A body-centric framework for generating and evaluating novel interaction techniques

Julie Wagner

LABORATOIRE DE RECHERCHE EN INFORMATIQUE

UNIVERSITÉ PARIS SUD

Inria informatics mathematics

UNIVERSITÉ PARIS-SUD

ÉCOLE DOCTORALE D'INFORMATIQUE DE PARIS-SUD

LABORATOIRE DE RECHERCHE EN INFORMATIQUE

INRIA SACLAY–ÎLE-DE-FRANCE

Discipline: Informatique

THÈSE DE DOCTORAT

soutenue le 6 Décembre 2012

par

Julie WAGNER

A Body-centric Framework for Generating and Evaluating Novel Interaction Techniques

Directeur de thèse :	Wendy MACKAY	Directeur de Recherche (Inria Saclay–Île-de-France)
Co-Directeur de thèse :	Stéphane HUOT	Maître de Conférences (Université Paris-Sud)
Composition du jury :		
Rapporteurs :	Mountaz HASCOËT	Maître de Conférences, HDR (Université Montpellier II)
	Celine LATULIPE	Associate Professor (University of North Carolina at Charlotte)
Examinateurs :	Steven FEINER	Professor (Columbia University, New York)
	Anne VILNAT	Professeur (Université Paris-Sud), Président du jury

Table of Contents

Abstract

This thesis introduces BodyScape, a body-centric framework that accounts for how users coordinate their movements within and across their own limbs in order to interact with a wide range of devices, across multiple surfaces. It introduces a graphical notation that describes interaction techniques in terms of (1) motor assemblies responsible for performing a control task (input motor assembly) or bringing the body into a position to visually perceive output (output motor assembly), and (2) the movement coordination of motor assemblies, relative to the body or fixed in the world, with respect to the interactive environment.

This thesis applies BodyScape to 1) investigate the role of support in a set of novel bimanual interaction techniques for hand-held devices, 2) analyze the competing effect across multiple input movements, and 3) compare twelve pan-and-zoom techniques on a wall-sized display to determine the roles of guidance and interference on performance.

Using BodyScape to characterize interaction clarifies the role of device support on the user's balance and subsequent comfort and performance. It allows designers to identify situations in which multiple body movements interfere with each other, with a corresponding decrease in performance. Finally, it highlights the trade-offs among different combinations of techniques, enabling the analysis and generation of a variety of multi-surface interaction techniques. I argue that including a body-centric perspective when defining interaction techniques is essential for addressing the combinatorial explosion of interactive devices in multi-surface environments.

Keywords: Multi-surface interaction, Body-centric design space, bimanual interaction, Tablet computer, BiTouch design space, BiPad, Multi-scale interfaces, Pan & Zoom, Navigation, Wall-sized displays, Mid-air interaction techniques..

Résumé

Cette thèse présente BodyScape, un espace de conception qui décrit la façon dont les utilisateurs coordonnent les mouvements de, et entre leurs membres lorsqu'ils interagissent avec divers dispositifs d'entrée et entre plusieurs surfaces d'affichage. BodySape introduit une notation graphique pour l'analyse des techniques d'interaction en termes : (1) d'assemblages de moteurs, qui accomplissent une tâche d'interaction atomique (assemblages de moteurs d'entrée), ou qui positionnent le corps pour percevoir les sorties du système (assemblages de moteurs de sortie); (2) de coordination des mouvements de ces assemblages de moteurs, relativement au corps de l'utilisateur ou à son environnement interactif.
Nous avons appliqué BodyScape à : 1) la caractérisation du rôle du support dans l'étude de nouvelles interactions bimanuelles pour dispositifs mobiles; 2) l'analyse des effets de mouvements concurrents lors de l'interaction; et 3) la comparaison de techniques d'interaction multi-échelle sur mur d'images afin d'évaluer le rôle du guidage et des interférences sur la performance.
La caractérisation des interactions avec BodyScape clarifie le rôle du support des dispositifs d'interaction sur l'équilibre de l'utilisateur, et donc sur le confort d'utilisation et la performance qui en découlent. L'espace de conception permet aussi aux concepteurs d'interactions d'identifier des situations dans lesquelles des mouvements peuvent interférer entre eux et donc diminuer performance et confort. Enfin, BodyScape révèle les compromis à considérer a priori lors de la combinaison de plusieurs techniques d'interaction, permettant l'analyse et la génération de techniques d'interaction variées pour les environnements multi-surfaces. Plus généralement, cette thèse défend l'idée qu'il est primordial d'adopter une approche centrée sur les fonctions corporelles engagées au cours de l'interaction, afin de maîtriser le nombre et la complexité croissants des dispositifs d'entrée disponibles lors de la conception de techniques d'interaction pour les environnements multi-surfaces.

Mots-clés: Interaction multi-surfaces, Espace de conception centré sur l'engagement corporel, Interaction bimanuelle, Tablettes tactiles, Espace de conception BiTouch, BiPad, Interfaces multi-échelle, Pan & Zoom, Navigation, Murs d'images, Techniques d'interaction gestuelles (mid-air).

Acknowledgments

I want to thank all those people who have accompanied me during the last 3 years of my Ph.D and who – not always knowingly – greatly influenced me in my ideas and opinions.
First of all, there are my two thesis advisors, Wendy E. Mackay and Stéphane Huot, two amazing researchers, that I had the pleasure to learn from.
Wendy, thank you for welcoming me in your lab and giving me the opportunity to experience working abroad. You gave me confidence, motivation and ideas for the next step, when I stepped into your office with my problems. I particular loved how you continuously reminded me to look for the 'intuition' inside of me that unconsciously drives my interests and research. I had great pleasure to learn from your HCI experience and 'magic' writing skills and will continue to apply everything you taught me to my future work.
Stéphane, thank you for being a friend and advisor, for being always there for me, especially in moments where I did not see the forest for the trees because you know 'rechts sind Bäume, links sind Bäume und dazwischen Zwischenräume' (a german song for children which sounds intimidating to french ears). You taught me useful work methods that I will continue to 'stufify' my work with. You were a huge support in pursuing my ideas and my research. I admire your research skills as much as your technical skills and enjoyed working with you.
A huge amount of my education is also thanks to the whole INSITU team. You are welcoming and supporting and I particularly loved the research discussions as well as discussions about anything else with all of you.
A special thanks goes to my fellow-Ph.D student Mathieu Nancel who wrote and submitted his thesis at the same time I did.
Mathieu, thanks for listening to all my 'little' problems and making me laugh every day. I enjoyed working, discussing ideas, programming and running experiments with you. The beers in the evenings were the cherry on the cream and our bizarre discussions about the world were relaxing.
I thank my parents Ernst and Farangis, my brothers Marcus, Marcel, and Jean for all their guidance as well as my good friends Michael Marcic, Anne Hunfeld, and Nina Bastian for always being there for me.
Last but not least, I thank the members of my jury Céline Latulipe, Mountaz Hascouët, Steven Feiner and Anne Vilnat for their time and effort on reviewing my thesis.

"We're in the middle of a period that I refer to as a period of 'combinatorial innovation'. So if you look historically, you'll find periods in history where there would be the availability of a different component parts that innovators could combine or recombine to create new inventions."

Hal Varian, Chief Economist at Google

1 Introduction

Computation has decentralized from a single box to lots of computers embedded into the environment and has changed size and shape. The body configuration shifted from a steadily seated configuration to a dynamically changing configuration involving large body parts in the interaction. I am interested in studying interaction in multi-surface environments, that demand interaction techniques with, across and beyond devices to support users' work strategies. Interaction design for such environments face a combinatoric explosion of on-device, with-device, free-hand and whole-body input. Previous work has studied individual interaction techniques. We face, however, an imbalance between the exploration of novel technology and interaction design and theoretical grounding. I propose a body-centric analysis of interaction by identifying key dimensions affecting human motor control. I argue that a body-centric taxonomy enables us to cope with the complexity by describing an interaction technique in terms of performed body input movements independent from the choice of technological implementation.

In the last decades, computers drastically changed from traditional personal computer (PC) interfaces to more physical forms: the computer shifted to either wearable, portable, or stationary devices; monitors became flat and large, and shifted from vertical orientation to horizontal orientation; the keyboard shrunk in size operated with one or two thumbs on a mobile device, turned into a purely virtual keyboard on larger interactive surfaces, or is completely replaced by using gestural input; the mouse became wireless, portable, or is completely replaced by pointing gestures. The technological part of Weiser's ubiquitous computing vision became true today [Wei99]: computation decentralized from a single box to many small devices of various form factors. Each form factor implies a different *spatial body-device relationship*.

With changing spatial body-device relationship, the body's role in interaction changes as well: steadily seated in front of the monitor of a desktop personal computer, the two *hands* perform small movements over keyboard and mouse; standing in front of a large display, the two *arms* perform large ***mid-air*** gestural movements; and when dancing and waving in front of a gaming console, the *entire body* is in motion. We observe an increasing *level of body involvement* into interaction techniques.

The shift in computation has an impact on the user's body configuration. With devices shrinking in size, they become portable: both device support and on-device interaction must be shared across parts of the user's body. As devices grow in size, users can work collaboratively with other people. They can stand or walk in front of a large display, use their hands to perform gestures in the air (mid-air gestures) or physically touch the display to manipulate their data. The *interactive environment* imposes specific body configurations *affecting interactive performance*.

I am interested in studying arising novel computer environments. Visionaries in human-computer interaction research explored the creation of new interaction setups that change the body's spatial relationship with a device, from augmented physical desks [Wel93] and tangible tabletop interfaces [FIB95] to entire rooms where users interact with multiple interactive surface technologies [Bro+97]. Computation shifts into the physical world and with constant technological advancement, we can choose from an extensive selection of devices such as game controllers, mobile devices, tablets, interactive tables and large displays. Some tasks, such as the exchange of files, persuades to design interaction *across* devices. Depth cameras and tracking systems, such as Microsoft Kinect[1] or the VICON tracking system[2], allow the detection of interaction going on *around and between* devices.

We have arrived at a *golden age* in human-computer interaction. We have explored a lot of different input possibilities and technologies. Now we have to take a step back in order to reflect on theories that help us understanding

[1]http://www.xbox.com/en-US/KINECT
[2]http://www.vicon.com/products/vicontracker.html

the differences among all those input possibilities and that help us to find yet unexplored areas for interaction design.

Historically, research in HCI faced a similar problem for the design of input devices: how to arrange a set of buttons, knobs, and sliders on a device in the best way? Card et al. [CMR91] applied a ***morphological approach*** to analyze input devices in terms of primitive body movements and their composition. Multi-surface environments, however, are complex environments where interaction can involve larger body parts than conventional input devices and where input can be performed with various concurrently performed body motions. In addition, body parts can have supplementary roles imposed by environmental factors, such as supporting a device or walking over to a large stationary device.

We need a model that describes (i) the assignment of various body parts to tasks within an interaction technique, such as pointing and ***zooming***; (ii) body-internal constraints such as restricted movements due to device support; and (iii) environmental factors that affect users' body configuration in the environment, such as the restriction to remain at a specific position in front of a stationary device. In contrast to buttons, knobs and sliders on input devices, body parts are interdependent: movements of one limb entails movements of neighboring limbs. This motion interdependence between body parts is commonly referred to as ***kinematic chain***. I propose a *body-centric morphological design space*, called *BodyScape*, that is grounded in Card's [CMR91] morphological approach, the kinematic chain model, and early psychological studies on human motor control [Bry92].

BodyScape describes an interaction technique according to the body's involvement into the interaction and the spatial relationship between body and the interactive environment. For instance, users can perform small circular gestures, *MircoRolls* [RLG09], with their right thumb on a mobile phone supported by the right hand. If the user positions the phone on a table in front of her and performs MicroRolls with the right thumb, is it still the same interaction technique? From a body-centric perspective, these are two different techniques for two reasons: (1) with the phone positioned on the table, users probably interact with the index finger rather than the thumb; thus, the *body involvement changes.* (2) Even if users would use their thumb, it would probably slow the interaction performance down due to the *changed spatial relationship between body and device*. A body-centric approach takes these differences into account.

The body-centric analysis considers the user's body involvement and its environment apart from the technological aspects of an input technique. Consequently, two techniques that are considered as different because they involve different devices can be the same in terms of the body's involvement. For instance, direct-touch interaction on an interactive tabletop using the right index finger is according to BodyScape the same as using the right index finger to touch a mobile phone fixed on a physical table. The body's involvement – a small gesture performed on a horizontal surface using the index finger – is equal. However, the large horizontal tabletop surface, in contrast to small mobile phone surfaces, offers a largest range of input gestures thanks to its physical size.

In this thesis, I demonstrate that we need to take into account how people *use* technology from a body-centric perspective in order to *explain* why some techniques are preferred over others, to *predict* the most suitable technique for a given spatial body-device setup, and to *propose* improving alternatives. The body-centric analysis is a tool to generate and evaluate novel techniques in multi-surface environments and beyond.

1.1 Thesis Statement

I promote a change in the way we are thinking about interaction techniques. From a user's perspective, underlying technology and implementation, and their complexity, do not really matter. I argue that analyzing interaction techniques from a body-centric perspective is essential for both the understanding and design of successful interaction in multi-surface environments. I propose a new framework, called BodyScape, that identifies key characteristics of human motor control with respect to the interactive environment in order to understand and, in the future, predict human performance and comfort. BodyScape allows us to compare and evaluate interaction techniques in terms of the body's involvement in the interaction, but also to predict potential sources of interference among body movements as well as to identify unexplored combinations that may help us generate new multi-surface interaction techniques.

1.2 Research Approach

My general research strategy is based on five steps:

1. I combine informal observations about other people interacting with

technology, informal complaints in everyday discussions with other people, own experiences from living with technology, and related work from multiple research areas such as psychology, ergonomic and human-computer interaction in order to identify interaction phenomena regarding one's body movement coordination.

2. I build on existing models in order to theoretical frame and identify properties that effect a phenomenon.
3. I conduct pilot studies to identify a certain aspect of the phenomenon, I use video prototyping to create a mental walk through an interaction idea, and/or I create software prototypes and iteratively refine them.
4. I then conduct controlled experiments in order to isolate each identified key dimension and investigate their effect on interaction performance and perceived comfort.
5. As a last step, I use the identified results to refine the theoretical part to provide a strong theoretical framework that is descriptive, generative, and predictive; or I provide concrete support in design choices for interaction designers.

1.3 Thesis overview

Chapter 2—"Multi-surface Environments: Users, Tasks, and Design Challenges" I briefly define what a multi-surface environment is and propose a short scenario that illustrates contemporary work strategies in a team meeting using single user environments (e.g. laptops). I highlight how the body configuration is dynamically changing in the physical environment during the meeting and how users involve the environment to accomplish their work strategies. I discuss how these contemporary work strategies can be augmented by existing multi-surface environments and interaction techniques. I introduce the WILD room as a test-bed of my research and describe the type of users we worked with, their tasks and work strategies. I illustrate how a given work strategy can be designed in multi-surface environments using various interaction techniques and highlight the need of a theoretical framework that enables designers to compare several alternative designs.

Chapter 3—"A Body-centric Design Space for Multi-surface Interaction" I present BodyScape, a body-centric morphological approach taking into account the user's body involvement and imposed restrictions of the environment when interacting across multiple surfaces. I discuss existing interaction techniques in the light of this framework. BodyScape predicts interaction effects between two concurrent ***input movements*** which ground the actual research questions that I investigate in the scope of this thesis.

Chapter 4—"The Effect of Support on Input Motor Assemblies" This chapter investigates a set of novel bimanual interaction techniques with the support hand on multi-touch tablets, called BiPad. BiPad allows me to explore the interaction effect between body parts that are handling both support and interaction since they are shared across one arm. In particular, I investigate the effect of device support on the interaction performance and on the perceived comfort.

Chapter 5—"Interaction Effects between Input Motor Assembly and Affected Body Parts" I apply BodyScape to generate a novel interaction technique which combines two existing ***atomic interaction techniques***: *on-body touch* interaction and *mid-air pointing*. I systematically study the performance of both techniques in isolation and combination. I present guidelines for efficient placement of on-body targets on the user's body, and demonstrate that on-body touch is *affecting* several body parts. This results in interaction effects between the body movements produced by each technique.

Chapter 6—"Interaction Effects between Two Input Motor Assemblies" I present twelve mid-air interaction techniques for a zoom task within the context of pan-and-zoom navigation on large displays. Techniques vary along two dimensions of BodyScape: the way two simultaneously performed input movements are composed and the number of body parts that are involved into the interaction. I show that simultaneously involving several body parts in two input movements results in noticeable interaction effects between the two produced body movements, and that involving smaller body parts in the interaction can increase the interactive performance for certain tasks.

"Complexity is one of the great problems in environmental design."

Christopher Alexander – Professor Emeritus at the University of California, Berkeley

2

Multi-surface Environments: Users, Tasks, and Design Challenges

*Multi-surface environments incorporate multiple input and output devices, sometimes operated by multiple users, to create compound interactive environments. This chapter begins with a brief definition, followed by a scenario to illustrate the full potential of multi-surface environments. I briefly introduce the WILD room (*wall-sized interaction with large datasets*) located in the INSITU lab. I present potential users and results from studies investigating their work strategies. I conclude with a list of user requirements and interaction design challenges, which serve as the foundation for the proposed design space and presented research questions in chapter 3.*

2.1 Introduction

What is a Multi-surface environment (MSE)? Does my two-monitor setup for my laptop count? On one level 'yes'; I use one device, a mouse, to control a single cursor across both display surfaces. However, for the purpose of this thesis, 'no', because this setup acts like a divided single screen, offering only a fraction of the potential of a true multi-surface environment. The

the challenge of "multiple"

word *"multiple"* – or *"multi"* – highlights the challenge of such environments: multiple *users*, multiple *computers*, and multiple *input devices* must all work in concert. Graphical interfaces and input are *distributed* across multiple people and devices. Multi-surface environments push the dimensions of conventional desktop environments – single user, input-, and output device – by physically multiplexing them.

the challenge of "surface"

The word *"Surface"* refers to interactive surfaces or devices in multiple shapes and sizes and the technical challenge to make them *being aware of each other*, which allows the exchange of data and interaction events, and the proper presentation of the graphical user interface on differently shaped devices. It refers also to the challenge of designing interaction techniques that allow a high degree of parallelism when accomplishing tasks. Interaction designers can pick from an extensive selection of individual interaction techniques performed by using direct touch interaction, input devices, tangible interaction or whole-body interaction: (i) users can directly touch the interface displayed on a large display (see Fig. 2.1a), an interactive ***tabletop*** (see Fig. 2.1b), or a hand-held device (see Fig. 2.1c); (ii) they can use tangible objects or input devices as mediators: e.g. by pointing with a chopstick to a brain model, users can navigate in different layers of the brain displayed on the large display (see Fig. 2.1d), and write on interactive paper to display notes during a discussion in front of the large display, making them visible to everyone (see Fig. 2.1e); and (iii) they can use mid-air gestures or entire body movements to trigger commands.

the challenge of "environment"

The word, *"Environment"* refers to all real-world *actions* going on in work environments that are not mediating input but are crucial to accomplish the overall task, e.g. gesturing while talking to co-workers. It refers also to all factors of the physical world that interaction across multiple surfaces entails. Each interaction technique imposes a different spatial relationship between the user's body and the interaction devices, which has implications on how the user coordinates movements to perform input and visually perceive feedback: direct-touch requires the user to remain within a limited space in front of the stationary device, such as a large display or tabletop; portable devices or tangible objects can be operated from distance to the stationary device, but require that some part of the body is dedicated to the device support; mid-air gestures can also be operated from a distance but often involve large body movements that can lead to *fatigue* effects. Visual output distributed across large displays and interactive tables restricts the body configuration to a place next to the table where the user's eyes can perceive both displays. Visual output presented on hand-held devices requires the support arm to maintain a steady spatial relationship between the display and users' eyes.

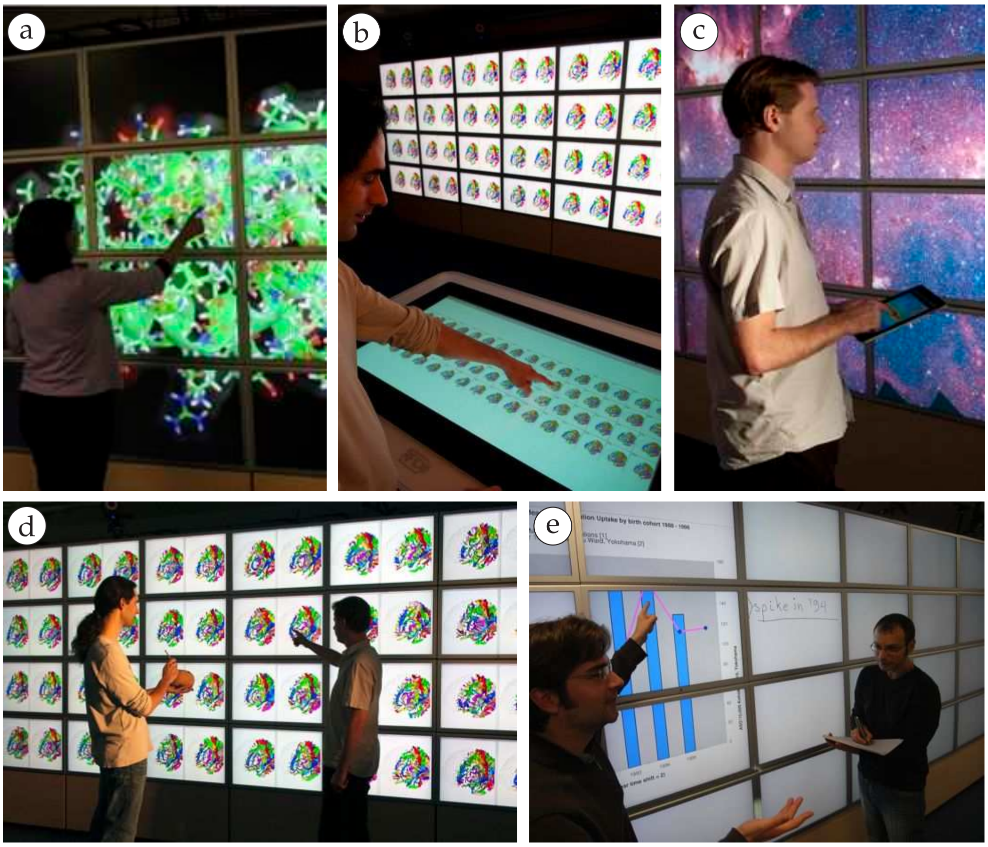


Figure 2.1: Five examples of multi-surface interaction: direct touch interaction a) on a large display, b) on an interactive table, and c) on a hand-held device; interaction mediated by d) pointing with a chopstick to a tangible brain model, and e) writing on augmented paper.

Research has explored many different ways of mediating users' input outside of the single-desktop environment. Each technique by itself entails certain body configurations on the user. For instance, the user in Figure 2.1e is standing next to his colleague in front of the large display and writes on ***augmented paper***. This implies that the paper notebook needs to be held by the user. An alternative is that he writes on augmented paper while being seated at a table. The paper would be then fully supported by the table and leaves the non-dominant hand free to perform further actions, e.g. like when combining pen-and-paper interfaces with mobile phones [Tsa12]. On a higher level, some might say that these two *ways* of interacting are the same interaction technique. However, the difference in users' body configuration can affect the user performance or comfort while writing on paper. Interaction in multi-surface environments is more physical than on conventional desktop machines due to the involvement of the body and the environmental factors, e.g., the spatial distribution of surfaces and the presence of other users. The design of interaction techniques faces a trade-off between the user's body involvement into an interaction and restrictions due to the body configuration imposed by the environment. The following scenario emphasizes these factors by illustrating a conventional collaborative work setup.

