Watson, Yoshifumi Kitamura, et al. 2004. The Tangible Pathfinder Design of a Wayfinding Trainer for the Visually Impaired. *Proc. Graphics Interface*: 2–3.
  26. Mathieu Simonnet, Dan Jacobson, Stephane Vieilledent, and Jacques Tisseau. 2009. SeaTouch: a haptic and auditory maritime environment for non visual cognitive mapping of blind sailors. *COSIT 2009, LNCS 5756*, Springer-Verlag, 212–226.  
[http://doi.org/10.1007/978-3-642-03832-7\\_13](http://doi.org/10.1007/978-3-642-03832-7_13)
  27. The Braille Authority of North America. 2010. *Guidelines and Standards for Tactile Graphics*. Retrieved from <http://brailleauthority.org/tg/web-manual/index.html>
  28. J Underkoffler and H Ishii. 1999. Urp: a luminous-tangible workbench for urban planning and design. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*: 386–383.  
<http://doi.org/10.1145/302979.303114>
  29. Fernando Vidal-verdú and Moustapha Hafez. 2007. Graphical Tactile Displays for Visually-Impaired People. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* 15, 1: 119–130.
  30. Zheshen Wang, Baoxin Li, Terri Hedgpeth, and Teresa Haven. 2009. Instant Tactile-Audio Map: Enabling Access to Digital Maps for People with Visual Impairment. *Proceeding of the eleventh international ACM SIGACCESS conference on Computers and accessibility - ASSETS '09*, ACM Press, 43–50.  
<http://doi.org/10.1145/1639642.1639652>
  31. Limin Zeng and Gerhard Weber. 2011. Accessible Maps for the Visually Impaired. *Proc. of IFIP INTERACT 2011 Workshop on ADDW*, 54–60.
  32. Limin Zeng and Gerhard Weber. 2012. ATMap: Annotated Tactile Maps for the Visually Impaired. *COST 2102 International Training School, Cognitive Behavioural Systems, LNCS Volume 7403, 2012*, Springer Berlin Heidelberg, 290–298.  
<http://doi.org/10.1007/978-3-642-34584-5>
  33. Haixia Zhao, Catherine Plaisant, Ben Shneiderman, and Jonathan Lazar. 2008. Data Sonification for Users with Visual Impairment. *ACM Transactions on Computer-Human Interaction* 15, 1: 1–28.