2.1.1 A Real-world Scenario

In this scenario, three co-workers have a meeting about a research article that has soon to be submitted to a conference. I emphasize *the changing body configuration* and *impact of the surrounding environment* during their work practices and present related work that can augment interaction and data sharing during meetings.

> *Françoise is head of a research lab in computer science and has a meeting in five minutes with two of her team members in her office. She accesses the latest document version of the article on her desktop machine and prints it out in order to carry it over to the round table in her office.*

Changing Body Configuration

> *Returning back from the printer in another room, her two collaborators, Jean-Michel and Anne are already* ***seated*** *at the table. Both brought their laptops and access the latest version of the article on their screens. They start discussing the paper, in particular one of the graphical illustrations.*
>
> *When the conversation arrives at the "abstract" section of the article, Françoise* ***walks*** *from the table to her desk in another corner of the room and tells Anne and Jean-Michel to join her. She shows on the large screen that is hooked up to her desktop computer a new version she was recently working on.*
>
> *Anne and Jean-Michel* ***stand*** *next to Françoise who is seated and bend down to see the displayed abstract. After giving Françoise feedback on her work, they all walk back and continue the meeting seated on the table.*

Anne, Françoise, and Jean-Michel started their meeting seated at a table. When Françoise wanted to share the "abstract" that is located on the desktop machine at another desk, they walk over and stand next to Françoise to discuss her work.

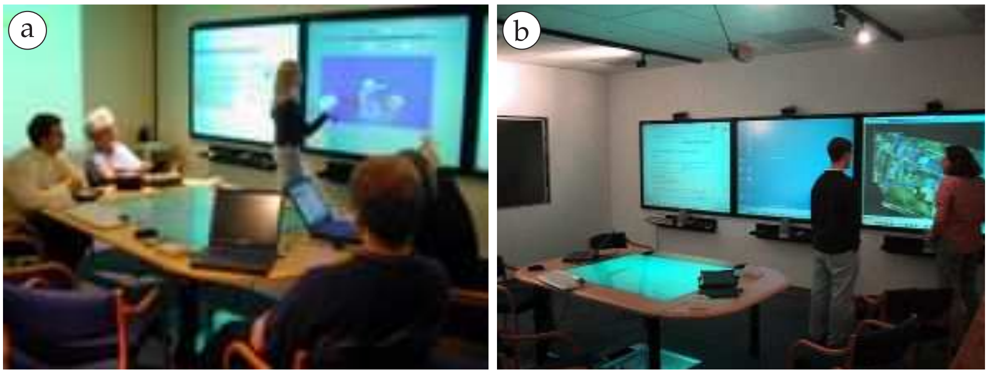


Figure 2.2: Two types of co-worker's body configuration during collaborative work in the iRoom [JFW02]: a) one person standing at the large display, everyone else seated; and b) two users discussing in front of the large display.

In multi-surface environments, such as *Collab*, Françoise can share her data by moving it over to a large display wall [Ste+87]. Using *Dynamo* [Iza+03], Françoise can place her data on large interactive surfaces that are accessible to everyone; and she can take away feedback and notes taken during the meeting. Systems such as *pointRight* enables her to control any device in the room using keyboard and mouse which enables Françoise to have complete control over the environment while remaining seated at a meeting table [Joh+02].

Wigdor et al. propose a table-centric control over the environment using an interactive table as input device. Another possibility is to use Rekimoto and Saitoh's *Augmented surfaces* [RS99] which allows Françoise to move graphical objects from a laptop onto a table or wall surface and among laptops in a workspace. Using *UbiTable* [SER03], she can spontaneously walk-up-and-use an interactive table to share her work and to take notes.

Multi-surface environments, such as *WeSpace* [Wig+09] or the *iRoom* [JFW02], support meeting work practices of small groups by mixing personal laptop devices with shared large horizontal or vertical displays. Users work practices imply various body configurations, as illustrated with figure 2.2 in the iRoom: a) one user is standing at the large display and the co-workers are seated; and b) two users are discussing next to the large display.

Impact of the Surrounding Environment

> *Françoise wants to discuss the general flow of the research article and arranges the 10 pages of her print-out on the table. They start discussing*

the arrangement of sections and figures in the article and support their statements by gesturing with their hands on the print-out.

Anne gets an idea to simplify a figure and sits down to create a quick draft of her idea. When she finished, she turns her laptop towards the others and points with her index finger to the specific area in the figure where she applied changes.

Françoise makes use of the large table to spread out her paper and arrange them in a certain way. Streitz et al. suggest that multi-surface interaction design, in particular for cooperative work of dynamic teams and changing needs, is based on an integration of information and architectural space. They propose digital furniture with build in displays, such as *InteracTable*, *DynaWall*, and *CommChairs* [Str+99] that allows the dynamic creation and allocation of workspaces in different parts of the room.

Anne uses spatial awareness to point out crucial information for the conversation. Gesturing has been demonstrated as very important influencing thought, understanding and creativity [GB10]. Baudel and Beaudouin-Lafon investigated free-hand gestures for controlling slides during a presentation [BB93]. They point out that the use of "unnatural" control gestures is crucial in order to enable the system to segment control gestures from people's natural way of gesturing (e.g., while talking or pointing out an object).

In summary, multi-surface environments seek to augment collaborative work practices by offering a set of interaction techniques involving several input devices operated by multiple body parts. They support work settings that are insufficiently assisted by traditional window-icon-menu-pointer (WIMP) graphical user interfaces[1]. For instance, nurses or class room teachers work often in collaboration with multiple people; their working environment is mobile and dynamically changing. When multiple people work collaboratively, they distribute interaction and scatter their data in the environment. Research in multi-surface environments seeks to find technological solutions to make people's work strategies more efficient. In order to support these work strategies, we need to design interaction techniques that account for users' changing body configuration and the relationship between the users' body and the interactive environment.

[1]Graphical user interfaces were designed at Xerox PARC in the 70's to support the work setting of executive secretaries, and are still in use today.

2.2 The WILD Room

The INSITU lab offers multiple interactive surface devices that are either stationary or portable with either horizontal or vertical screen orientation. The *Wall-sized Interaction with Large Datasets (WILD) room* (see Fig.2.3) is centered around a very-high resolution vertical wall-sized display assembled by 32 monitors mounted in a 4×8 grid. The 32 monitors are connected to 16 computers located in the cluster room, and are controlled with a *frontal* computer in the main room. A VICON motion capture system uses ten infra-red cameras around the room to track reflective markers that can be mounted on the users' body, or on arbitrary rigid objects and devices. An interactive table offers a large horizontal interactive surface, and several tablets and phone-sized multitouch hand-held devices can also be used. Figure 2.3 illustrates an outline of the spatial arrangement of devices in the room. Further technical specifications are described in Appendix A—"WILD: Technical Equipment".

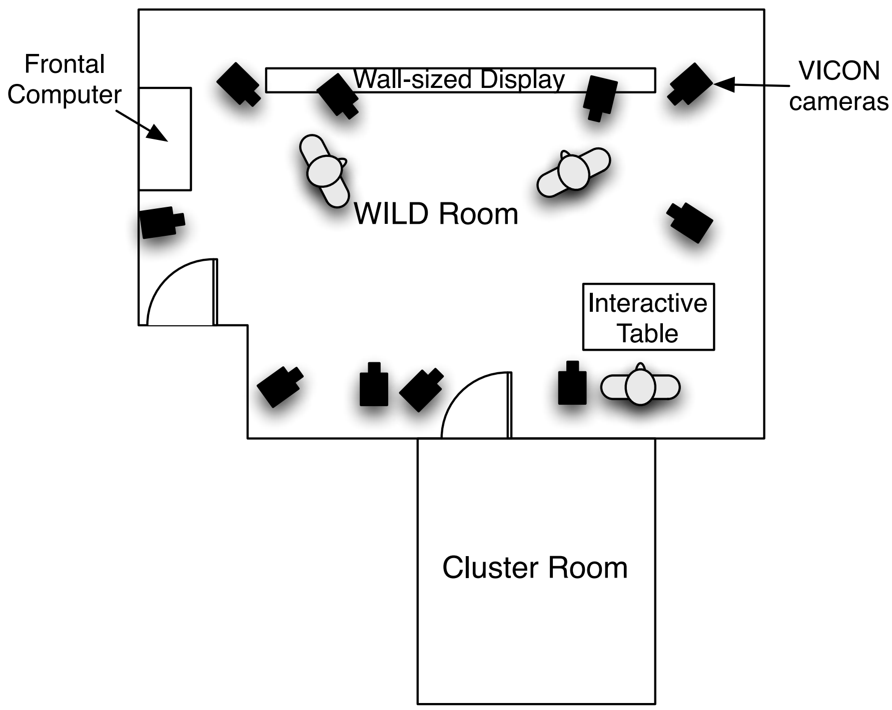


Figure 2.3: An outline illustrating the spatial arrangement of the wall-sized display, the interactive table and the ten VICON tracking cameras in the WILD room.

2.2.1 User-centered Research Approach

The WILD room serves as a testbed for the research I present in this thesis. The overall research strategy of the INSITU research group with WILD is to push the limits of technology – both hardware and software –, and to ground the design process on the work practices of *extreme users* who work with very large datasets and can benefit from using a wall-sized display ([Bea+12]).

Users

Researchers from the *Paris-Sacklay campus* in astrophysics, particle physics, chemistry, molecular biology, neuroscience, mechanical engineering and applied mathematics were chosen as extreme users of the WILD platform. They were invited to participate in an initial "show-and-tell" workshop where they presented specific examples of their tasks along with their data analysis processes and tools.

Tasks

The group of microbiologists showed their data and explained their study on how one molecule docks with another. Some of them work with a large molecular model, some work with interactive 3D models of molecules, and others require access to online databases, websites, and research articles. In the WILD environment, they envision to smoothly shift among different representations of each molecule and being able to transfer and arrange them across displays, so that multiple colleagues can work in the same room or even collaborate remotely.

Work Strategies

From these sessions, we identified four work strategies for these groups of researcher with the WILD platform:

Navigation Some researchers, such as the biologists and astrophysicists, have to visualize very large datasets such as simulations of molecules with hundreds of thousands of atoms, or a gigapixel image of deep space taken from a space telescope and containing thousands of galaxies. They require interaction techniques to navigate, e.g. by ***panning*** and zooming, into those images.

Comparison Other researcher groups, such as the neuroscientists and astrophysicists need to compare large numbers of related images, such as pathological brain scans or observations of regions of the sky at different wavelengths. They benefit form a large scale visualization platform and need to be able to move in front of the large display in order to see their data from different perspectives.

Juxtaposition All researchers need to layout some juxtapositions of data in various forms, and from different sources, such as articles, raw data tables, formulas, graphs, etc..

Communication They also need to communicate with remote or co-located collaborators requiring to share and collaboratively explore their work.

2.3 Interaction Design Challenge

The above described work strategies of potential WILD users highlight the need for location independent input techniques. Users need to physically move in front of the wall-sized display in order to explore their data. This can be achieved by designing several alternative interaction techniques using three types of input: by (1) using portable devices, such as touch-enabled hand-held devices or 3D mouse devices, by (2) using mid-air free-hand gestures or by (3) combining both portable devices with mid-air gestures. The problem is how to select a suitable interaction technique from the variety of possible designs.

To illustrate the complexity of this issue, I discuss various alternative interaction techniques for the high-level user task of arranging visual data on wall-sized displays. WILD users need to arrange data as part of their comparison and juxtaposition task. They could apply two different interaction techniques: *drag-and-drop* and *pick-and-drop*. Drag-and-drop [Col+05] simply moves the graphical representation of the data to another position on the wall: it remains visible on the wall during the entire interaction. Using pick-and-drop [Rek97], users can pick up the object, remove it from the wall-display, and drop it back to another position (or even on another surface), making the object reappear. The advantage of pick-and-drop is that multiple objects can first be picked up before being dropped somewhere else.

In order to design a pick-and-drop interaction technique, the first step is to understand what users need to control within a task. Figure 2.4 illustrates the high-level ***task*** "pick-and-drop" has two ***subtasks***, "pick" displayed data up and "drop" it somewhere else. Each subtask can be further decomposed into two **control tasks**: to pick up data, user need to point to it and apply a "pick" command; to drop data, users need to point as well and apply a "drop" command.

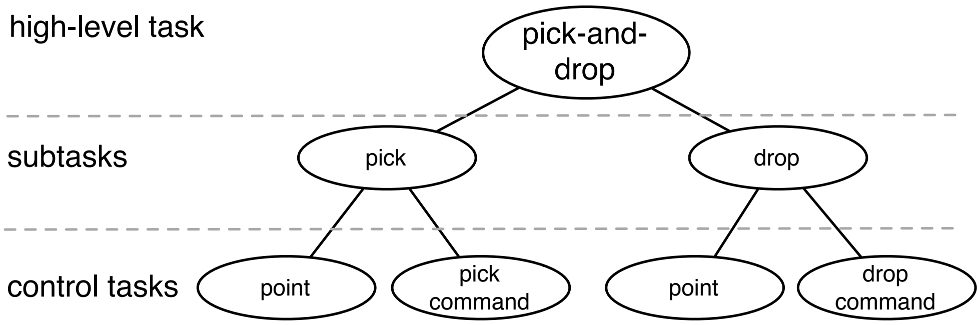


Figure 2.4: The decomposition of a high-level "pick-and-drop" task into "pick" and "drop" subtasks. Each subtask is further decomposed into control tasks.

I consider three alternative designs of mid-air interaction techniques for a high-level "pick-and-drop" task: (1) *pick-and-drop projector*, (2) *pick-and-drop pointer*, and (3) *pick-and-drop fingers*.

Pick-and-drop projector is inspired by Boring et al.'s *Touch projector* [Bor+10] where users use a mobile phone as a "magic lens" [Bie+93]. When holding the device between their eyes and a large display, users see virtual objects from the distant display on the phone's display and can pick them up by touching them. Picked up data then turns into a small icon on the right frame of the phone's display. By pointing the phone to another area of the screen and touching the icon, users can drop their data at another position.

With *pick-and-drop pointer*, users can perform mid-air pointing with a mobile phone device and control a cursor on the wall display. Touching the phone's surface while the cursor on the wall-sized display is located within a virtual object picks the object up and represents it as small icon on the phone's display. When users touch this icon while the cursor is positioned somewhere else on the wall-sized display, they drop their data.

Pick-and-drop fingers is a free-hand gesture interaction technique. Users perform mid-air pointing using the index finger of the dominant arm in order to control a cursor on the wall display. While the cursor is positioned above a virtual object on the wall-sized display, users can assign the virtual object to either the middle, ring or pinky finger by touching the appropriate finger using their thumb.

The question is then *How to decide which technique is best suited?* The answer depends on two subsequent questions: (1) how well is an interaction technique suited within a *sequence of interactions* that users perform and (2) how suitable is the *assignment of body parts* to achieve maximal interaction performance and comfort?

Interaction sequence If users use pick-and-drop projector to drop data at a certain position and then need to use an augmented pen-and-paper interface to further interact with this data, users might feel it impractical to *switch between devices*. If users use the free-hand pick-and-drop fingers technique to arrange a virtual object and then require a hand-held device to change properties of the object, it would result in an uncomfortable *depositing and relocating of the hand-held device*. It is therefore important to investigate an interaction technique within the context of the overall task [ALM04].

Assignment of body parts Pick-and-drop projector requires the user to hold a portable device and might result in *fatigue* effects. Pick-and-drop fingers involves two concurrent body movements, index finger pointing and thumb-finger pressing. One movement might interfere with the other. Pick-and-drop pointer requires that users switch attention: they visually attend the wall-sized display when "picking" and have to switch visual attention between the wall-sized display and the mobile phone when "dropping".

2.4 Conclusion

Multi-surface environments push the dimensions of conventional environments and support collaborative work across multiple devices. Interaction and graphical user interface are spatially distributed: Interaction techniques can involve multiple body parts and can affect the body's configuration, e.g. when users need to support devices or remain within arm-reach position to a touch-enabled stationary display; Graphical output is distributed as well and in order to visually perceive output, users need to turn the head, their torso, or even the entire body in case the visual output is on their back.

In such environments, users also have specific work strategies to accomplish a high-level task such as comparison and virtual navigation. In order to design suitable interaction techniques, to compare their trade-offs, and to systematically predict and investigate possible problems (i.e., during simultaneous or consecutive input movements), we need to decompose the overall task into several subtasks, each performed at *different moments in time*. Subtasks can be decomposed into elementary control tasks that are all performed *at the same time*. For instance, when user perform the high-level "pan-and-zoom" task, they either perform "zooming" or "panning" at a moment in time. The subtask "zooming" itself can be further decomposed into the simultaneous execution of two control tasks, "the center of zoom" and the "zoom direction" (zoom in/out). Figure 2.5 shows the decomposition of a high-level user task into several control tasks.

Similarly, an interaction technique can be decomposed into several atomic interaction techniques that each allows to perform one control task. The question is then how can we describe, compare and generate an individual atomic interaction technique for a given subtask? How can we analyze various simultaneously performed body input movements? And how can we predict if two interdependent body movements interfere with each other, possibly causing discomfort and slowing down the interactive performance?

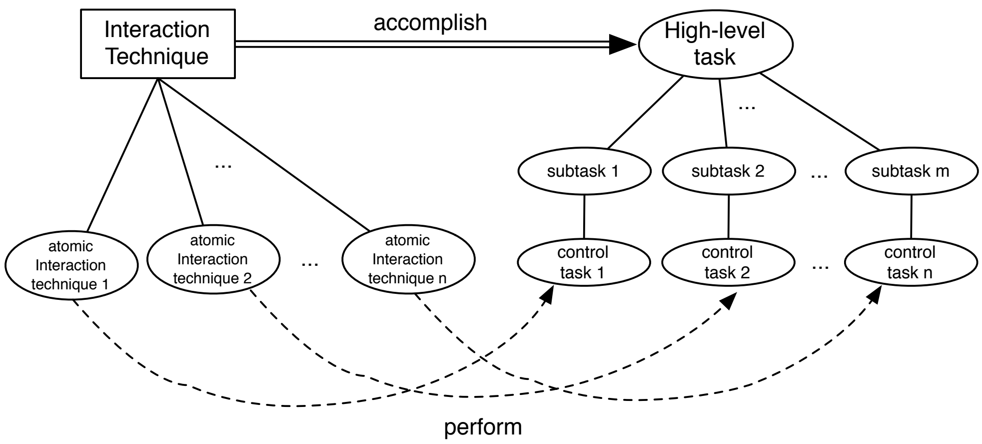


Figure 2.5: The decomposition of an interaction technique used to accomplish a high-level task. Interaction techniques can be decomposed into atomic interaction techniques each responsible for one control task.

In the next chapter, 3—"A Body-centric Design Space for Multi-surface Interaction", I address these questions with a body-centric analysis of interaction techniques. This approach isolates body interaction movements from the technical implementation of interaction techniques, by focusing on the user's body involvement and the restrictions that are imposed by the interactive environment onto the user's body configuration. This enables us to compare interaction techniques composed of atomic interaction techniques performed at the same time[2].

In the following chapters, I will present how this approach can be applied to systematically investigate practical problems and inform the design of body-centric interaction techniques:

- Chapter 4 introduces a novel interaction technique for bimanual interaction while holding a tablet device. I investigate how shared support and interaction across one arm affects interactive performance and perceived comfort.
- Chapter 5 presents a systematic analysis of interaction techniques combined by two atomic interaction techniques, on-body touch and mid-air pointing, and investigates how these two techniques require to perform body movements that interfere with each other.
- In chapter 6, I present twelve techniques for pan-and-zoom navigation tasks on large displays and demonstrate that it is crucial to consider which body parts are involved into the execution of an atomic interaction technique, as well as their level of interdependence.

[2] I will discuss how to account for sequences of interaction technique in future work (chapter 7)

"[...] we are in a new golden age for HCI. [...] the state of current technology and the spirit of the Maker Movement suggest a means for making progress on one of HCI's oldest structural problems: how to ground the field, accelerate its progress, and make it cumulative by fashioning theories and incorporating them into practice."

Stuart Card – (Consulting Professor at Stanford University)

3

A Body-centric Design Space for Multi-surface Interaction

I propose a body-centric approach to classify interaction techniques for multi-surface environments in terms of level of body restriction. I decompose the user's body in groups of body limbs associated with specific roles while performing an atomic interaction technique. I call these groups input and output motor assemblies. More complex tasks are accomplished by combining atomic interaction techniques, thus involving multiple motor assemblies: they can be in series, in parallel, or overlapping with each other. Based on the physical relationships between the body and its environment, I introduce a scale of body restriction and its impact on the compatibly of atomic interaction techniques when performed simultaneously. I present the body-centric design space, illustrate it with examples, and discuss the implications for the design of multi-surface interaction. I conclude with three research questions that I investigate in chapters 4, 5, and 6.

3.1 Introduction

The distribution of interaction and graphical user interfaces (GUI) in multi-surface environments poses challenges in the design of suitable interaction techniques. Previous work explored (i) gestural interaction *on* interactive surface such as hand-held devices [RB10], tabletops [Mic+09], or large displays [MLG10]; (ii) gesture-based interaction *with* hand-held devices such as device shaking [WMH07] or device deformation [SPM04]; and (iii) interaction *between* devices, e.g. by a through-body transmission paradigm where users touch both a tabletop and a large display to transmit a virtual object [WB10]. Body-tracking systems allow the interaction *without any device* at all, using free-hand gestures [Zig+10; HRH12] or the entire body [DSK09]. New sensors even detect precise touches on the user's body [HTM10; Lin+11]. And all these possibilities can be combined together to create novel interaction techniques.

This previous research work mostly focus on new technologies and ***point designs***. They introduce particular instances of novel interaction techniques that are not grounded in a particular problem at hand. Without theoretical grounding, it is difficult to compare existing techniques and to reason about their trade-offs. We need a systematic ***framework*** that supports design decisions and research analysis; that helps reasoning, comparing and creating appropriate techniques for the problem at hand; we need to be able to discard possibilities leaving a manageable set of techniques that can eventually be tested.

We have on the one hand a multitude of different devices, sensors and possibilities to *track* a specific input action of the user, and on the other hand a limited manageable number of physical actions that a human body can perform. Whereas devices rapidly evolve and change in size and shape, I argue that a *body-centric approach* that analyses users' ***motor*** capability provides knowledge that can be generalized to a number of user-device setups. Moreover, since whole-body interaction has become an active research area in recent years, knowledge about motor capabilities can provide implications for *device-less* interaction techniques. In this chapter, I introduce a body-centric ***morphological design space analysis*** for studying interaction in multi-surface environments. I present related taxonomies which inspired my work and introduce a novel body-centric perspective illustrated with examples. I conclude with a discussion regarding design implications.

3.2 Related Work

The first morphological approach for modeling input devices was proposed by Mackinlay et al. [MCR90] and later improved by Card et al. [CMR91]. They decompose input devices into a *vocabulary of primitive movements*, e.g. linear and rotary, and propose three *composition operators*, i.e. *merge*, *layout* and *connect*. This approach allowed them to create a multi-dimensional parametric space made of all the combinations of primitive movements in each direction of a three-dimensional space. This morphological approach helps finding abstractions for generating a design space. It supports the systematic testing of novel device designs and provides a way to reason about their effectiveness. Later, it has also been successfully applied to structure gestural interaction techniques based on accelerometers [SBC11] and menu systems [NHB09].

This design space also inspired an interesting study in the context of my work, demonstrating that the assignment of muscle groups to operate an input device is a critical factor [ZMB96]. Input device designers have to trade-off between using different muscle groups: small muscle groups have more dexterity; larger muscle groups used *together* offer a greater range of movement. But interaction in a multi-surface environment results in a more complex situation where we can (1) assign multiple muscle groups into an atomic interaction technique and (2) combine multiple atomic interaction techniques to more complex tasks. Examples of atomic interaction technique include: mid-air pointing on large displays with the arm [VB05], a mid-air gesture with both arms to control a continuous value [Bai+12; HRH12], a foot tap to answer an incoming call [Ale+12] and using the entire body to control the zoom direction by shifting one's balance point [DSK09].

While Shoemaker et al.'s [Sho+10] pioneering work introduced high-level design principles and guidelines for body-centric interaction on large displays, using a lower level morphological approach would allow us to describe atomic interaction techniques by their level of *body involvement* and to analyze their compatibility when combined with other atomic interaction techniques (i.e., to control a more complex task). This addresses a gap in the research literature with respect to the theoretical grounding of atomic interaction techniques and their *combination*, allowing us to analyze the complex nature of distributed input and output in multi-surface environment in terms of the implications on the users' body configuration. It provides a means to systematically analyze which body part is responsible for which control task and to detect potential interferences between input and output related body movements.

3.3 BodyScape: A Framework for Body-centric Interaction

In most human-computer dialogs, the human body must perform two coordinated movements: (1) perform a physical action to mediate input and (2) move the head into a position to perceive visual output[1]. Input is commonly *directed* towards virtual objects in the graphical user interface.

body-relative vs. world-fixed input and output

In multi-surface environments, input and output are physically distributed in the environment: the graphical user interface and the performed input movements are either *fixed* on stationary devices, or *relative* to the body (e.g., when using portable devices); the input can be performed using direct touch or mid-air gestures from a distance. This *spatial relationship between body, input and visual output* has implications for the *degrees of freedom* of the users' body configuration for a given interaction technique and has implications for the assignment of further body input movements performed in parallel.

On a finer level of granularity, we need to also consider how the body coordinates movements internally. When the body performs a complex multi-joint movement, e.g. reaching out for a glass of water, it involves a group of ***body limbs***, e.g. the dominant upper arm, forearm, palm and fingers, which together form a ***kinematic chain***. Each limb is part of a chain: limbs of the upper and lower extremities are either proximal (close) or distant from the torso with respect to other limbs: e.g., the forearm is more ***proximal*** than the palm and the fingers are more ***distal*** than then upper arm. According to the kinematic chain, each body limb contributes to the overall performance of a movement because it is physically connected to nearby body limbs by joints: movements produced with a distant limb, e.g. the forearm, are accompanied by movements of all proximal limbs with respect to the forearm, e.g. the upper arm. Following Guiard's terminology [Gui87], I define body limbs that are involved in the relevant movements of a human-computer interaction as ***motors***. For instance if the dominant arm performs a mid-air gesture, it consists of a set of motors within a kinematic chain: upper arm, forearm, palm, fingers and thumb.

BodyScape is a ***body-centric framework*** that builds upon these concepts to describe interaction techniques in terms of the body's involvement in the interaction and the level of body restriction imposed by the environment. The basic primitive is a ***motor assembly***, which is a group of motors (movable body limbs) that the user adjusts to control input or view output.

[1] I do not consider auditory feedback since sound perception does not depend on the body position in most cases, except in rare environments featuring finely tuned spatial audio.

3.3.1 Motor Assemblies

Input Motor Assembly

As seen in chapter 2, I define a high-level user task as a set of subtasks that users accomplish *in sequence*, e.g. “pick” and “drop”. Each subtask can again be divided into one or more control tasks that users perform *simultaneously*. For instance, when the users select a tool from a palette, they need to perform two control tasks: (1) pointing to the tool and (2) performing a selection command. The user accomplishes each control task with an interaction technique that consists of one or more atomic interaction techniques, each one involving exactly one ***input motor assembly*** to produce the required movements (see Fig. 3.1)

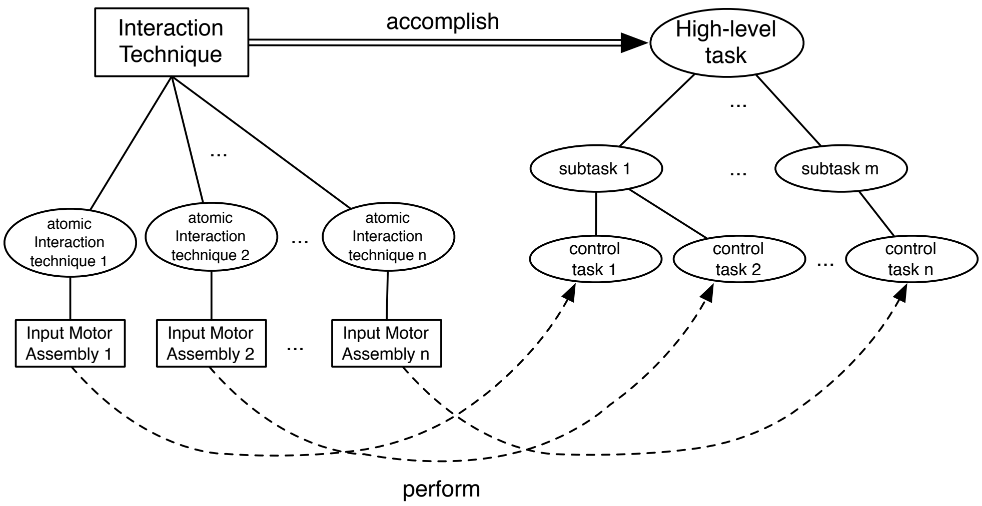


Figure 3.1: Mapping of interaction technique to user task: each input motor assembly controls the input of a control task.

An *input motor assembly* is thus a group of motors that handle a control task, for example, when performing a mid-air pointing task the motors of the dominant arm form an input motor assembly for this control task.

The implementation of a mid-air pointing technique directly affects which motors are contained in the corresponding input motor assembly. For example, a mid-air pointing technique[2] could be implemented so that the cursor position is controlled by (i) the forearm, (ii) the index finger or (iii) the ori-

[2] Mid-air pointing techniques are commonly implemented by ***ray-casting*** from a specific limb on the dominant arm, e.g. the index finger, to the display using vision-based tracking systems.

entation of a hand-held device. Consequently, the input motor assembly is either (i) upper arm and forearm or (ii) + (iii) the entire dominant arm. Note that due to the kinematic chain principle, movements of distant extremities affect movements of the proximal extremities.

Output Motor Assembly

An **output motor assembly** is a group of motors that is responsible for bringing the eyes into an appropriate position to enable visual perception of output. Each atomic interaction technique may include one or more input motor assemblies – one per atomic interaction technique but only one output assembly. The output motor assembly can be *the neck* when the visual output is positioned within a small angular rotation of the head; with larger angle, the output motor assembly involves motors from the head down in the interaction: it might require the shoulders, the hips or even the entire body to turn as well, e.g. when visual information is presented in the back of the user.

3.3.2 The Spatial Relationship between the Body and the Surrounding Interactive Environment

We move and coordinate input and output motor assemblies with respect to the environment: our senses and motor skills are in a constant dialogue that exchanges feedforward information [CR11; LRE06; Gib86] about possible actions and potential body postural adjustments, and feedback about what has changed in the world. Our body is our *point of reference* and our movements are coordinated within the *frame of reference* provided by our bodies.

Input and output motor assemblies are coordinated with respect to the physical environment, e.g furniture and interactive devices, and with respect to the virtual environment, e.g. graphical objects. For example, when performing mid-air pointing, the pointing arm remains in a steady position oriented towards a virtual object displayed on a stationary display. In contrast, pointing on a touch-enabled mobile device is coordinated relative to the body. Even though the body changes position and orientation, the touch gestures – the action of the thumb – remains identical with respect to the device surface.

Input Motor Assembly: Body-relative vs. World-fixed

Input motor assemblies can be coordinated relative to the body (***body-relative***) or relative to the world (***world-fixed***). These kind of relations between the body and the environment have already been introduced in previous research [Bil+98; Fei+93], but they were never applied to a body-centric description of interaction as we do with BodyScape. Figure 3.2 illustrates both types of relations as a user interacts with a mobile phone. Figure 3.2a is body-relative: the dominant hand supports the device with the palm and interacts with the thumb (the input motor assembly) on the interactive surface. The device support maintains a consistent relationship between the users' thumb and the device surface. However, this relationship between the thumb and the surface is not affected by the user's position and orientation within the room. Thus, the input motor assembly coordinates its movements relative to the body.

body-relative

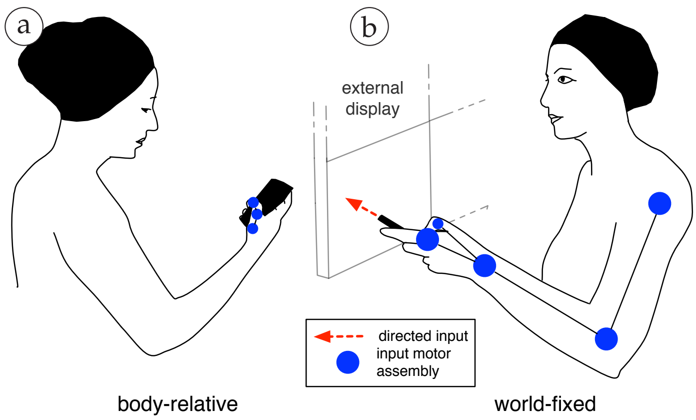


Figure 3.2: input motor assembly coordination: a) body-relative (thumb interacts on a surface that moves relative to the body) and b) world-fixed (user needs to stabilize input motors towards a fixed target).

Figure 3.2b is world-fixed: the user points with the mobile phone to a *fixed* target on an external display. The input motor assembly now consists of the entire arm which must remain stable with respect to a fixed external object on the display. This object does not adjust its position according to the user's body movements, e.g. a jittering arm, which makes it harder to coordinate world-fixed than body-relative movements.

world-fixed

Output Motor Assembly: Body-relative vs. World-fixed

Visual output may appear relative to the user's body, e.g. skinput projects an image on the user's arm [HTM10], or, more commonly, fixed in the world, e.g. on an external display. The visual output motor assembly seeks to establish and maintain a spatial relationship between the eyes and the visual output. This can be achieved by *turning the head* and – if that is not sufficient – in *turning the torso* or the *entire body*. The user can coordinate output motor assembly movements *relative to the body* or *fixed in the world*. Figure 3.3 (a) shows a user interacting with a mobile phone: neck, shoulder and arm contribute to maintaining a spatial relationship between the eyes and the display. However, this relationship is not affected when the user turns and moves within the environment. The output motor assembly can also coordinate movement towards a fixed external display in the environment. Figure 3.3 (b) shows the user interacting with the wall using a mobile phone. In this example, the neck allows the head to turn towards the virtual object on the external display.

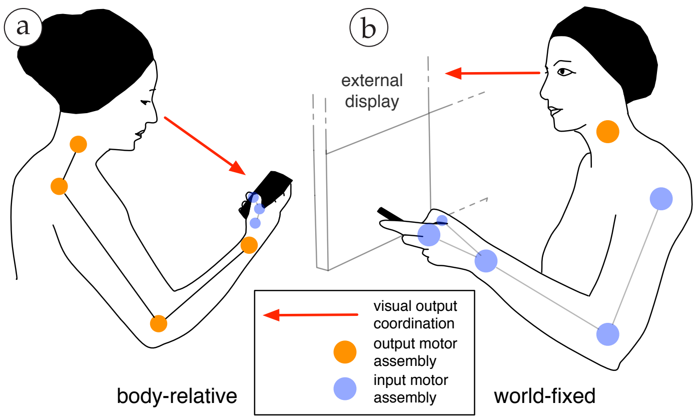


Figure 3.3: Output motor assembly coordination: a) body-relative (output motor assemblies keep steady relationship between head and display) and b) world-fixed (neck keeps head oriented towards fixed object on external display.)

3.3.3 Body Restriction in The Environment

Interaction techniques can then be described in terms of input and output motor assemblies and their coordination with respect to the interactive environment (body-relative vs. world-fixed coordination). These factors have implications for the overall configuration of the body when performing a particular interaction technique. They can restrict the configuration of the body to a specific position and orientation within the environment with varying degree. Thus, we can classify atomic interaction techniques according to the level of restrictions they impose on the body and can inform about the compatibility of two interaction techniques.

Input motor assembly We can analyze input motor assemblies along a dimension, from body-relative to world-fixed. For instance, when users carry handheld devices, they freely interact with them from anywhere. Some technologies even make an input device unnecessary: PUB [Lin+11] enables on-body touch interaction and PinStripe [Kar+11] detects pinching and rolling gestures on the users' clothes. Hand-held devices, on-body touch, and clothing interaction all allow *body-relative input*. However, when a hand-held device or a limb is tracked in 3D, e.g., mid-air pointing on a display, the arm must be held in a specific position relative to a target fixed in the world.

In the case of world-fixed coordinated movements, we need to also distinguish between *mid-air* and *touch*. First, body movement coordination is affected by the physical connection with the environment [DL04]: an established touch connection can reduce body sway of complementary body actions; and second, touch-based techniques with world-fixed objects, e.g. a target on a tabletop, requires that the body remains within arm reach from the stationary device. Mid-air interaction restricts the body in position and orientation as well but provides a wider range in front of the device within which the user can operate. For body-relative input, mid-air or touch does not make a difference regarding body restrictions. It plays however a role in the feedback about users' actions: for instance, when users perform mid-air gestures by pointing to their hip in order to invoke a command [Sho+10], they perceive proprioceptive feedback about their body position but they can need visual attention to verify their action; when users point and touch the hip by using on-body touch detecting technology, they perceive – in addition to proprioceptive feedback – tactile feedback about their action [GIB62].

Output Motor Assembly Multi-surface environments are inevitably affected by the separation of the visual outputs [SB05; TC03] which are divided across multiple devices. Users must shift their gaze and switch their attention to the output devices that are relevant to their current task. Visual output *relative to the body* is independent of the user's location in the environment, e.g. the screen of a hand-held device. It does not constrain the user's location and orientation, but it affects the body's internal posture, since it includes a larger part of the body into the output motor assembly. Figure 3.3)a shows that mobile devices require that the support – the arm, shoulder and neck – remain in a position that allows the visual perception of feedback on the mobile phone. Conversely, visual output *fixed in the world* requires the user's head to be oriented towards its physical location, e.g, on a wall projector. When output is fixed in the world, users' locations and body configurations are restricted to the positions that allow them to see the visual output most effectively.

The Continuum The input and output coordination dimensions define a qualitative scale of restrictions that a given interaction technique imposes on the user's body (see Fig. 3.4). These measure the degree of freedom that are required as the whole body moves to accommodate a particular interaction task. As illustrated in Fig. 3.4, constraints from the *Input* dimension have more impact on body restriction than those from the *Visual Output*. In particular for input fixed in the world, *touching* a surface is more constraining than moving the hands in the air (*mid-air*). Consequently, in Fig. 3.4, the horizontal division between mid-air and touch of input that is coordinated relative to the body indicates that mid-air or touch makes no difference with respect to the body restriction; Conversely, the vertical division between mid-air and touch interaction of input that is coordinated fixed in the world shows that mid-air is less restricting than touch.

Note that *body restriction* is not necessarily negative. For example, assigning a dedicated fixed display area for each user in a collaborative multi-surface environments restricts their operating area, which could reduce visual occlusions, collisions, conflicts or privacy concerns [SCI04].

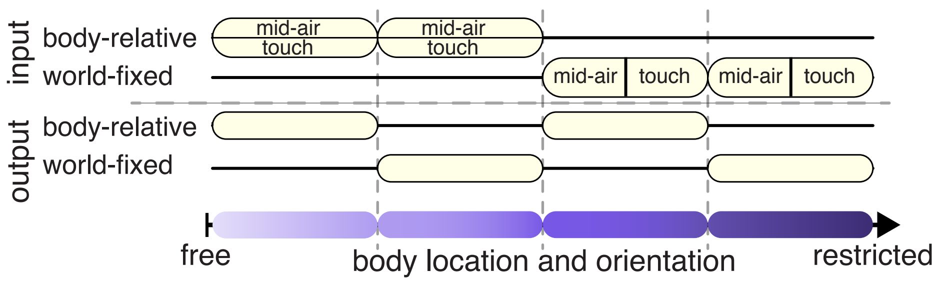


Figure 3.4: Different combinations of body-relative and world-fixed input and output motor assemblies affect the orientation and location of the body along a continuum from free to fully restricted. For world-fixed input motor assemblies, mid-air input is more restrictive than direct touch input. For body-relative input motor assemblies, mid-air or touch makes no difference with respect to the body restriction.

Classifying Atomic Interaction Techniques using BodyScape

Figure 3.5 illustrates the BodyScape dimensions with several *body-centric* atomic interaction techniques. These atomic interaction techniques allow users to perform elementary actions like moving a cursor or selecting a target.

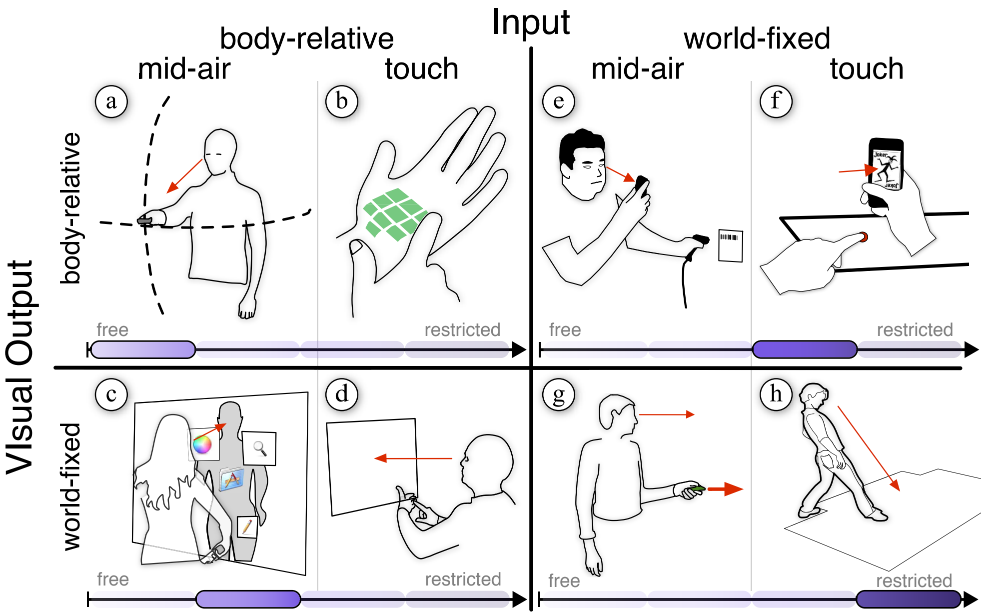


Figure 3.5: Atomic body-centric interaction techniques according to their input and output characteristics: a) Virtual Shelves [LDT09] ; b) Skinput [HTM10] ; c) Body-centric interaction techniques for wall-sized displays [Sho+10] ; d) PalmRC [Dez+12] ; e) Scanning an object with feedback on a mobile device ; f) Pick-and-Drop [Rek97] ; g) Mid-air pointing ; h) Multitoe [Aug+10] .

Upper-left Quadrant: Body-relative Input – Body-relative Output The least restricted combination is obviously when both input and output are relative to the body, since the user can move freely in the environment while still being able to interact and get visual feedback.

VirtualShelf [LDT09] permits short-cuts on a mobile phone by orienting the device in mid-air within a spherical space in front of the user (Fig.3.5a). The limbs of the dominant arm form the input motor assembly. *Armura* [HRH12] extends this approach with wearable hardware that detects mid-air gestures of both arms and projects visual feedback on the user's body. Skinput [HTM10] (Fig. 3.5b) enables touch input on the users' forearm and provides body-relative visual output with a projector mounted on the user's shoulder. The non-dominant arm, the input motor assembly, points to a target projected on the dominant forearm, which is part of the output motor assembly.

Upper-right Quadrant: Relative Input – Fixed Output Fixing the output in the environment constrains user's orientation and, if the distance to the display matters, her location. Shoemaker et al. introduced body-centric interaction techniques for large displays [Sho+10] where the user select tools by pointing towards body parts, e.g. the stomach, and pressing the button of a hand-held device. The pointing arm forms the input motor assembly. The user's shadow is displayed on the wall display indicating the location of the tools on the body. In Krueger et al.'s *Videoplace* [KGH85], users interact using their entire body and see the feedback, a virtual silhouette of their own body, on an stationary display. Thus, it requires the user to remain in a restricted body configuration in front of the screen (Fig. 3.5c).

PalmRC [Dez+12] (Fig. 3.5d) allows free-hand operations on a TV set. Users press imaginary buttons on their palm [GHB11] and perceive visual feedback on the fixed TV screen. The pointing arm is the input motor assembly. However, since the dominant arm is pointing to the other hand, this interaction technique introduces the need to distinguish between body parts that are *involved* in the interaction, the input motor assembly, and body parts that are *affected* by the interaction. I will further discuss this case in the next section.

Lower-Left Quadrant: Fixed Input – Relative Output An input fixed in the world is more constraining since it requires to stand in a defined perimeter that limits movements. In this case, touch is more constraining than mid-air. For example, while limited, the detection range of a Kinect device[3] is less constraining than having to stand at the edge of an interactive table.

A simple example of a mid-air fixed input with a relative output is when a user is scanning a barcode while watching the feedback on a mobile device (Fig. 3.5e). With touch interactions, these kind of input/output combinations are common when transferring an object from a fixed surface to a mobile device, like in Fig. 3.5f (Pick and Drop [Rek97]). Both examples assign the dominant arm to perform input and the non-dominant arm that carries the handheld device to maintain a steady relationship between device and eyes.

[3]http://www.xbox.com/en-US/KINECT

Lower-Right Quadrant: Fixed Input – Fixed Output In this situation, the location and visual attention of the user are constrained by both the input and the output. This is the most constraining combination, especially with *Touch* input.

Mid-air world-fixed input and output is a common combination for pointing on todays wall-sized display, using the "laser pointer" metaphor. Even if the interaction is performed at a distance, it is fixed in the world since it requires to stand at an appropriate location in order to be able to directly point toward an object on the surface (Fig. 3.5g). Conventional touch interaction, with a tabletop or a large display, requires to be in front of the surface. This is even more restrictive with Multitoe ([Aug+10]) since it enables visual output and touch interaction on the floor with the feet (Fig. 3.5h).

3.3.4 Compound Techniques in Multi-Surface Environments

The distributed nature of multi-surface environments often forces users to combine several atomic interaction techniques to perform a complex higher-level task (i) in sequence, e.g., selecting an object first on one touch surface and then on another; (ii) or in parallel, e.g., touching an object on a fixed surface while simultaneously touching another one on a handheld device.

Sequential Combination A sequential combination refers to the temporal sequence of atomic interaction techniques. The combined techniques can be interdependent (sharing the same object, or the output of one being the input of the other), but the first action should be ended before the second one starts. It could consist in selecting an object on a tactile surface (touch and release) and then applying a function onto this object with a menu on a mobile device. In BodyScape, this kind of compound technique does not change the body restrictions that are imposed by each of the atomic techniques in the sequence, nor the body parts they are involving. However, when designing such sequences, one have to consider their characteristics to avoid awkward situations for the user, e.g., moving back and forth in the environment, constantly switching attention between fixed and relative displays, switching a device from one hand to another.

Parallel Combination Parallel combination consists in performing two techniques at the same time. It could consist in touching two tactile surfaces simultaneously in order to transfer an object from one to the other [WB10]. This has to be considered differently than sequential combinations in BodyScape, in order to determine the impact on user's body restrictions, and potential conflicts between movements produced by multiple input motor assemblies. The body movement restriction of a parallel combination of techniques depends on the more restrictive of the combined techniques: combining a motor assembly that coordinates its movements fixed in the world with one that moves relative to the body will result in a composition of two motor assemblies that are coordinated fixed in the world. In the following, I present the three possible types of parallel composition of input motor assemblies into a compound technique.

Composing Input Motor Assemblies to a Compound Technique

Input motor assemblies (IMA) can be composed in three ways: *in series*, *separated* or *overlapping*. Figure 3.6 illustrates three motor assembly compositions using a simple *point-and-select* task on a large display.

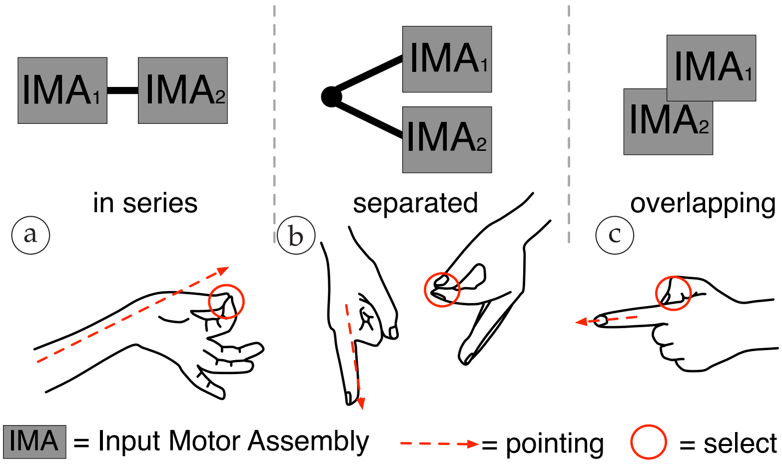


Figure 3.6: Input motor assemblies(IMA) can be composed *in series*, *separated* from each other, or *overlapping*.

Two IMAs in series In Fig. 3.6a, the user combines two input motor assemblies ***in series***. One performs mid-air pointing gesture and the other performs a *select* operation. The corresponding input motor assemblies are the dominant forearm and upper arm for the pointing task (IMA_1) and the fingers for the selection task (IMA_2): IMA_1 moves the screen cursor (note that the forearm's orientation is tracked) to the virtual object and the fingertips of both thumb and index finger press against each other to select the object on the display (see Fig. 3.6a). These input motor assemblies are independent, with no shared motors (body limbs).

Two separated IMAs An example for assigning the two control tasks to two input motors is to assign them to two separated hands (two ***separated*** IMAs): the pointing motor assembly includes the dominant index finger and, due to the kinematic chain, the entire dominant arm (IMA_1) to give the full range of movements. The selection motor assembly includes the index finger and the thumb (IMA_2) of the non-dominant hand (see Fig. 3.6b). Fingers need to be pressed one against the other to trigger the *selection* command.

Two overlapping IMAs Two motor assemblies are ***overlapping*** when they *share* body limbs. For instance, the dominant arm and index finger (IM_1) are responsible for pointing while the selection is performed with the dominant hand's thumb pressing on the index finger (IM_2) (see Fig. 3.6c). The index finger is shared by both input motors.

Input Motor Assemblies and Affected Body Parts

An interaction technique involves an input motor assembly into the interaction. However, in some cases, this input motor assembly can *affect* other body parts that are not directly involved in the interaction. For instance, when users touch their own bodies (*on-body touch* interaction technique) [Lin+11; HTM10] (see Fig. 3.9b), one arm is involved in a pointing task to an *on-body target*. Touching targets on the other arm *affect* this body part. When users interact with an interactive floor [Aug+10] or take a call on mobile phones by tapping with the foot [Ale+12], they shift their balance to the other leg, thus *affecting* it. Another example is when the user both supports and interacts with a hand-held device. The interacting thumb can be simultaneously affected by the support and involved in the interaction. This can lead to interferences effects when combining multiple input motor assemblies to a compound technique performed in parallel.

Note that I sometimes refer to *involved* and *affected* body parts in the following. Involved body parts are in this context all motors (body limbs) that are contained in the input motor assembly.

Interaction Effects of Compound Interaction Techniques in Parallel

When interaction techniques are combinations from two or more parallel atomic interaction techniques, they assign multiple input motor assemblies that interact simultaneously. Input movements can interfere with each other depending upon how the input motor assemblies are composed. In addition, atomic interaction techniques can also affect multiple body parts.

Figure 3.7 illustrates several state of the art atomic interaction techniques, according to their level of restriction of the body configuration, and the sum of body parts that compose the input motor assembly and that are affected by the interaction. To simplify the user's body model, I divide the human body into five groups of body parts: both arms and legs and the torso. Each atomic interaction technique assigns exactly one input motor assembly that can include one to five body parts. In addition, the input motor assembly can affect up to five body parts. An example for an atomic interaction technique that involves and affects all body parts is interactive dance performances [Lat+10], since dancers perform complex movements that have an impact on the body's balance.

Most of the atomic interaction techniques illustrated in Figure 3.7 were already introduced in the previous section. This graphical representation also illustrates the implications of combining multiple atomic interaction techniques to a compound technique with the example of *Touch projector* [Bor+10], in blue in Fig. 3.7. Touch projector consists of using a handheld device as a lens to select objects on a distant display: the user orients the device towards the target (mid-air fixed input and output) and simultaneously touches the tactile screen of the device in order to select the object (touch relative input and output). Touch projector is thus considered a "mid-air fixed input and output" technique in the design space. The advantage of minimizing body restrictions with a technique relative to the body is thus waived by the requirement of a fixed input. It is however beneficial in the case of touch projector, since it enables direct interaction with a display at a distance, preventing moving to the display or using another interaction device.

With this approach, I expect to ease the study and to quantify possible interaction effects between movements of either two input motor assemblies, or when one input motor assembly affects body parts that are themselves parts of another input motor assembly (interferences between input and output motor assemblies are out of the scope of this thesis and will be addressed in future work).

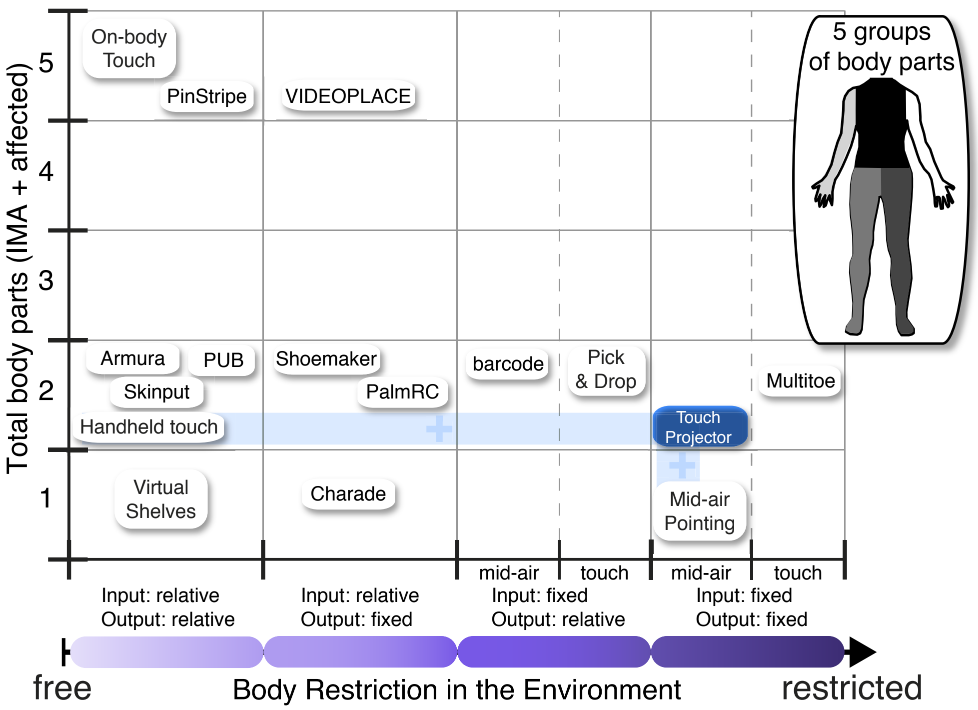


Figure 3.7: Atomic interaction techniques according to level of body restriction and total number of body parts involved to the input motor and affected by the interaction. Touch projector is a compound interaction technique combining hand-held touch and mid-air pointing.

Interaction Effects Between Input Motor Assembly and Affected Body Parts: Figure 3.8 shows a user both supporting and interacting on a hand-held tablet with the non-dominant hand, using two different device holds: either the thumb or the fingers of the support hand reach the interactive surface. This interaction technique allows users to perform bimanual interaction on off-the-shelf hand-held tablets.

The first research question that I investigate in chapter 4—"The Effect of Support on Input Motor Assemblies" is:

> **RESEARCH QUESTION 1:**
> How does device support affect input motor assemblies performance and perceived comfort of use if both support and interaction are shared across one arm?

Another example of an interaction technique where interaction effects might occur between input motor assembly and affected body parts is illustrated

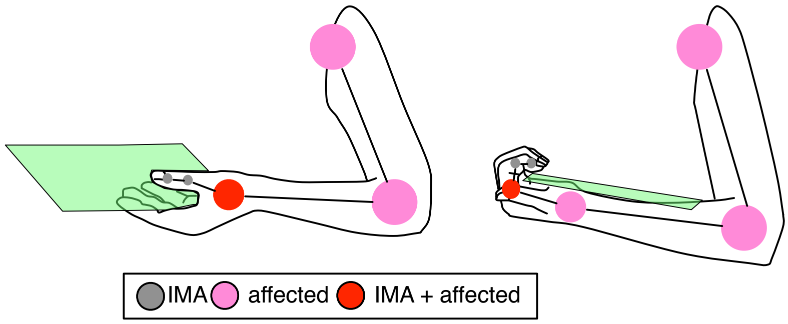


Figure 3.8: Two example holds of a hand-held tablet with the non-dominant hand where the arm shares both device support and interaction. Interaction effects can occur between body parts that support and those that interact.

in Figure 3.9. A user performs a mid-air pointing task (see Fig. 3.9a). This control task involves the dominant arm as motor. Figure 3.9b shows a user performing an on-body touch interaction involving the non-dominant arm and affecting the dominant forearm. Figure 3.9c shows a user performing those two interaction techniques simultaneously: the dominant arm is then both involved and affected.

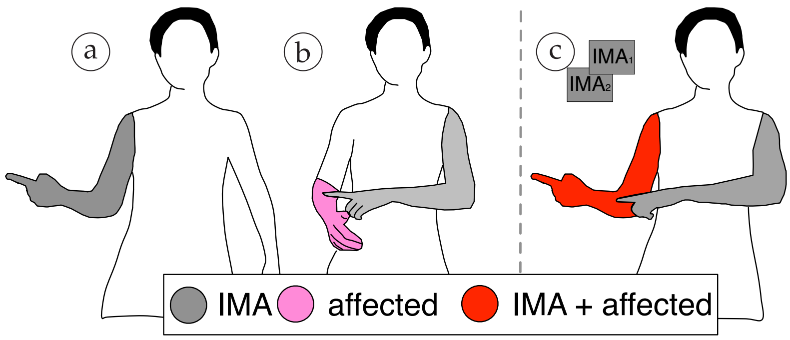


Figure 3.9: Interaction effect between involved and affected body limbs: a) dominant arm involved to mid-air pointing, b) non-dominant arm affecting dominant arm by on-body touch and c) interaction effect between dominant arm being simultaneously involved into mid-air pointing and affected by on-body touch.

The second research question that I investigate in chapter 5—“Interaction Effects between Input Motor Assembly and Affected Body Parts” is:

RESEARCH QUESTION 2:
How do two input motor assemblies affect each other when one input motor assembly, performing on-body touch interaction, affects body parts of a second input motor assembly, performing mid-air pointing at the same time?

Interaction Effect Between two Input Motor Assemblies: Two overlapping input motors can produce input movements that interfere with each other. Figure 3.10a shows a user performing a mid-air pointing task and a circular zoom gesture (see Fig. 3.10c) with one arm. Figure 3.10b shows a user performing both control tasks separated across two hands. The circular movements of the thumb can interfere with the steady pointing gesture of the entire arm.

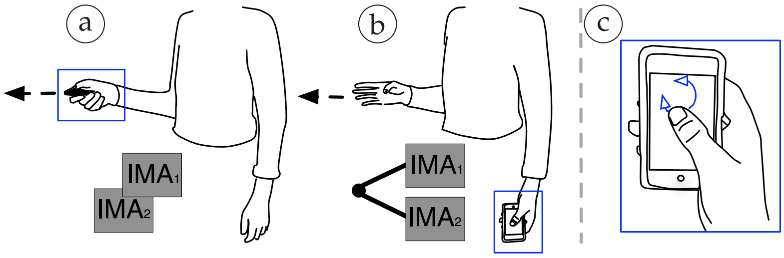


Figure 3.10: Interaction effect between two input motor assemblies: mid-air pointing and circular zoom gesture performed a) with one arm and b) with two arms. c) shows the circular zoom gesture performed by the thumb in both cases.

The third research question that I investigate in chapter 6—"Interaction Effects between Two Input Motor Assemblies" is:

RESEARCH QUESTION 3:
Are interaction techniques that separate the two input motor assemblies more performant than those where both overlap?

3.4 Formal Description of Interaction Techniques with BodyScape

Previous work of Foley et al. [FWC90] on a taxonomy of interaction techniques consider devices that output the same information to be equivalent, despite the potential individual subjective advantages they can present to the user. For example, positioning a cursor in x and y direction using an integral mouse device is considered similar to using an Etch-a-Sketch, where users control both dimensions using individual knobs. BodyScape takes the involved body parts of a technique into account and, therefore, considers those two techniques as being different. BodyScape attempts to assist the designer of interaction techniques in effectively seeing and systematically studying interaction effects between two or more concurrent input movements.

In order to inform the analysis process of interaction effects between input motor assemblies and affected body parts presented in the following chapters, I introduce a formal description of compound interaction techniques. While the longterm goal is a formal description and model which can predict the performance of interaction sequences in multi-surface environments, the first step in the context of this thesis is a graphical notation that helps identifying interaction effects between two concurrent input movements produced by two or more input motor assemblies.

A compound interaction technique T can be decomposed into n atomic interaction techniques ($T = \{AT_1...AT_n\}$), of which each *involves* exactly one input motor assembly[4] and can *affect* one or more input motor assemblies (IMA) (see Fig. 3.11). One motor assembly can be involved in two atomic interaction techniques: e.g. users can perform a symmetric bimanual interaction technique on an interactive table involving both arms to change both the size and the position of a virtual line.

Motor assemblies can be either coordinated world-fixed or body-relative: this is indicated in the figure by the type of outline of the motor assembly: double-lined outline refers to world-fixed and single-lined outline refers to body-relative coordinated motor assemblies.

Motor assemblies always have a spatial relationship with each other, indicated by one of the motor assembly composition types (*in serial*, *separated*, and *overlapping*). Figure 3.11 shows the relationship of MA_3 to all other motor assemblies. To simplify the representation, I dispense with all other relationships for each motor assembly.

[4] An atomic interaction technique involves exactly one motor assembly of each type, input and output. The output motor assembly is omitted here since I am only investigating interaction effects between input motor assemblies in the scope of this thesis.

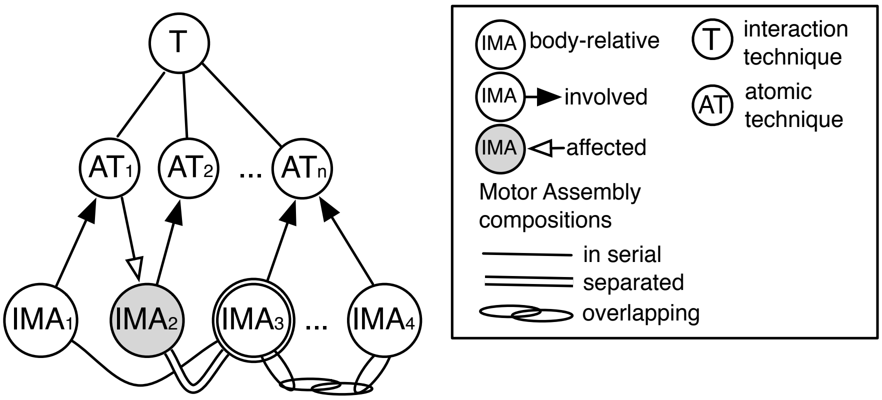


Figure 3.11: The decomposition of an interaction technique into a set of atomic techniques, each involving exactly one input motor assembly (IMA) and affecting none or m IMAs. Each IMA maintains a relationship to the other IMAs, expressed by the IMA composition types (only illustrated for IMA_3).

3.5 Conclusion

This chapter introduced BodyScape, a body-centric framework for analyzing interaction techniques in multi-surface environments. BodyScape makes it possible to formally describe an interaction technique in terms of atomic interaction techniques, each of them handling a control task in the interactive environment. Each atomic interaction technique involves exactly one input motor assembly and can effect multiple input motor assemblies, or other body parts that are not actively involved into the interaction. In addition, BodyScape allows to classify atomic interaction techniques in terms of the restrictions they impose on the user's body configuration.

I presented three ways of composing input motor assemblies and presented two types of interaction effects that might be likely to occur: (i) between two input motor assemblies; or (ii) between input motor assemblies and affected body parts. The following chapters are dedicated to presenting several point designs in the context of BodyScape and investigate these interaction effects. Chapter 4—"The Effect of Support on Input Motor Assemblies" and chapter 5 will explore interaction effects between input motor and affected body parts. Chapter 6 will investigate possible interaction effects between two input motor assemblies.

3.6 Contributions

1. A body-centric framework, BodyScape, that describes interaction techniques independently of the technical aspects of their implementation. BodyScape builds upon two essential factors: the input motor assembly involved in performing a control task in the interactive environment and the effects of the interactive environment on the input motor assembly, i.e., the spatial relationship between the body and the device(s).

2. A formal notation of a ***body-centric interaction technique*** in terms of atomic interaction techniques, each involving exactly one input motor assembly to perform a control task.

3. A body-centric taxonomy classifying related literature regarding shared characteristics of body-relative and world-fixed input and output.

4

The Effect of Support on Input Motor Assemblies

This chapter investigates the role of support in restricting input motor assemblies, in particular with hand-held tablet devices. I present BiPad, a toolkit for designing interfaces for bimanual interfaces on multi-touch tablets in which one of the hands both supports and interacts with the device. I report results of a preliminary study about how people hold tablets and identify five holds that enable the support hand to curl around the device in order to reach the interactive surface with either the thumb or fingers. I present a theoretical design space, that I call BiTouch, which extends Guiard's kinematic chain theory. BiTouch provides a "microscopic" view of a specific input motor assembly in order to identify how both input and support are distributed across motors for each tablet hold. Results show that supporting the device restricts motors within the kinematic chain, affecting the motion of input motors.

4.1 Introduction

In this chapter, I introduce novel bimanual interaction techniques for hand-held tablets where the supporting arm simultaneously interacts with the device. Using the BodyScape design space, I investigate how the device support affects the input motor assembly. According to BodyScape notation, a BiPad interaction technique T combines two *atomic techniques* AT_1 and AT_2 (see Fig. 4.1). Both AT_1 and AT_2 are conventional hand-held touch interaction techniques, where AT_1 involves the dominant arm as input motor (IMA_1) and AT_2 involves the non-dominant arm as input motor (IMA_2). BodyScape highlights that the non-dominant arm is also affected by AT_2, since it helps supporting the device. This two-way relationship between IMA_2 and AT_2, where IMA_2 is both involved and affected, suggests potential interferences between the support and the interaction which might have an impact on user performance and comfort (red frame in Fig. 4.1). I am thus interested in studying how interaction and support can be shared by the non-dominant arm and how support affects interaction[1].

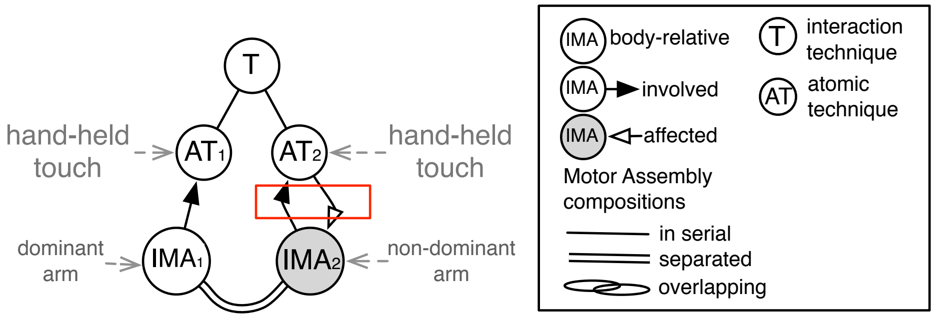


Figure 4.1: BiPad (T) is composed by atomic hand-held touch techniques, each with associated input motor assemblies IMA. The non-dominant arm shares device support and interaction with the IMA_2 input motor assembly which is both involved and affected (red frame).

4.2 Motivation for Bimanual Interaction with the Support Hand

Although commercial tablets offer intuitive interaction techniques such as a *swipe* to displace an object or a *tap* to select an item, they do not fully exploit the range of interaction possibilities found in the research literature regarding bimanual interaction. Moscovich and Hughes has demonstrated that bimanual interaction can increase users' performance using multi-touch enabled

[1]This work is a collaborative work with both my advisors, Stéphane Huot and Wendy Mackay, and was published as a full paper at ACM CHI'12 [WHM12]. This chapter illustrates how BodyScape can be applied to the study of the particular case of bimanual interaction with the support hand. The published paper is included in appendix C—"Selected Publications" (page 149).

surfaces [MH08] and beyond [Bra+08; KAD09]; and Benko et al. showed that users' input precision can be increased [BWB06]; Further, Leganchuk et al. and Wu and Balakrishnan demonstrated that bimanual interaction can enhance the user experience [LZB98; WB03]. However, these studies assume that both hands are free to interact, e.g. on a stationary multi-touch surface or a small multi-touch device placed on a table. I am interested in hand-held tablets which require at least one hand to support the device, thus restricting the ability to interact.

4.2.1 Related Work

Desktop-based bimanual interaction techniques increase both performance and accuracy [BH99; Bie+93; KBS94] and are more convenient when performing highly demanding cognitive tasks [LZB98; Hin+98]. Commercially available PDAs and smart phones are designed primarily for one-handed interaction [PRM00; KB06] due to their small size. Most interaction is accomplished with the index finger, although some techniques use the thumb, since it can reach the screen from most carrying positions [HL07; KB08; RHL08; KBS05]. Additionally, Froehlich et al. proposed to use the outer frame of the phone to improve pointing accuracy or to disambiguate among actions [FWK07]; Similarily Roth and Turner's *BezelSwipe* enriches the interaction vocabulary [RT09].

Several research prototypes offer the potential for bimanual interaction by adding hardware. For example, *HandSense* [WB09] uses capacitive sensors to distinguish among six different grasping styles. One could create simple bimanual tasks by allowing these grasps to modify the actions of the dominant interaction hand. An alternative is *HybridTouch* [SH06], which adds a rear touchpad to a PDA to enable simultaneous front and back interaction: while simultaneously supporting the device, users can simultaneously scroll on the back of the device using their non-dominant hand's index finger while interacting on the front of the device using a stylus.

Wobbrock et al. investigated how different hand positions on the front or back of a handheld device affect interaction performance with the index finger or the thumb [WMA08]. They found that the index finger performed best in all conditions, front or back, and that horizontal movements were faster and more accurate. Although useful for comparing thumb and finger performance on small devices, additional research is needed to understand bimanual interaction on larger portable devices, such as multi-touch tablets.

To date, most bimanual interaction techniques require additional hardware, e.g. to detect touches on the back or sides of the device. For example, *RearType* [Sco+10] includes a physical keyboard on the back of a tablet PC. Users hold it with both hands while entering text, thus avoiding an on-screen keyboard and graphical occlusion by the fingers. *Lucid Touch* [Wig+07] is a proof-of-concept of a see-through tablet that supports simultaneous touch input on the front and on the back of the device. Users hold the device with both hands, with thumbs on the front and remaining fingers on the back. The see-through concept refers to the implemented effect that the hands on the back of the device are displayed semi-transparently on the device screen but other objects behind the screen are not. A mounted camera on the back of the device can detect the shape of the support hands and the device display the virtual representation of the real hand-position on the back of the device. Since users can reach the entire screen, they can perform multi-touch interaction with both support hands without graphical occlusion. However, the arm-mounted camera makes this approach impractical.

Lucid Touch

I am interested in investigating bimanual interaction with the support hand on off-the-shelf tablets without the use of additional hardware, in particular how device support and interaction are shared across the user's arms. Bimanual interaction had been studied extensively on stationary devices or unsupported hand-held devices and, thus, does not inform about the various ways people hold tablets.

4.3 Preliminary Study: How do People Hold Tablets?

Studying how people naturally hold tablets is tricky. Field studies require mounting additional electronic sensors to determine the user's hold. Doing so, however, means changing the device shape- and weight-factor, which can affect the study's outcome. Instead, I asked users to perform a distractor task while observing how they held the tablet.

Participants Six men and two women, average age 30. Four owned iPads, four had never used a tablet.

Apparatus Apple iPad 1 (display: 9.7”, weight: 680 g, dimensions: 19$\times$ 24.3 $\times$ 1.3 cm).

Procedure I told participants that I was interested in how pointing and scrolling performance varies as people sit, stand and walk, given different tablet orientations. This was intentionally misleading, since we were really studying how they unconsciously held the tablet while interacting with it. The true experiment was a [2×3] within-subjects design with two factors: tablet *orientation* (landscape, portrait) and *stance* (sit, stand, walk), with tablet *hold* as the dependent measure. The distractor tasks were *pointing* (tapping five successive on-screen targets) and *scrolling* (moving a slider's thumb-wheel from one end to the other). Pointing targets were distributed across six equal squares on the screen; slider positions included the four screen borders and horizontally and vertically in the screen center.

Participants were asked to hold the iPad comfortably and perform each task as quickly as possible. They were allowed to adopt a new hold only when beginning a new block of distractor tasks. Sessions lasted approximately 45 minutes. At the end, I debriefed each participant as to the true goal of the study to learn how they chose to hold the tablets. I first asked them to reproduce the holds they had used and then to adapt them so that the fingers or thumb of the support hand could reach the touch screen. I asked them to rate comfort and ease of interaction when using the support hand to interact and whether they had suggestions for other holding positions.

Data collection I videotaped each trial and coded how participants supported the tablet with the non-dominant hand, wrist or forearm. I collected touch events, including those that occurred outside experiment trials and while reading instructions. I also measured completion time per trial.

4.3.1 Results

I did not find a single, optimal hold and found significant differences according to experience. All four novices used the same uncomfortable position: the fingers, thumb and palm of their non-dominant hand supported the center of the tablet, like a waiter holding a tray. Novices found this tiring but worried that the tablet would slip if they held it by the border. None found other holds. In contrast, the four experts easily found a variety of secure, comfortable holds. I identified ten unique holds, five per orientation, all of which involved grasping the border of the tablet with the thumb and fingers. Fig. 4.2 shows these five holds in portrait mode, with the thumb on the bottom, corner or side, or the fingers on the top or side.

Table 4.1 shows how these holds were distributed across the six conditions: most common was F-side (41%), least common was T-side (9%). The lat-

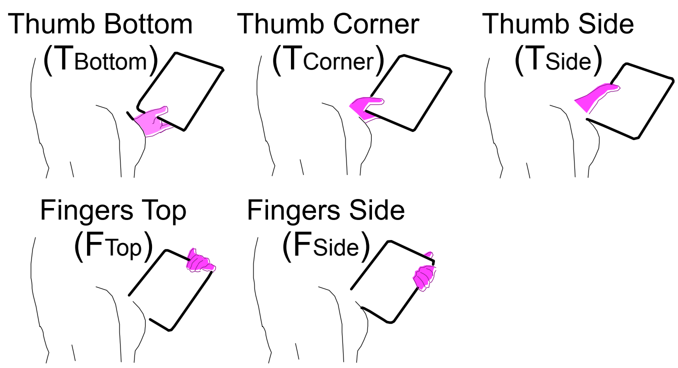


Figure 4.2: Five spontaneous holds (portrait orientation).

Table 4.1: Total holds per condition (expert users)

	F_{side}	T_{bottom}	F_{top}	T_{corner}	T_{side}
Landscape	3	4	4	4	1
	8	4	0	4	0
	4	4	7	0	1
Portrait	8	3	1	0	4
	8	4	0	4	0
	8	1	3	1	3
	41%	21%	16%	14%	9%

ter was deemed least comfortable, especially in landscape mode, but participants felt that they could use it for a short time. Experts tried nine of ten possible holds in the sitting and walking conditions, but only six when standing, omitting F-top or T-side in both orientations. Individuals varied as to how many unique holds they tried, from three to eight of ten possible. All switched holds at least once and two switched positions often (50% and 66%) across different blocks of the same condition.

4.3.2 Design Implications

First, tablets can feel heavy and users are more comfortable when they can change orientation or swap the thumb and fingers. We should thus seek a small set of roughly equivalent bimanual interactive holds that are easy to shift between, rather than designing a single, 'optimal' hold. Second, users can use the thumb and fingers of the support hand for interaction. We can thus create interactive zones on the edges of the tablet, corresponding to the holds in Fig. 4.2. Fig. 4.3 shows these zones in portrait and landscape mode. Although changes in the form factor of a tablet, such as its size, shape or weight, may affect these holds, users are still likely to shift between holds for comfort reasons, just as when reading a book or holding a notebook.

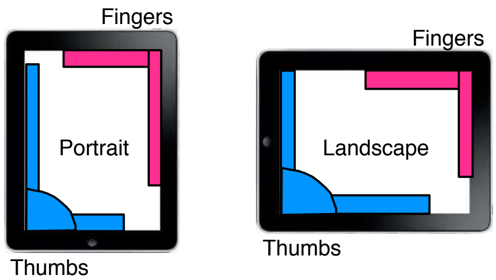


Figure 4.3: Five support-hand interaction zones.

4.4 The BiPad Prototype

4.4.1 BiPad Toolkit and Applications

In order to design and study bimanual interaction with handheld tablets, I implemented the BiPad toolkit[2] that enables developers to add bimanual interaction to off-the-shelf multi-touch tablets. BiPad creates and manages five interactive zones in each device orientation (illustrated in Figure 4.4a in portrait orientation), where the fingers or the thumb of the supporting hand can interact. These correspond to the holds identified in the preliminary study. The BiPad toolkit, written in Objective-C for the Apple iOS operating system, supports the development of bimanual applications as follows:

BiPad applications consist of one or more views, widgets and controllers, similar to standard iOS applications. The framework lays out the interface in the main view to control overlay feedback and advanced input management required to enable BiPad interaction. The application defines BiPad-enabled functions that can be mapped to interactions with the support hand. For example, a text editing application could define `shift` and `num` functions equivalent to pressing the shift or number keys of a virtual keyboard.

BiPad zones appear on the sides and corners of the screen (see Fig. 4.4b). Applications can define various interactions for the support hand and modify

[2]The BiPad toolkit can be downloaded at the http://insitu.lri.fr/Bipad/Bipad

the default visual representation, e.g., buttons for taps and guides for chords. Zones are displayed as 80-pixel strips, of which the 40 outermost are semi-transparent, on top of the edges of the application view. Zones may be permanently or temporarily visible and the user's hand position determines which is active. Temporarily visible areas shrink automatically when not in use, displaying only a narrow semi-transparent strip of pixels on the appropriate side. Touching once on the outer part of a shrunken BiPad zone causes it to slide out and enables interaction. If a zone contains interaction widgets and is configured to be temporarily visible, it does not shrink completely but remains semi-transparent (see Fig. 4.4c).

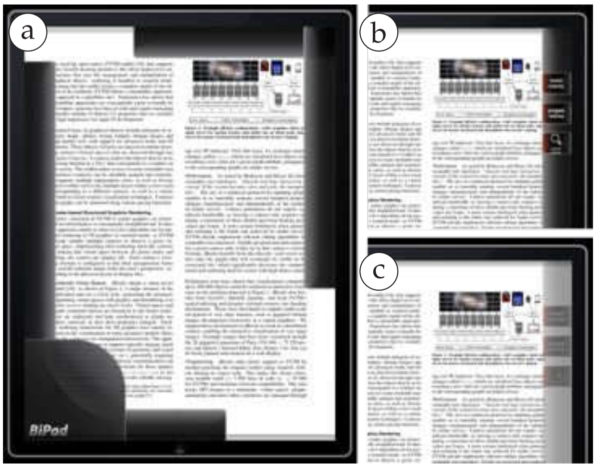


Figure 4.4: Screenshot of an interface created with BiPad: a) five activated interactive zones in portrait device orientation, b) bipad zone for F_{side} remains c) semi-transparent if not in use, i.e. when user interacts with the pdf content.

4.4.2 BiPad Interaction Techniques for the Non-Dominant Support Hand

Users have access to a larger input vocabulary when using different motions and gestures. BiPad introduces three predefined interaction techniques for the support hand: bimanual *Taps*, *Chords* and *Gestures*. Bimanual *Taps* involve a press-and-release action on a button within a BiPad zone, using a finger or the thumb (see Fig. 4.5a). Bimanual *Chords* involve multiple fingers pressing down simultaneously within a BiPad zone (and are obviously not possible with the thumb). Figure 4.5b shows how pressing the *'stroke'* button with the index finger adds additional finger positions below. The user can adjust the stroke size by holding down a second finger on the appropriate button.

Bimanual *Gestures* involve sliding the thumb or finger, starting from a BiPad

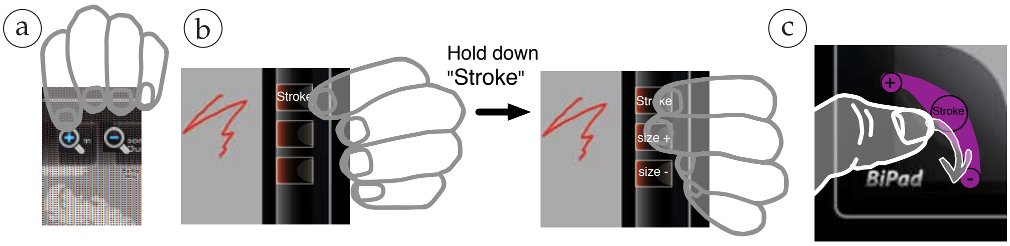


Figure 4.5: BiPad interaction techniques: a) Taps on buttons. b) Chords with multiple fingers. c) Gestures in multiple directions.

zone or from an edge related to a BiPad zone, as in Bezel Swipe [RT09]. In the border zones, *Gestures* are limited to orthogonal movements from the edge, but offer additional degrees of freedom for the thumb in the corner (up-to-down, right-to-left and diagonal). Small stroke shapes indicate the direction of the gesture and its function (see Fig. 4.5c).

The application defines which BiPad interaction(s) will trigger which function in which zone(s). Applications can specify several interaction techniques for the same function depending upon which BiPad zone (and therefore *Hold*) the user registers. For example, an application might specify that a *Tap* with a finger on the F_{side} zone and a downward *Gesture* with the thumb in the T_{corner} zone will both shift modes for the dominant hand, triggering a pop-up menu rather than selecting an on-screen object. Finally, to activate a Bipad zone for the first time, the user double taps its visible area. This is normally only necessary when the user changes the hold.

4.4.3 Sample BiPad Applications

With the BiPad toolkit, I implemented three applications that illustrate how to add bimanual interaction to handheld tablets (see Fig. 4.6).

BiPDF (see Fig. 4.6a) is a PDF reader that uses standard touch gestures to navigate through pages, scroll or zoom the document. A pie menu contains additional commands, e.g. first/last page. As with many tablet applications, the user must touch and dwell to activate the menu instead of executing a gesture. We added a bimanual tap that speeds up interaction: while the user is touching the screen with the dominant hand, a tap on a BiPad button activates the menu immediately.

BiText (see Fig. 4.6b) lets users create custom bimanual shortcuts for text entry, e.g. a button for the '`space`' key and a quasi-mode button for the soft keyboard's '`keypad`' key. Although the dominant hand can also reach these keys, it requires extra movement. The user can also assign any key from the keyboard to a BiPad button by simultaneously pressing the two. Modifier keys, such as the '`keypad`' key become quasi-modes: they activate the mode as long as they are being pressed. Two other BiPad buttons accept or reject the suggestions from the standard text completion engine, reducing movements by the dominant hand.

Figure 4.6: BiPad applications: a) trigger and navigate in menus with BiPDF b) quasi-mode modifier key in BiText c) panning and zooming with BiMaps.

The previous examples refer to two-handed interactions based on temporal multiplexing. BiPad can also handle spatially multiplexed tasks. *BiMap* (see Fig. 4.6c) lets users zoom in and out by pressing buttons with the support hand. They can select part of the map larger than the view port by (i) selecting with the dominant hand; (ii) simultaneously controlling the zoom factor with the non-dominant hand; and (iii) continuing to change the selection with the dominant hand.

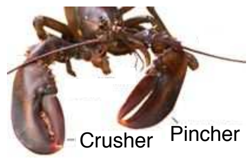
4.5 Classifying Tablet Holds

In the preliminary study, we identified five physical holds of a tablet device that allow fingers and thumb to reach the interactive surface. From a body-centric perspective, the BodyScape design space and classification informs that a BiPad technique involves two separated input motor assemblies which interact in parallel to perform a compound task (see Fig. 4.1). At a finer level of granularity we can identify two families of holds that show major differences in the level of body involvement: holds where the *fingers* reach the front of the device (F_{top} and F_{side}) require a larger part of the body to support the device than holds where the *thumb* reaches the front (T_{bottom}, T_{corner}, and T_{side}). In addition, all holds with fingers on the front offer more fingers to interact than only the thumb. Guiard's *kinematic chain theory* [Gui87] offers a detailed analysis on *limb level* to describe these differences. In Guiard's model, the limbs represent motors that create motion. The motors of an arm, upper arm, forearm, palm and fingers, are assembled to form a kinematic chain and contribute to the overall performance of an action, e.g., reaching for a glass of water. When device support is distributed across multiple motors, these motors can be blocked by the rigidity of the device. This section investigates in detail the role of the support limb with respect to the kinematic chain.

4.5.1 Kinematic Chain Theory: Modeling Bimanual Action

Guiard introduced a theoretical framework for the study of *asymmetric* bimanual interaction [Gui87], human skilled manual activities that involve two hands, each playing a different role in the task. The two main assumptions of his model are that (1) the two hands act as two abstract motors that create motion and that (2) this two motors cooperate with each other as if they were two assembled motors in the kinematic chain: the motions of the non-dominant hand frame the motions of the dominant hand.

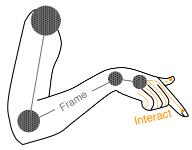
Guiard identifies three types of everyday manual activities: *uni-manual* (e.g. brushing one's teeth), *asymmetric bimanual* (e.g. playing the violin), and *symmetric bimanual*, where the two hands play essentially the same role either *in phase* (e.g, lifting a weight) or *out of phase* (e.g. in juggling). He is particularly interested in the case of asymmetric bimanual activities and how we assign roles to each hand. Guiard illustrates the principle with lobster claws: one claw, either left or right (it depends on the species), is usually larger than the other claw. For a "left-handed" lobster, the right claw is usually larger and stronger than the left claw, and the left claw is usually sharper than the right one. An action of a lobster starts with the larger and stronger claw; it crushes and holds prey; the smaller claw serves to pinch the prey into smaller pieces. This observation is consistent with Guiard's *higher order principles governing the asymmetry of human asymmetric bimanual gestures*: (1) *Right-to-Left spatial reference in manual motion*, (2) *Left-hand precedence in action*, and (3) *Left-Right contrast in the spatial-temporal scale of motion*.



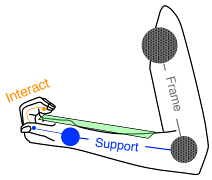
left-handed lobster

The significant aspect of Guiard's theory for my work is the abstraction of limbs, such as the hands, into *abstract motors* that both contribute to the overall performance of a bimanual gesture. Guiard proposes the kinematic chain as a general model, in which the shoulder, elbow, wrist and fingers work together as a series of abstract motors. Each consists of a proximal element, e.g. the elbow, and a distal element, e.g. the wrist, which together make up a specific link, e.g. the forearm: the forearm acts as one motor within the kinematic chain of the arm that contributes to the overall arm movement. Proximal elements frame the action of distal elements: in this case, the distal wrist must organize its movement relative to the output of the proximal elbow, since the two are physically attached.

Frame, Interaction:



Frame, **Support**, Interaction:



In a bimanual action, such as the interaction on an interactive table, we see the kinematic chain in action: the forearms *frame* the action of the hands; the hands *frame* the action of the fingers; and the fingers interact. The two hand motors are free to move and each limb in the *arm chain* has either the role of framing or interacting. In the case of supporting a hand-held device, however, some motors in the chain of the arm play another role: they are responsible for the support of the device. The support introduces a *restriction* to movements produced by the motors. For instance, when holding a tablet with forearm and palm, the support anchors the movements of forearm and palm motors: the upper arm motor *frames* the movements of the forearm motor; the forearm and wrist are *supporting* the device and are therefore joined into one motor, unable to move the joint in between; and the fingers *interact*.

4.5.2 BiTouch Design Space: Framing, Support, Interaction

The spatial relationship between body and device changes with varying holds of hand-held devices. In the identified holds of our preliminary study, different body parts are involved into the support of the device. This introduces a variable restriction in movement: as the number of body parts needed to support the device increases, the more joints in between those body parts are blocked in movement by the rigidity of the device.

BiTouch is a design space that identifies *framing*, *support* and *interaction* as major dimensions and investigates the degrees of freedom left for the interacting limbs after framing and support had been allocated. BiTouch is a *microscopic view* into input motor assemblies which allows to analyze the distribution of support on the considered motors, and to investigate the effect of support on the motors that perform the input movement.

Figure 4.7 shows three bimanual alternatives, based on the location of tablet support within the kinematic chain: the palm or forearm of the non-dominant arm (see Fig. 4.7a, 4.7b); shared equally between the palms of both hands (see Fig. 4.7c). In each case, the most proximal links control the spatial frame of reference; support links are always intermediate between framing and interaction links; and the most distal links use whatever remains of the thumb and fingers to interact.

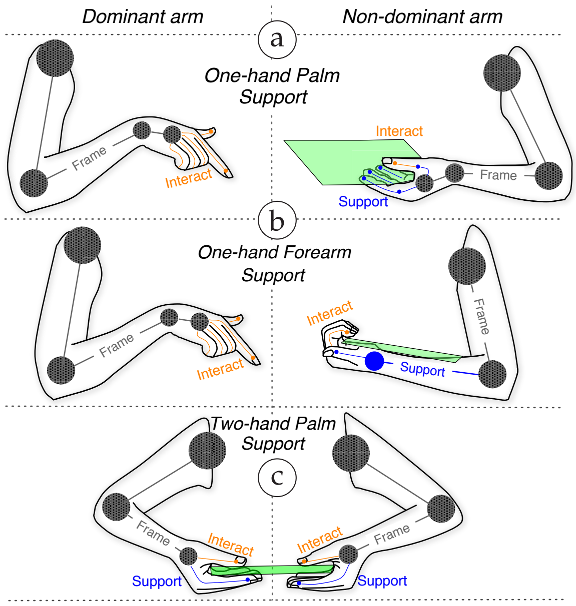


Figure 4.7: The user creates a spatial *frame*, *supports* the device, and *interacts* with it. Different holds offer different trade-offs with respect to interaction possibilities and comfort.

In the preliminary study, I identified ten user-generated support holds that permit the thumb or fingers to reach the interactive area. Each poses trade-offs between comfort and degrees of freedom available for interaction. For example, supporting the tablet with the forearm (see Fig. 4.7b) provides a secure, stable hold but forces the fingers to curl around the tablet, leaving little room for movement. In contrast, holding the tablet in the palm (see Fig. 4.7a) gives the thumb its full range of movement, but is tiring and less stable.

Table 4.2 summarizes the key dimensions of the BiTouch design space, according to framing, support and interaction functions of the kinematic chain. Each is affected by the relationship between specific characteristics of the human body, the physical device and the interaction between them.

Framing is handled at the most proximal *locations* within the kinematic chain and may be *distributed* over multiple parts of the body.

Support always occurs in *locations* within the kinematic chain, distal to the frame. Support may be completely *distributed* over one or more body parts, symmetrically or not; shared with an *independent* support, e.g. a table or lap; or omitted, e.g. interacting on a freestanding interactive table.

Table 4.2: Trading off *framing*, *support* and *interaction* functions of the kinematic chain with respect to the body and the device.

Framing	
Location:	proximal link in the kinematic chain
Distribution:	1 – n body parts
Support	
Location:	none or middle link in the kinematic chain
Distribution:	0 – n body parts
Independence:	0% – 100% body support
Interaction	
Location:	distal link in the kinematic chain
Distribution:	1 – n body parts
Degrees of freedom:	0% – 100% body movement
Technique:	touch, deformation,...

Interaction is always handled at the most distal *location* in the kinematic chain, immediately after the support link. Interaction may be *distributed* across one or more body parts, often incorporating the thumbs or sets of fingers. The *degrees of freedom* available for interaction depend upon what remains after framing and support functions have been allocated, e.g. a finger tip, and the inherent movement capabilities of the body part, e.g. the pinky has little independent movement compared to the index finger. Possible *interaction techniques* are affected by all of the above, as well as the technical capabilities of the device. For example, touch sensors might appear on the front, side or back of the device, or the device itself might be deformable.

The BiTouch design space allows us to describe all of the user-generated holds from the preliminary study, as well as many from the literature, e.g. bimanual interaction on free-standing interactive tabletops. It also suggests directions for designing new bimanual interaction techniques. For example, although the hold in Figure 4.7c did not appear in the preliminary study, it becomes an obvious possibility if we examine ways to share support across hands. Similarly, once we understand which thumbs or fingers are available for interaction and what constrains their potential movement, we can design novel interaction techniques. In fact, hands that interact as well as support the device have fewer degrees of freedom available for movement. I thus expect the non-dominant support hand to be capable of limited interaction, e.g. mode switches or menu choices, that frame the interaction of the freer dominant hand.

Additionally, the BiTouch design space can now be used to classify the five holds corresponding to the five BiPad interactive zones on the edge of the tablet. All holds where the thumb interacts in a BiPad zone have in common that the arm and forearm frame the support, the palm and fingers fully support the devices, and the thumb has full degrees of freedom to perform either bimanual taps or gestures (the intrinsic limitation of having only one thumb in each hand makes it impossible for us to perform *Chords* in this situation). All holds where the fingers interact in a BiPad zone have in common that the arm is framing the support, the forearm, palm, and parts of the fingers fully support the device, and the tips of the fingers interact using one of the three techniques of BiPad (bimanual taps, chords or gestures). Since parts of the fingers are involved into the support by curling around the device, their degrees of freedom are reduced.

4.6 Evaluating Interaction with the Support Hand

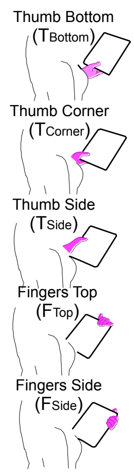
I ran a controlled experiment to determine if BiPad bimanual interaction techniques outperform a common one-handed technique and whether BiTouch successfully identifies which bimanual holds are the most comfortable and efficient.

I asked participants to stand while holding a multi-touch tablet, using one of the holds identified in the preliminary study. I then asked them to perform a series of bimanual *Taps*, *Gestures* or *Chords*, using the thumb or fingers of the non-dominant support hand to modify the actions of the dominant hand. The key research questions were:

Q1 Are two-handed BiPad techniques faster than a one-handed baseline technique to perform a similar task?

Q2 What are the trade-offs among the different bimanual holds, orientations and interaction techniques?

Participants Nine men, three women, all right-handed, aged 22-35. Six own a touch-screen phone, one owns a tablet PC.



Apparatus iPad 1 (display: 9.7" , weight: 680 g, dimensions: 190 $\times$ 243 $\times$ 13 mm), running BiPad.

Procedure I conducted a $[2 \times 5 \times 3]$ within-subjects design with three factors: ORIENTATION (portrait, landscape), HOLD (F_{side}, F_{top}, T_{bottom}, T_{corner}, T_{side}), corresponding to the five BiPad interaction zones, and TECHNIQUE (tap, chord, gesture), i.e. 30 unique conditions, plus the no-BiPad control, a standard one-handed task. I discarded eight conditions as impossible or impractical:

Chords can only be performed with the F_{side} and F_{top} HOLD (in both *Orientations*) since a single thumb cannot perform multi-finger interactions.

Gestures were discarded in the F_{side} and F_{top} landscape conditions, since the short edge of the tablet cannot be held steadily on the forearm.

Trials were organized into blocks of 6 trials according to TECHNIQUE, ORIENTATION, and HOLD. Participants were asked to stand and support the tablet with a specified hold. In each trial, the participant touched four successive 80-pixel circular targets with the index finger of the dominant hand while holding the tablet with the non-dominant hand. Targets were arranged randomly around the center of the screen. The first target of a series was always green and one randomly chosen target of the following three targets was red. When the red target appeared, the participant was instructed to use the specified technique to turn the target from red back to green before touching it with the dominant hand.

The four techniques for changing red targets to green include the three BiPad techniques: *Tap*, *Chord*, *Gesture*, and the no-BiPad control condition. The three chords use the index finger and one or both of the remaining fingers of the support hand (middle or ring finger). Gestures slide toward the center of the screen, except for T_{corner}, where the thumb slides up-down, down-up or diagonally. In the no-BiPad control condition, the user touches a button at the bottom of the screen with the dominant hand. The task was chosen to support both pointing and bimanual interaction, including mode switches and quasi-modes.

Participants began with the uni-manual no-BiPad control condition, followed by the bimanual BiPad conditions (ORIENTATION, HOLD, TECHNIQUE) counterbalanced across subjects using a Latin square. Although this simplifies the experimental design, it does not account for potential order effects between uni-manual and bimanual conditions. On the other hand, all of today's tablets are one-handed and it is unlikely that performing a bimanual task prior to a uni-manual one would improve performance on the latter. Indeed, the more likely effect would be a drop in performance due to fatigue. To ensure that participants were familiar with the basic task and both conditions, I asked them to perform a three-trial practice block in portrait mode prior to each no-BiPad condition and to each TECHNIQUE×HOLD condition. They were also allowed to perform a one-trial recall prior to each TECHNIQUE×ORIENTATIONS×HOLD so they the could find a comfortable position for the assigned hold.

To begin an experimental BiPad block, participants touched the specified BiPad zone to register the support hand. Participants were asked to maintain this hold throughout the block and perform each task as quickly as possible. At the end of each condition, they evaluated how comfortable it was to interact with the support hand using that hold. Each session lasted approximately 45 minutes.

In summary, I presented two orientations for no-BiPad, all 10 holds for bimanual taps, eight for bimanual gestures (no landscape thumb holds) and four for bimanual chords (fingers only). I thus collected 216 trials per participant:

- 6 replications of the *no-BiPad* control condition in both ORIENTATIONS (landscape, portrait): 12 trials;
- 6 replications of the *Tap* technique in all HOLD and ORIENTATION conditions: 60 trials;
- 6 replications of the three *Chord* techniques in both ORIENTATIONS for finger-based HOLDS (F_{side}, F_{top}): 72 trials;
- 6 replications of each of the three *Gesture* techniques:
  - two-finger-based HOLDS (F_{side}, F_{top}) in portrait ORIENTATION: 12 trials;
  - two thumb-based HOLDS (T_{bottom}, T_{side}) in both ORIENTATIONS: 24 trials;
  - one thumb-based HOLD (T_{corner}) in both ORIENTATIONS: 36 trials.

Data Collection I videotaped each trial and recorded three temporal measures:

1. Trial time: from the appearance of the first target to the selection of the final target;

2. BiPad reaction time: from the appearance of the red target to the first touch in the BiPad area;

3. BiPad completion time: from the appearance of the red target to the successful execution of the BiPad interaction.

Participants evaluated the perceived comfort for each combination hold/bimanual technique with a 5-point Likert scale (1 = very uncomfortable; 5 = very comfortable).

4.7 Results

I conducted a full factorial ANOVA and handled 'participant' as a random variable, using the standard repeated measures REML technique from the JMP statistical package.

4.7.1 Q1: Bimanual BiPad vs. one-handed interaction

I compared the mean trial time of BiPad techniques to the no-BiPad control condition, using the TECHNIQUE×ORIENTATION ×Random(PARTICIPANT) ANOVA model. I found a significant effect for TECHNIQUE ($F_{3,33} = 16.16$, $p < 0.0001$) but no effect for ORIENTATION ($F_{1,11} = 0.30$, $p = 0.60$). However, I did find a significant interaction effect between TECHNIQUE and ORIENTATION ($F_{3,33} = 8.23$, $p = 0.0003$).

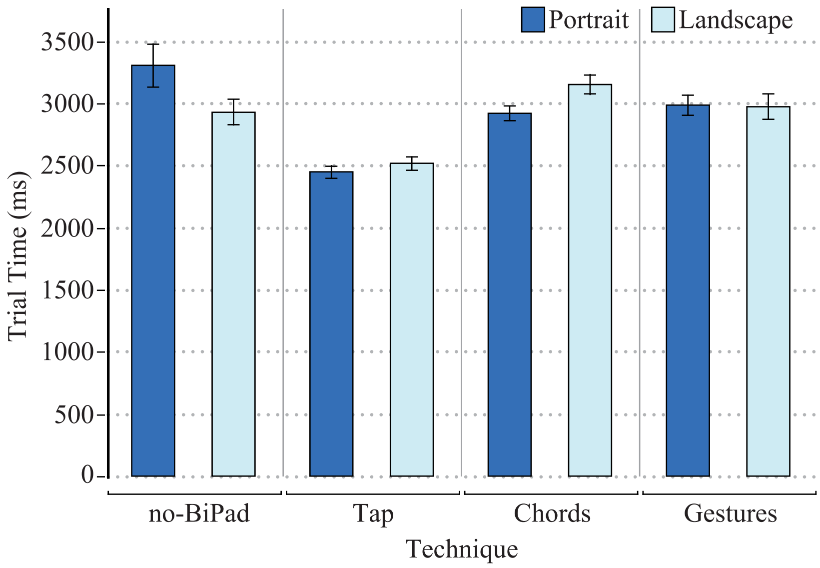


Figure 4.8: Mean *Trial Time* for each TECHNIQUE by ORIENTATION.

This can be explained by the faster performance in landscape mode for the one-handed no-BiPad condition (see Fig. 4.8): participants performed 11.4% faster ($F_{1,11} = 4.6$, $p = 0.04$) because the distance to reach the button is shorter. Thus, while bimanual taps are significantly faster than the control condition for both orientations (25.9% in portrait and 14% in landscape), bimanual gestures and chords are only significantly faster than no-BiPad in portrait mode (10.4% and 11.7% resp.). In landscape mode, the differences between no-BiPad and bimanual gestures and chords are not significant.

Bimanual taps are significantly faster than bimanual gestures and chords in both device orientations (17.3% and 16.1% in portrait, 14.7% and 19.7% in landscape). Participants significantly preferred bimanual taps (3.5) over bimanual chords (3.3) and gestures (2.7) ($F_{2,22} = 17.5$, $p < 0.0001$). Overall, BiPad techniques were more efficient than the one-handed technique I compared them with.

4.7.2 Q2: BiPad tradeoffs: HOLD×ORIENTATION by TECHNIQUE

BiPad *Taps* I ran an ANOVA with the model HOLD×ORIENTATION× Random(PARTICIPANT) on trial time for BiPad taps. I found significant effects for HOLD and ORIENTATION ($F_{4,44} = 3.10$, $p = 0.02$ and $F_{1,11} = 5.37$, $p = 0.04$) and no interaction effect ($F_{4,44} = 0.65$, $p = 0.63$).

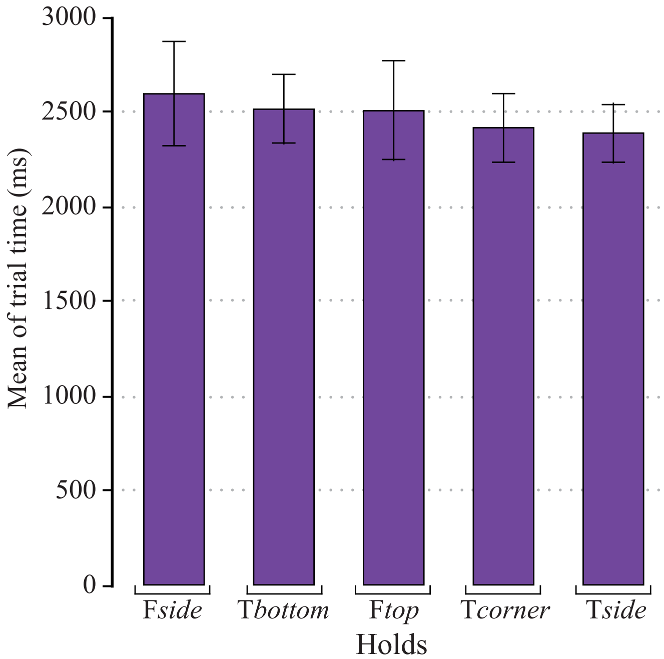


Figure 4.9: *Tap* performance according to HOLD.

For HOLD, Tukey post-hoc tests revealed only one significant result: placing the fingers on the right is slower than placing the thumb on the left side of the tablet for right-handed participants (see Fig. 4.9). For ORIENTATION, a Student's t-test reveals that portrait is significantly faster (LSM $= 2447.31\,ms$) than landscape (LSM $= 2515.99\,ms$).

Performance among bimanual taps is very similar across conditions, making them suitable for all ten holds. The only significant difference is between fingers and thumbs with a side hold. However, although the F_{side} hold is slightly slower, participants preferred it to the T_{side} hold: fingers are more stable than thumbs and cause less fatigue.

BiPad *Gestures* As I discarded the two bimanual holds with fingers placed on the right and top of the device in landscape mode, I examined trial Time for each ORIENTATION condition separately for the remaining eight holds. HOLD has a significant effect on the performance time in both portrait ($F_{4,44} = 4.14$, $p = 0.01$) and landscape ($F_{2,22} = 4.75$, $p = 0.02$).

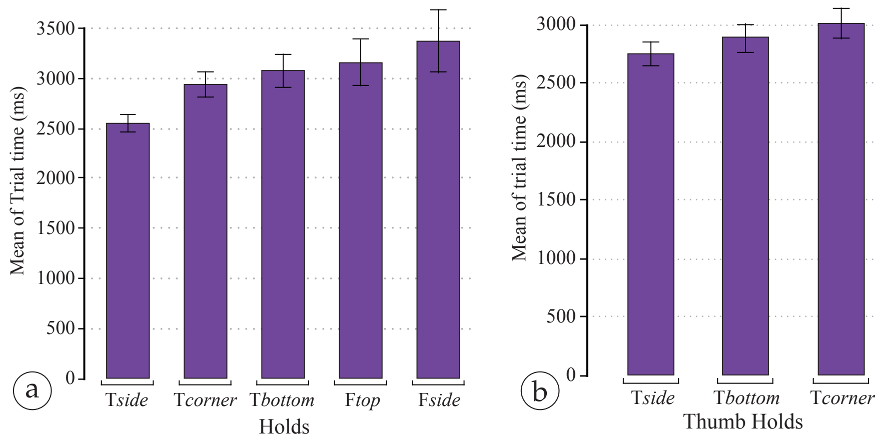


Figure 4.10: *Gesture* performance according to HOLD (a) in *Portrait*, (b) in landscape

In *Portrait*, post-hoc Tukey HSD tests show that, for a right-handed user, performing gestures with the fingers on the right side of the device is significantly slower than with the thumb on the left side (see Fig. 4.10a). Participants preferred performing gestures with the fingers or with the thumb on the side of the device. In fact, gestures are most difficult to perform when the support hand is placed on the top or bottom of the device when held in portrait mode.

In landscape, where only the *Thumb* placements were tested, performing gestures while supporting the tablet with the thumb on the bottom of the device is significantly faster than in the corner (see Fig. 4.10b). However, since gestures were performed in both ORIENTATION conditions with the thumb, I also compared performance according to thumb holds in both orientation conditions (HOLD×ORIENTATION×Random(PARTICIPANT)).

I found no significant effect of HOLD and ORIENTATION but a significant interaction effect for HOLD×ORIENTATION ($F_{2,22} = 15.08, p < 0.0001$). This is because performing gestures with the thumb is significantly faster in portrait, when the support hand is on the side, but significantly slower when the thumb is on the bottom, in which case landscape is faster. The difference between orientations is not significant when the thumb is placed in the corner (see Fig. 4.11).

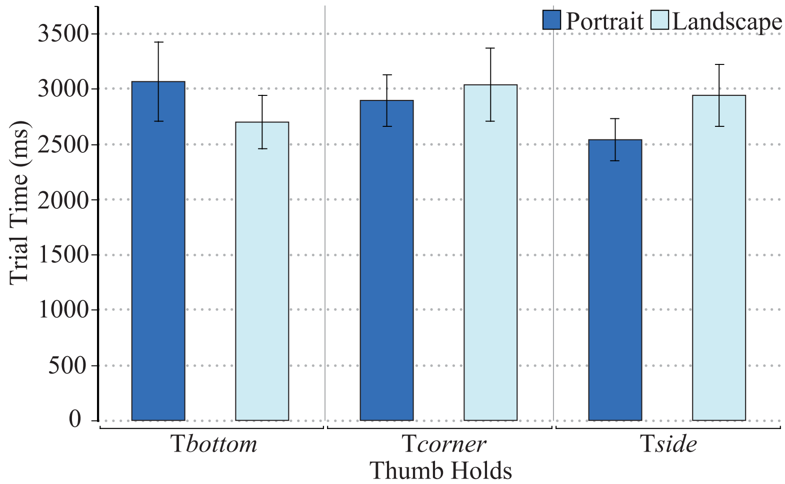


Figure 4.11: *Gesture* performance for the *Thumb* according to HOLD and ORIENTATION.

The latter effect is interesting and can be explained by the principle of a lever. The greater the distance between the balance point and the most distal support link, the heavier the tablet is perceived. This is considered less comfortable and users find it more difficult to perform gestures. The exception is when the thumb is in the corner: the distal point of the support is equally close to the tablet's balance point in both orientations, thus the two holds are not significantly different. This explanation correlates with the participants' comfort ratings and comments. They preferred to perform gestures with the thumb on the side in portrait and on the bottom in landscape but had no preference for orientation when the thumb is in the corner. Compared to other BiPad techniques, however, gestures were perceived as relatively uncomfortable and practical only for rapid or occasional use.

BiPad *Chords* I ran an ANOVA with the model HOLD×ORIENTATION×CHORD TYPE×Random(PARTICIPANT) on *Trial Time*. I found no significant effects of HOLD and ORIENTATION and no interaction effects. For CHORD TYPE, I found a significant effect ($F_{2,22} = 9.09$, $p = 0.01$): holding the index finger down together with the middle finger is significantly faster ($2875\,ms$) than holding down three fingers ($3095\,ms$) or the index and ring finger together ($3131\,ms$).

Participants did not express any significant comfort preferences with respect to chords. However, some participants reported that chords are difficult to perform at the top of the device, especially in landscape mode, due to tension in the arm. Two users could only perform two-finger chords since their third finger could not easily reach the screen.

4.8 Discussion and Conclusion

In summary, these results demonstrate not only that hand-held touch tablets can support bimanual interaction, but that it outperforms all tested unimanual interactions in almost all of our experimental conditions. I created a set of 22 bimanual interaction techniques that combine the ten holds identified in the preliminary study with bimanual taps (10), chords (4) and gestures (8). BiPad techniques offer users trade-offs in performance, comfort and expressive power that let users transition smoothly among them.

In our general body-centric approach, BiPad interaction techniques consist of two atomic touch interaction techniques (AT_1 and AT_2 in Fig. 4.1). Each of this atomic technique involves one arm into the interaction (IMA_1 or IMA_2). However, the dominant arm, IMA_2, is simultaneously involved and affected by the corresponding atomic technique due to the device support and this have an impact on its performance and on user comfort. Whereas participants were more performant in holds where the thumb interacts on the front (T_{side}) than in holds where the fingers interact on the front (F_{side}), they perceived it as less comfortable (see Fig. 4.12).

The support chain varies in *length* between the two holds (see Fig. 4.12, blue bar): T_{side} holds support the device using the palm, F_{side} holds support the device with the forearm and palm. A F_{side} hold has a long support chain and was perceived as more comfortable than T_{side} which has a short support chain: I therefore hypothesize that the perceived comfort is correlated with the *length of the support chain* in the kinematic chain.

The support chain varies also in the *degrees of freedom* left for the interaction chain (see Fig. 4.12 red bar): T_{side}'s palm support leaves the thumb free to move; the F_{side}'s forearm and palm support, however, requires the fingers to curl around the devices which restricts their movements. T_{side} was more performant than F_{side}: I therefore hypothesize that the interaction performance correlates with the degrees of freedom of body movements in the interaction chain.

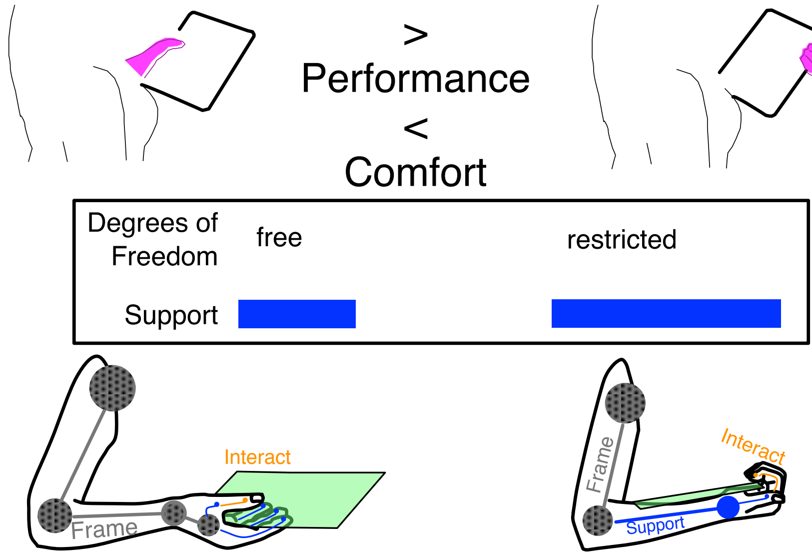
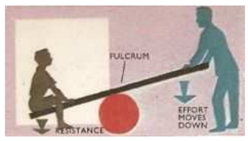


Figure 4.12: Holding the tablet with the thumb on the side (T_{side}, on the left) has an inverse correlation between performance and comfort with holds where the fingers are on the side (F_{side}, on the right). Users perform faster in T_{side} but perceive it less comfortable: the input motor assembly has more degrees of freedom but the support is distributed across fewer body limbs.

The next effect of support on input motor assemblies that we identified in this study relies on the *lever* principle. At its simplest a lever is a rigid bar that can be turned freely around a fixed point (*fulcrum*). A load that can be lifted or moved is called *resistance* and the force required to move the load is called *effort*. Consider the example of a kid (resistance) sitting on one side of a seesaw (lever). The fulcrum is in the center of the seesaw. The lever length goes from the kid to a man on the other side trying to lift the kid. The effort he needs to lift the kid depends on the distance between the kid and the fulcrum: the larger this distance, the more effort is required. When we hold a tablet in one hand, the tablet acts as a lever: the fulcrum is the most distant body part of the support link, the finger tips behind the tablet, and the resistance is the weight of the tablet, concentrated at its *balance point*. The palm and part of the thumb need to apply an effort to hold the tablet which depends on the lever length.



example of a lever

For all BiPad holds where the thumb is on the front of the device (T_{side}, T_{corner}, and T_{bottom}), performance is affected by the distance between the support position in the kinematic chain and the balance point of the tablet, the

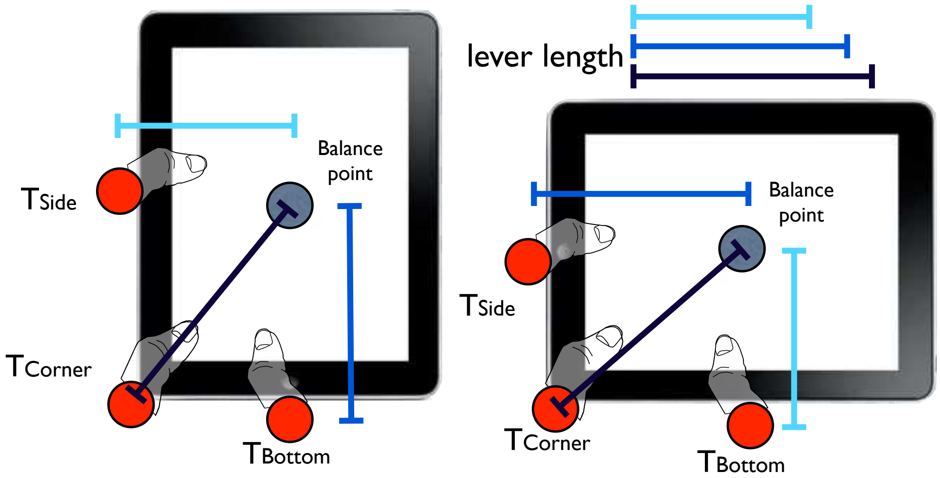


Figure 4.13: The tablet acts as lever. Users perform slower when the lever length increases since they perceive it as heavier.

lever length. The balance point of an iPad is roughly in the center of the device. Figure 4.13 illustrates the three lever lengths for both device orientations for each thumb hold. T_{side} in portrait orientation has the same lever length as T_{bottom} in landscape orientation (see light blue line in Fig. 4.13); similarly, the lever length for T_{bottom} in portrait orientation is identical to the lever length of T_{side} in landscape (see dark blue line in Fig. 4.13); and the lever length of T_{corner} is identical in both orientations (see black line in Fig. 4.13). In portrait device orientation, T_{side} is significant faster than T_{corner} (T_{bottom} is not significantly different from T_{side} and T_{corner}). In landscape orientation, we see a similar result regarding the length of the lever: T_{bottom} is significantly faster than T_{corner} with T_{side} being not significantly different from the two other holds. The longer the distance becomes, the heavier users perceive the tablet, and the less performant is the bimanual interaction with the support hand.

4.9 Contributions

1. Identified ten ways of holding a tablet, five per device orientation, and found that users shift between holds in order to avoid *fatigue* effects.
2. Implemented a toolkit to design interfaces for bimanual interaction with the support hand on off-the-shelf tablets.
3. BiTouch extends Guiard's kinematic chain theory by taking the role of support into account and provides a microscopic view into an input

motor assembly in order to analyze the precise distribution of support and interaction.

4. Device support restricts motion of input motor assemblies. The more body parts are involved into the support, the more comfortable is the interaction perceived by users but the less performant is the interaction.

5

Interaction Effects between Input Motor Assembly and Affected Body Parts

This chapter investigates the interaction effects that can occur when atomic interaction techniques are combined in parallel. Touching one's own body (on-body touch) and gesturing with the hands in the air (mid-air pointing) offers a rich set of commands that can be triggered by touching on-body targets, such as the elbow or shoulder, in an eyes-free and hand-free maneer. This compound technique involves the dominant arm which forms one input motor assembly (IMA_1) to perform mid-air pointing, and the non-dominant arm which forms the second input motor assembly (IMA_2) to perform on-body touch. BodyScape predicts that IMA_1 might interfere with IMA_2. We can expect a subsequent decrease in performance and precision, since IMA_1 is performing one atomic interaction technique while being affected by IMA_2 which performs the second atomic interaction technique.

5.1 Introduction

BodyScape predicts that interaction effects between two atomic interaction techniques in a compound task can occur between input motor assembly and affected body parts. The previous chapter investigated an atomic interaction technique that simultaneously involves and affects the input motor assembly. BiPad enables hand-held touch interaction with the support hand: the interaction technique simultaneously involves the non-dominant arm as an input motor assembly and affects it by supporting. This chapter investigates the case where the input motor assembly of one atomic interaction technique affects *another* input motor assembly involved into another task[1].

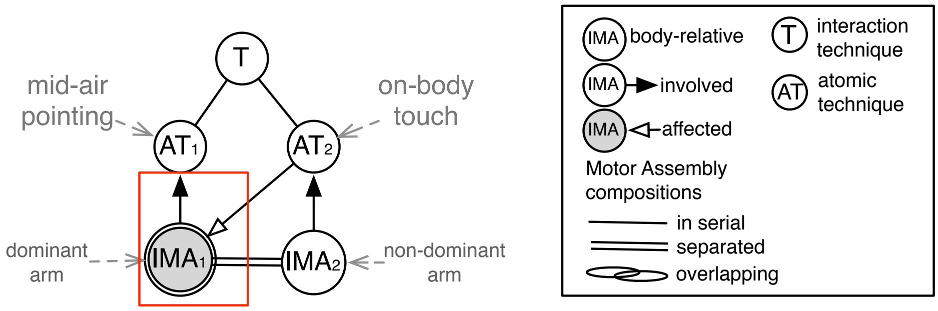


Figure 5.1: A compound technique T involves two atomic interaction techniques, mid-air pointing (AT_1) and on-body touch (AT_2). Mid-air pointing involves the dominant arm (IMA_1) as input motor assembly. On-body touch involves the non-dominant arm (IMA_2) and affects all body parts where on-body targets are located, and especially IMA_1 which is performing mid-air pointing (red frame).

Figure 5.1 illustrates a compound technique T that combines two atomic interaction techniques AT_1, mid-air pointing, and AT_2, on-body touch, allowing to perform two control tasks. Mid-air pointing involves the dominant arm [2] (IMA_1) and coordinates its movements according to a fixed target in the world. On-body touch involves the non-dominant arm as input motor assembly (IMA_2). In addition, on-body touch (IMA_2) affects IMA_1 when *on-body targets* are located on the dominant arm (red frame in Fig. 5.1). This compound technique allows users to point to distant fixed virtual objects on stationary devices using the dominant hand and to trigger commands by pointing (relative to the body) to on-body targets in order to apply the command to the object. It has the advantage to be device-free, and to rely on simple on-body touch gestures to improve interaction. However, as suggested by the BodyScape representation in Figure 5.1, potential interferences between input motor assemblies should be studied in details.

[1]This work is the result of a collaborative effort with Mathieu Nancel (Ph.D student at INSITU), Sean Gustafson (Ph.D student at Hasso Plattner Institute in Potsdam, Germany), and both my advisors Stéphane Huot and Wendy Mackay. The paper is currently in submission process and appears in appendix C—“Selected Publications” (page 159).

[2]According to Guiard’s observation on bimanual interaction [Gui87] the dominant hand is able to perform more precise actions than the non-dominant hand.

5.1.1 Related Interaction Techniques

WILD users require mid-air interaction techniques in order to interact with data displayed on the wall-sized display (see chapter 2). Controlling large displays raises challenges that are addressed by a variety of technologies: e.g., many blackboard-sized displays, such as *SmartBoards*, are touch sensitive. However, this approach does not scale well to wall-sized displays, since users cannot physically reach the entire surface.

Mid-air interaction allows users to interact at any distance from the screen using freehand techniques [VB05; STB07], or by *physical navigation* such as zooming a large image simply by walking towards it [BNB07].

Mid-air pointing in combination with on-body touch techniques provides users the means to apply commands to a specific screen area or virtual object. The dominant hand performs a pointing gestures [Gui87] and the non-dominant hand touches *targets* on the body, such as the elbow. Klemmer et al. argue that using the body enhances both learning and reasoning [KHT06] and whole-body interaction has been applied successfully in gaming [Sho+11], when controlling multimedia dance performances [Lat+10] and for skilled, hands-free tasks, such as surgery ([Wac+08]).

Previous work demonstrated that interaction with virtual objects benefits from kinesthetic memory and proprioceptive cues: Li et al. [LDT10] found that users could more easily remember where they placed items on *Virtual Shelves*, as they oriented a mobile device within the hemisphere before them, in their own personal space. Mine et al. [MBS97] explored the body-relative gestures, such as a over-the-shoulder toss gesture for deletion, to compensate for the lack of haptic contact with real objects in immersive virtual environments. Shoemaker et al. extract the user's silhouette to provide visual feedback about the user's body position [Sho+10]. Users can select menu items by pointing a Wii remote to specific parts of the body, such as the torso, without touching them. Interaction with virtual objects also benefits from direct touch [For+07]. Gustafson et al. found that users retain spatial memory of the location of icons they habitually select on a mobile touch screen [GHB11]. They can transfer this knowledge to accurately select the associated locations of these icons on the palms of their hands. Ängeslevä et al. found that certain body locations have special significance and are easier to remember, specifically hip pockets, the stomach, the head and the heart [Ang+03].

I am interested in investigating body-relative touch input, so as to gain the benefits of kinesthetic memory and proprioception. Lin et al. use an ultrasonic sensor mounted on the user's wrist to track up to seven different input locations [Lin+11]. They found that haptic feedback increases the accuracy of on-body touches and that users can discriminate among six distinct locations on the forearm. *PinStripe* [Kar+11] detects pinching and rolling gestures with clothing made of smart fabrics. All of these techniques represent several point design in BodyScape. By combining these atomic interaction techniques to more complex techniques, we can support the accomplishment of more complex tasks in multi-surface environments.

5.2 Studying On-body Touch and Mid-air Pointing as a Compound Technique

We can use the BodyScape to look systematically at different types of body-centric interaction techniques, both in their atomic form and when combined into compound interaction techniques. The work with users in complex multi-surface environments of the INSITU group (see Chapter 2—"Multi-surface Environments: Users, Tasks, and Design Challenges") highlighted the need for interaction techniques that go beyond simple pointing and navigation. Users need to combine techniques as they interact with complex data spread across multiple surfaces. BodyScape presented in chapter 3 suggests a number of possibilities for both atomic and compound interaction techniques that we can now compare and contrast.

We chose two techniques, illustrated in Figure 5.2d, on-body touch input, and 5.2g, mid-air pointing input, both with visual output on a wall display, which is where WILD users typically need to interact with their data. Although the latter has been well-studied in the literature [VB05; MJ01], we know little of the performance and acceptability trade-offs involved in touching on-body targets to control a multi-surface environment. We are particularly interested in using on-body touch for secondary tasks such as confirming a selection, triggering an action on a specified object, or changing the scope or interpretation of a gesture.

We are interested in how both atomic interaction techniques compare with each other: mid-air pointing is a world-fixed mid-air gesture that restricts movement more than on-body touch which is body-relative (Fig. 5.2g vs. 5.2d). On-body touch does not only involve the arm into the interaction, but also *affects* body parts where on-body targets are located. Figure 5.3 classifies related interaction techniques regarding their *level of body restriction* and the *number of involved and affected body limbs*. To simplify, I have divided the body into five groups of body parts: dominant/non-dominant

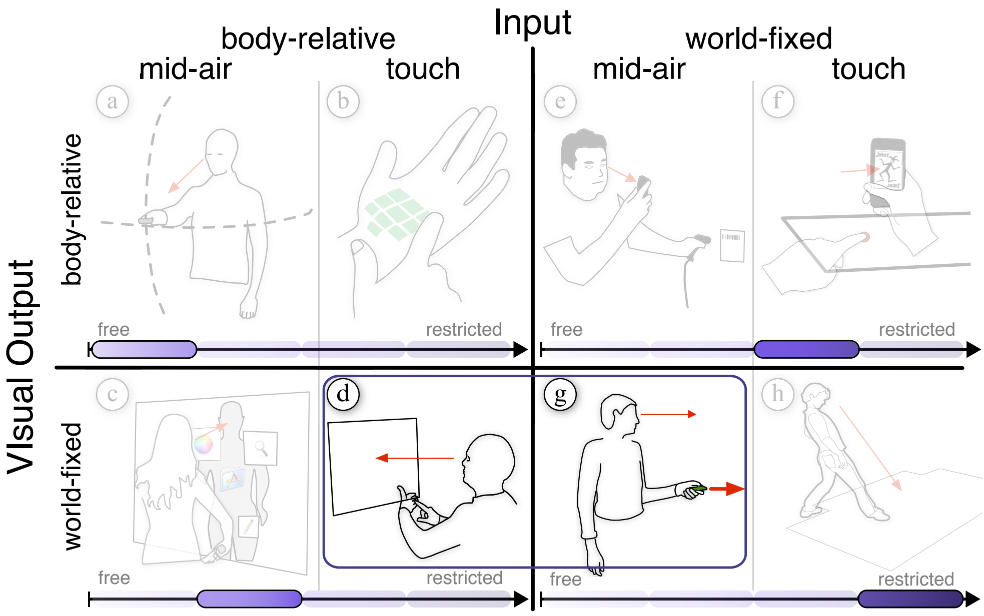


Figure 5.2: Two techniques introduced in the BodyScape framework in chapter 3 (Fig. 3.5) are combined into a compound technique: d) body-relative on-body touch interaction while visually attending to a large display and g) mid-air pointing towards a world-fixed virtual target.

arms, dominant/non-dominant leg and torso. Each atomic interaction technique has exactly one input motor assembly, but can affect other body parts.

The compound technique combining mid-air pointing and on-body touch is illustrated in purple in figure 5.3. mid-air pointing on large displays is *world-fixed input and output*: it is located on the right (restricted) side of the scale. Mid-air pointing involves only one body part, the dominant arm, as input motor assembly. On-body touch is *body-relative input and output*: it is located on the left (unrestricted) side of the scale. Since the non-dominant hand can not touch itself, on-body touch performed on the entire body affects the four other body parts. Thus, on-body touch involves and affects in sum five body parts. It obtains for compound technique performed in parallel that the more restricted atomic interaction technique dominates the overall level of restriction: the compound technique remains on the (restricted) side of the scale and has a total of five involved and affected body parts.

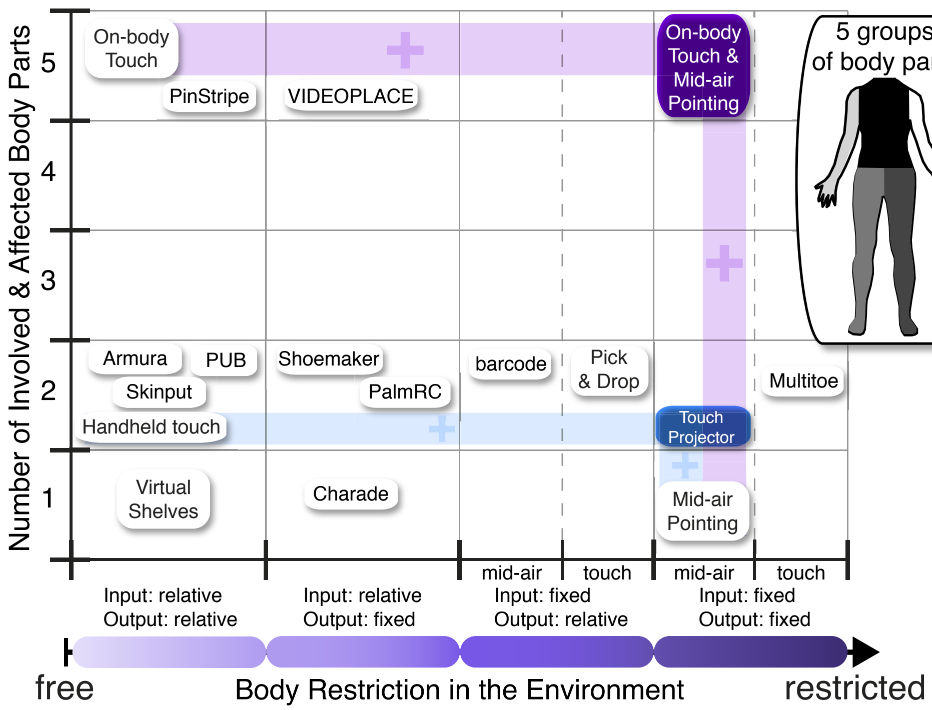


Figure 5.3: Atomic interaction techniques classified by their level of body restriction in the environment and the total involved and affected body parts (see Fig. 3.7). The compound technique, illustrated in violet, combines two extreme cases of this framework: on-body touch offers body-relative input with a high sum of involved and affected body parts; mid-air pointing is a world-fixed input just involving the dominant arm.

5.3 Experiment

The relationship between body parts that are *involved* into the interaction (forming an input motor assembly) and body parts that are *affected* by the interaction is dynamically changing and depends on which on-body targets are touched. Figure 5.4 illustrates on-body touch interaction for three on-body targets: the non-dominant arm is involved into the interaction and touches, (a) the torso (b) the other arm or (c) the leg. MID-AIR POINTING (see Fig. 5.4d) involves the dominant arm. The combination of MID-AIR POINTING and ON-BODY TOUCH shows that the dominant arm is both involved into the interaction and affected by the non-dominant arm (see Fig. 5.4e).

Creating *compound* interaction techniques is of interest to increase the size of the command vocabulary and offer users more nuanced control. However, because this involves coordinating two controlled movements, we need to

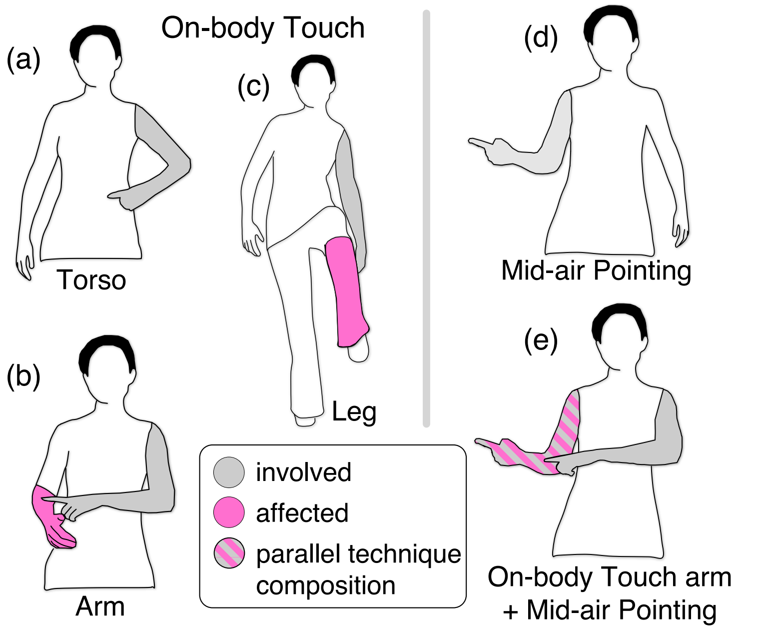


Figure 5.4: Body parts involved when touching the (a) torso, (b) arm, (c) leg; (d) mid-air pointing; and (e) in parallel, when the dominant hand points in mid-air and non-dominant hand touches the dominant arm.

understand any potential interaction effects. The following experiment investigates the two atomic interaction techniques, MID-AIR POINTING and ON-BODY TOUCH, which also act as baselines for comparison with a compound technique that combines them.

1. *Which on-body targets are most efficient and acceptable?* Users can take advantage of proprioception when touching their own bodies, which enables eyes-free interaction and suggests higher performance. However, body targets differ both in the level of motor control required to reach them, e.g., touching a foot requires more balance than touching a shoulder, and in their social acceptability, e.g., touching below the waist [Kar+11].
2. *What performance trade-offs obtain with compound body-centric interaction techniques?* Users must position themselves relative to a target displayed on the wall and stabilize the body to point effectively. Simultaneously selecting on-body targets that force shifts in balance or awkward movements may degrade pointing performance. In addition, smaller targets will decrease pointing performance, but may also decrease ON-BODY TOUCH performance.

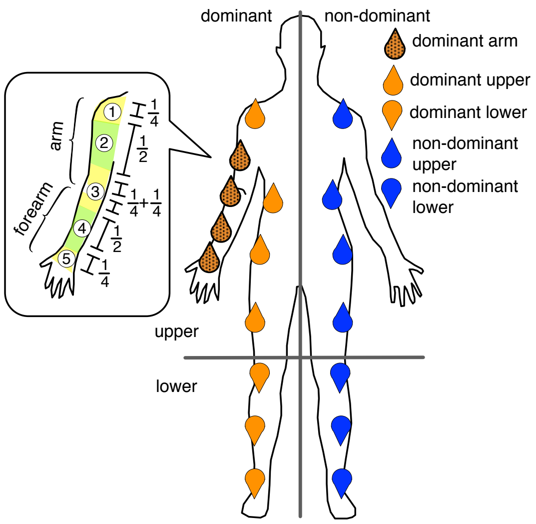


Figure 5.5: 18 body targets are grouped into five categories.

5.3.1 Method

Participants We recruited sixteen unpaid right-handed volunteers (13 men, average age 28); five had previous experience using a wall-sized display. All had good to excellent balance (median 4 on a 5-high Likert scale) and practiced at least one activity that requires balance and body control. All wore comfortable, non-restrictive clothing.

Apparatus We used ***On-body Touch Hi***, a high-fidelity prototype detecting on-body touch presented in appendix B—"On-body Touch Prototype". To track pointing at targets on the wall-sized display, users held a wireless mouse with mounted VICON tracking markers in their dominant hand[3]

Based on pilot studies, we defined 18 body target locations distributed across the body (Fig. 5.5), ranging in size from $9cm$ on the forearm to $16cm$ on the lower limbs, depending upon location and density of nearby targets, grouped as follows:

Dominant Arm: 4 targets on dominant arm (D ARM = upper arm, elbow, forearm, wrist)

Dominant Upper Body: 4 targets on dominant side of upper body (D UPPER = thigh, hip, torso, shoulder)

[3] All subjects were right-handed, so "dominant" refers to the right hand or side.

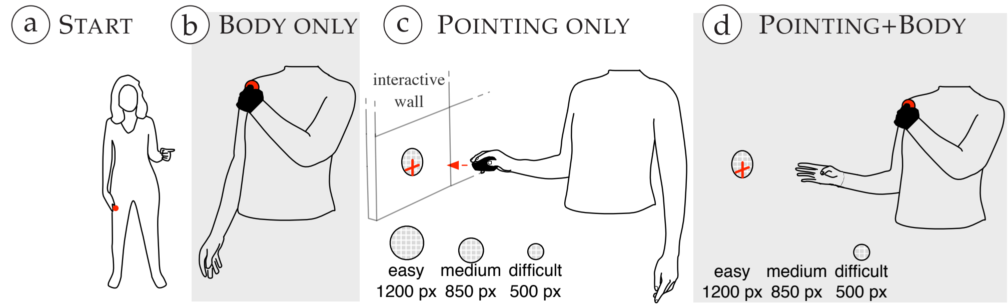


Figure 5.6: Starting position (a) and experimental conditions: (b) BODY ONLY (c) POINTING ONLY (d) POINTING+BODY

Non-dominant Upper Body: 4 targets on non-dominant side of upper body (ND UPPER = thigh, hip, torso, shoulder)
Dominant Lower Leg: 3 targets on dominant side of lower leg (D LOWER = knee, tibia, foot)
Non-dominant Lower Leg: 3 targets on non-dominant side of lower leg (ND LOWER = knee, tibia, foot)

Arms and legs involve multiple limb segments: forearms, upper arms and lower legs, each with three potential targets (Figure 5.5, left). Targets at the center of a limb segment cover one half its length, whereas targets at the ends of a limb segment, e.g. the elbow and shoulder, cover one quarter of its length. Most of end targets overlap with neighboring limb segments, e.g., target *3* in Fig. 5.5a, and thus cover the same percentage of the body. The exceptions are body targets located at the extremities: wrists (*5*) and feet.

Wall pointing tasks varied in difficulty from *easy* (diameter of the circular target was $1200px$ or $30cm$) to *medium* ($850px$ or $21.25cm$) to *hard* ($500px$ or $12.5cm$). Wall targets were randomly placed $4700px$ ($117.5cm$) from the starting target.

Data Collection We collected timing and error data for each trial, as follows:

TRIAL TIME**:** from trial start to completion.
POINTING REACTION TIME**:** from appearance of on-screen target to cursor displacement of more than $1000px$.
POINTING MOVEMENT TIME**:** from initial cursor movement to entry into goal target.
CURSOR READJUSTMENT TIME**:** from leaving goal target to relocating cursor onto goal target.

BODY REACTION TIME: from appearance of trial stimulus to leaving starting position.
BODY POINTING TIME: from leaving start position to touching on-body target.
BODY ERRORS: number of incorrect touches detected on body target[4]; includes list of incorrect targets per error.

We debriefed participants at the end of the experiment and asked them to rank on a Likert scale: (i) perceived comfort of each body target according to each MID-AIR POINTING condition ('1=very uncomfortable' to '5=very comfortable'); and (ii) social acceptability of each on-body target:"*Would you agree to touch this body target in a work environment with colleagues in the same room?*" ('1=never' to '5=certainly').

Procedure Each session lasted about 60 minutes, starting with a training session, followed by blocks of trials of the following conditions, counterbalanced across subjects using a Latin square.

BODY ONLY: Non-dominant hand touches one of 18 on-body targets (atomic interaction technique − 18×5 replications = 90 trials)
POINTING ONLY: Dominant hand points to one of three target sizes (atomic interaction technique − 3×5 replications = 15 trials)
POINTING+BODY: Combines touching an on-body target with selecting a wall target (compound technique − (18×3)×5 replications = 270 trials)

Participants were thus exposed to 75 unique conditions, each replicated five times, for a total of 375 trials. BODY ONLY and POINTING+BODY trials were organized into blocks of six, with the location of body targets randomized and no two successive trials involved the same body target group. POINTING ONLY trials were organized into blocks of five and all wall pointing trials were counterbalanced across difficulty. The two atomic interaction techniques, BODY ONLY and POINTING ONLY serve as baseline comparisons for performance with the compound interaction technique, POINTING+BODY. Participants were instructed to perform trials as quickly and accurately as possible.

BODY ONLY (Fig. 5.6b): The starting position involves standing comfortably facing the wall display, with the non-dominant hand at the thigh (Fig. 5.6a). The trial begins when a body-target illustration appears on the wall. The participant touches that target with the index finger of the non-dominant hand as quickly and accurately as possible. Participants were asked to avoid crouching or bending their bodies, which forced them to lift their legs to reach lower-leg targets. The trial ends only when the participant successfully selects the

[4] Includes both system detection and user errors.

correct target; all intermediate incorrect selections are logged.

POINTING ONLY (Fig. 5.6c): The starting position involves standing comfortably facing the wall display and using the dominant hand to locate a cursor within a circular target displayed in the center of the wall. The trial begins when the starting target disappears and goal target appears between $0.5s$ and $1s$ later, to reduce anticipatory movements and learning effects. The participant moves the dominant hand to move the cursor to the goal target and selects by pressing the left button of the mouse bearing the optical marker used for pointing. The trial ends only when the participant successfully clicks the mouse button while the cursor is inside the goal target.

POINTING+BODY (Fig. 5.6d): The starting position combines the above, with the non-dominant hand at the thigh and the dominant hand pointing to the starting target on the wall. The trial begins with the appearance of a body-target illustration and the goal target on the wall display. The participant first points the cursor at the goal target, then completes the trial by touching the designated on-body target. The trial ends only when the on-body touch occurs while the cursor is inside the goal target on the wall. As in the BODY ONLY condition, multiple body parts may be involved, sometimes with adverse effects.

Training Participants began by calibrating the system to their bodies, visually locating, touching and verifying each of the 18 body targets. They were then exposed to three blocks of six BODY ONLY trials, with the requirement that they performed two on-body touches in less than five seconds. They continued with three additional blocks to ensure they could accurately touch each of the targets. Next, they were exposed to all three levels of difficulty for the POINTING ONLY condition: easy, medium and hard, in a single block. Finally, they performed three additional blocks of the compound POINTING+BODY technique.

5.3.2 Results

We conducted a full factorial ANOVA on the BODY ONLY condition, with PARTICIPANT as a random variable based on the standard repeated measures (REML) technique from the JMP 9 statistical package. We found no fatigue or learning effects.

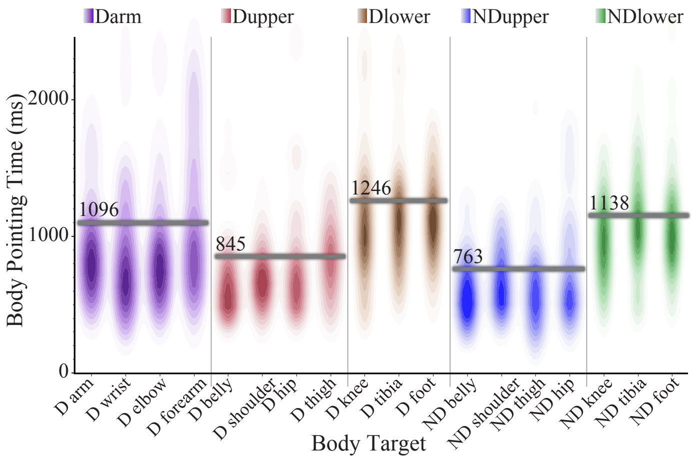


Figure 5.7: Mean BODY POINTING TIME is faster for both upper body targets (D UPPER and ND UPPER) compared to other targets. Horizontal lines indicate group means; performance within groups is consistent.

5.3.3 Q1: Efficiency & Acceptability of On-body Targets

Our first research question asks which on-body targets are most efficient and which are socially acceptable. Figure 5.7 shows the times for touching all 18 on-body targets, grouped into the five body areas. We found significant effects of BODY TARGET on BODY POINTING TIME: touching lower body targets is slower. Since BODY POINTING TIME is consistent for targets within a given target group, we report results according to target group, unless otherwise stated.

Overall, we found a main effect of BODY TARGET GROUP on TRIAL TIME ($F_{4,60} = 21.2$, $p < 0.0001$). A post-hoc Tukey test revealed two significantly different groups: body targets located on the upper torso required less than $1400ms$ to be touched whereas targets on the dominant arm and on the lower body parts required more than $1600ms$. Results are similar for BODY POINTING TIME with a significant effect of BODY TARGET GROUP only for the D UPPER group ($F_{3,45} = 5.07, p = 0.004$), specifically, targets on the dominant thigh are touched more slowly than those on the shoulder or torso. For BODY REACTION TIME, despite a significant effect, values are very close for each BODY TARGET GROUP ($530ms \pm 20ms$).

Participants were able to quickly touch on-body targets with an accuracy of 92.4% on the first try. A post-hoc Tukey test showed that targets on the dominant arm were more prone to errors than other body target areas (14.8% vs. 6% for dominant and non-dominant upper body and 2.9% for non-

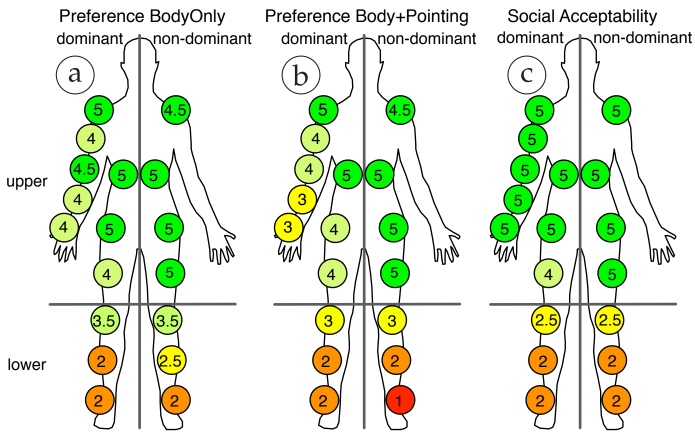


Figure 5.8: Median preference and acceptability rankings of on-body targets (from green = acceptable to red = unacceptable).

dominant lower body targets). Most errors obtained when targets were close to each other, i.e. when the participant's hand touched the boundary between the goal and a nearby target or when the dominant arm was held close to the torso, making it difficult to distinguish between the torso and arm targets. Touching lower body parts is, not surprisingly, slower, since these targets are further from the starting position and require more complex movements. However, the difference is small, about $200ms$ or 12% of global trial time.

Qualitative measures of Preference and Social Acceptance

Figure 5.8a shows that participants' preferences (median values of Likert-scale) for and within each BODY TARGET GROUP were consistent with performance measures: targets on the upper body parts were preferred over lower body parts (consistent with [Kar+11]) and targets on the torso were slightly more preferred than on the dominant arm.

Interestingly, preferences for non-dominant foot and the dominant arm decrease when on-body touch interaction is combined with mid-air pointing (Fig. 5.8b). The latter is surprising, given that the most popular location for on-body targets in the literature is on the dominant arm. This suggests that interaction designers should explore alternative on-body targets as well. Social acceptability varies from highly acceptable (upper body) to unacceptable (lower body) (Figure 5.8c).

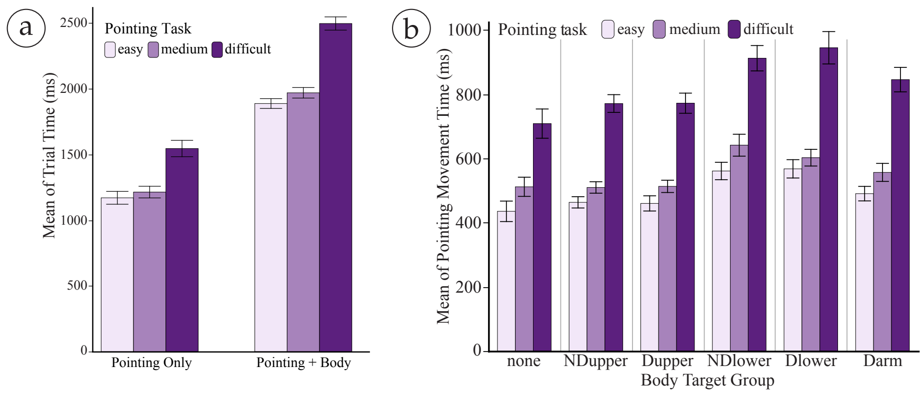


Figure 5.9: TRIAL TIME for (a) BODY ONLY and (b) POINTING+BODY, by pointing difficulty; and (c) Interaction Pointing×Body on POINTING MOVEMENT TIME.

5.3.4 Q2: Performance Trade-offs for compound techniques

The second research question examines the effect of combining two atomic interaction techniques, in this case BODY ONLY and POINTING ONLY, into a single compound technique. We treat these atomic interaction techniques as baseline values to help us better evaluate the compound task.

Pointing Only task

Not surprisingly, hard pointing tasks are significantly slower ($1545ms$) (TRIAL TIME ($F_{2,30} = 40.23$, $p < 0.0001$) than medium ($1216ms$) or easy ($1170ms$) tasks, which are not significantly different from each other (Fig. 5.9a). POINTING REACTION TIME is also significantly slower for difficult ($498ms$) as opposed to medium ($443ms$) or easy ($456ms$) tasks. POINTING MOVEMENT TIME is significantly different for all three levels of difficulty: hard ($708ms$), medium ($511ms$) and easy ($435ms$).

Participants made few errors but occasionally had to relocate the cursor inside the goal target before validating the selection with the mouse. This occurred rarely (1.8% of all trials), but CURSOR READJUSTMENT TIME was significantly more likely for difficult pointing tasks (15%) ($F_{2,30} = 8.02$, $p = 0.0016$) and accounts for the differences in TRIAL TIME and POINTING MOVEMENT TIME.

Compound POINTING+BODY task

Figure 5.9a shows that the combined MID-AIR POINTING and ON-BODY TOUCH task is significantly slower than MID-AIR POINTING alone for all levels of difficulty. TRIAL TIME is significantly slower for difficult MID-AIR POINTING ($2545ms$) than both medium ($1997ms$) and easy ($1905ms$) tasks. In fact, the easiest compound task is significantly slower that the hardest POINTING ONLY task.

BODY TARGET GROUP also has an effect on TRIAL TIME ($F_{4,60} = 34.1, p < 0.0001$) with the same significant groups as for BODY ONLY: TRIAL TIME is significantly faster when touching upper body targets (ND UPPER $= 1794ms$, D UPPER $= 1914ms$) than lower body targets (ND LOWER $= 2267ms$, D LOWER $= 2368ms$) or the dominant arm (D ARM $= 2401ms$). BODY REACTION TIME is faster than POINTING REACTION TIME, regardless of pointing difficulty.

How ON-BODY TOUCH affects MID-AIR POINTING?
Although we can see that the individual techniques are both more efficient than the compound technique, the question is why? Just how does ON-BODY TOUCH affect MID-AIR POINTING? Figure 5.9b shows interaction effects between the two elements of the compound tasks, by both BODY TARGET GROUP and pointing difficulty. While POINTING MOVEMENT TIME is close to the pointing baseline (BODY TARGET GROUP “none”) for all difficulties when MID-AIR POINTING is combined with ON-BODY TOUCH on the upper body parts, we observe a stronger negative effect for the lower body parts and the dominant arm, especially for difficult pointing tasks.

This impact of ON-BODY TOUCH on the MID-AIR POINTING task does not only relate to the movement phase but also cursor readjustments. For the combined POINTING+BODY task, 31% of the trials required the participants to relocate the cursor inside of the target before validating the selection with a body touch, compared to only 6% for POINTING ONLY. Thus, we found significant effects of MID-AIR POINTING ($F_{2,30} = 59.64, p < 0.0001$), BODY TARGET GROUP ($F_{5,75} = 23.03, p < 0.0001$) and MID-AIR POINTING$\times$BODY TARGET GROUP ($F_{10,150} = 8.45, p < 0.0001$) on CURSOR READJUSTMENT TIME. As shown in Figure 5.10, CURSOR READJUSTMENT TIME increases significantly for each level of difficulty of MID-AIR POINTING but selecting body targets on some BODY TARGET GROUP, especially in D LOWER and D ARM, affects the body configuration and requires even more time to relocate the cursor inside of the on-screen target.

This result reveals (1) touching the dominant arm while pointing affects the

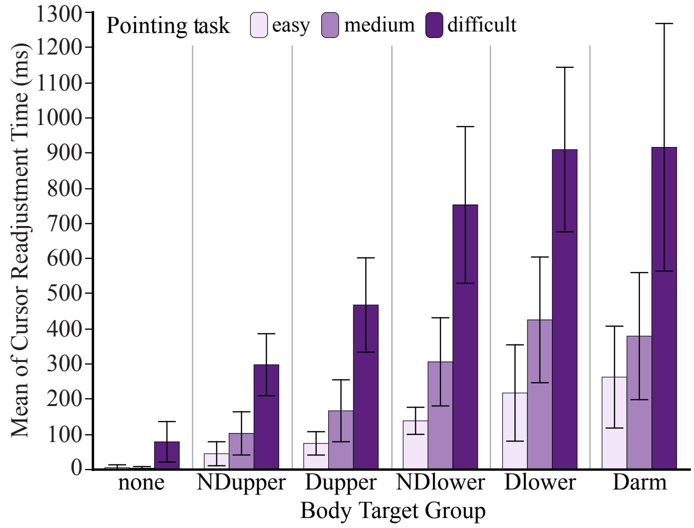


Figure 5.10: Effect of Pointing difficulty and BODY TARGET GROUP on CURSOR READJUSTMENT TIME.

precision of pointing and requires "force-balance" (targets on D ARM); and (2) touching targets on the lower body parts affects the precision of pointing and requires "movement-balance" (targets on ND LOWER and D LOWER). Overall, since the impact of both D LOWER and D ARM is similar, we observe that maintaining force-balance is as difficult as maintaining movement-balance during the pointing task, and that the difficulty in movement-balance is not only caused by standing on one leg, but by simultaneously crossing the body's sagittal plane (difference between D LOWER and ND LOWER).

How MID-AIR POINTING **affects** ON-BODY TOUCH**?**
Similarly, we studied the effect of MID-AIR POINTING on ON-BODY TOUCH by performing an ANOVA with the model POINTING[none /easy/medium/difficult]×BODY TARGET GROUP. We did not find any effect on BODY REACTION TIME. On BODY POINTING TIME, we did find a significant effect of BODY TARGET GROUP ($F_{4,60} = 38.69$, $p < 0.0001$), of MID-AIR POINTING ($F_{3,45} = 78.15$, $p < 0.0001$) and a significant MID-AIR POINTING×BODY TARGET GROUP interaction ($F_{12,180} = 2.28$, $p = 0.01$). The main effect of BODY TARGET GROUP is similar to the baseline (with ND UPPER and D UPPER significantly faster than all other groups). The main effect of MID-AIR POINTING is also similar to those observed before, showing that difficult pointing tasks make simultaneous body touching slower than medium or easy pointing task. Obviously, these are all significantly slower than the BODY ONLY baseline.

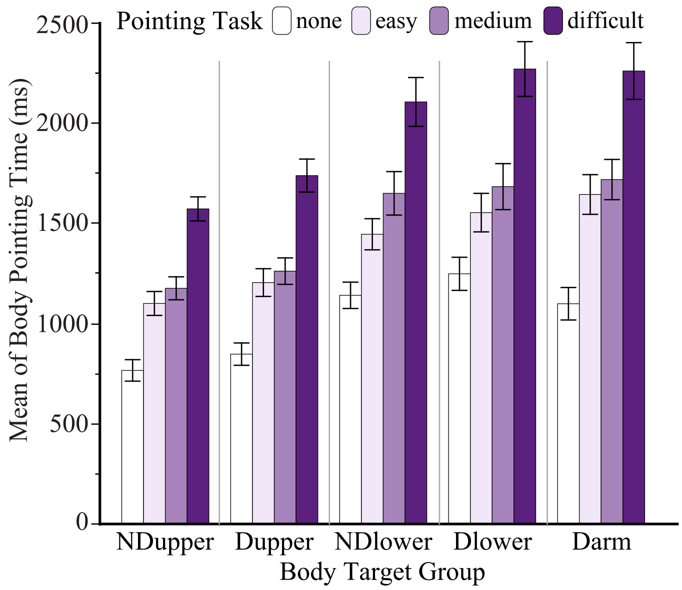


Figure 5.11: Interaction Pointing×Body on BODY POINTING TIME.

More interesting, the MID-AIR POINTING×BODY TARGET GROUP interaction effect reveals the actual impact of MID-AIR POINTING on ON-BODY TOUCH. As shown in Figure 5.11: (i) the increasing difficulty of the pointing task increases BODY POINTING TIME. In fact, despite how our task required body target selection to be the last action, the reaction times indicate that both tasks start almost simultaneously (ON-BODY TOUCH even before MID-AIR POINTING); (ii) this increase of difficulty does not change the difference between the groups of targets, but amplifies these differences. ND UPPER and D UPPER are still the group of targets that require less time to be touched.

Adding the MID-AIR POINTING task also changes how BODY TARGET affects BODY POINTING TIME. Recall that for the BODY ONLY baseline, BODY POINTING TIME for body targets inside of a same BODY TARGET GROUP were not significantly different, except for one target of D UPPER. However, an ANOVA with the model MID-AIR POINTING×BODY TARGET grouped by BODY TARGET GROUP shows the obvious significant effect of MID-AIR POINTING but also of BODY TARGET for of all BODY TARGET GROUP, and some significant MID-AIR POINTING×BODY TARGET interaction effects.

These results suggest that, whereas BODY TARGETS of a BODY TARGET GROUP are as easy to reach when ON-BODY TOUCH is performed alone, differences appear when combined to another task, and that it is even more significant as the difficulty of the other task increases. The summarized noticeable results are: (i) for the lower parts of the body, targets on the *feet* and the *tibias* are now slower to touch than targets on the *knees*. In fact, the lower the target is, the more it requires to change the body balance to be reached, and participants were more careful in order to not impair their pointing precision; (ii) for the D ARM group, the target located on the *upper arm* is touched around $400ms$ faster than the one located on the *forearm*, suggesting that participants were taking care to not displace the cursor by touching their forearm too precipitately.

BODY TARGET GROUP and MID-AIR POINTING also have an effect on the ON-BODY TOUCH error rate ($F_{4,60} = 12.77$, $p < 0.0001$ and $F_{3,45} = 3.41$, $p = 0.0253$). Post-hoc Tukey test shows that difficult pointing tasks have significantly less body pointing errors (5.2%) than easy and medium (8.1%), but the BODY ONLY baseline is in between, not significantly different to any other conditions (7.3%). Although surprising, this lower ON-BODY TOUCH error rate while performing difficult pointing tasks could be explained by an increased attention of the participants due to the difficulty of the task. For BODY TARGET GROUP, a post-hoc Tukey test shows that targets on the dominant arm are more error prone than those in others groups (15.6% against about 6% for dominant and non-dominant upper, and 2.7% for non-dominant lower). This result is similar to the BODY ONLY baseline and suggests that targets on the dominant arm are the less reliable in both conditions, probably because of detection problems due to their proximity. However, the analysis of the list of erroneous touches give a more interesting insight into the reason for such detection errors. We classified errors into five groups:

1. **Vicinity (57%):** touching the boundary between the goal and a nearby target.
2. **Dominant arm position (33.3%):** dominant arm and torso targets are too close.
3. **Symmetry (4.2%):** touching a symmetrical target with respect to the sagittal plane.
4. **Default target (1.3%):** repeatedly pressing the default position target after stimuli appear.
5. **Completely wrong (4.2%):** no relationship between touched and goal targets.

We found more dominant arm position errors (mostly on the D ARM and D UPPER groups) compared to the BODY ONLY condition. In fact, we observed that participants kept their dominant arm close to their torso in order to help stabilize the cursor while touching body targets, likely causing more recognition errors when they had to touch a target on the dominant arm or the torso. We also observed several *symmetry* and *completely wrong* errors that did not occur in the BODY ONLY condition, probably because participants were sometimes confused by performing two tasks simultaneously. These errors should decrease for frequent users as they become familiar with both the system and the on-body targets.

In summary, the compound POINTING+BODY task involves interaction effects between the two atomic interaction techniques, which not only incur a time penalty when the tasks are performed simultaneously, but also degrades pointing performance for MID-AIR POINTING (fixed in the world) when combined with a body-relative technique that involves and affects multiple limbs. However, our results also reveal that ON-BODY TOUCH on the lower parts of the body significantly impair the movement phase of pointing, and that the overall negative impact increases with the difficulty of the pointing task, especially when targeting on the pointing arm.

When users performed on-body touch individually, touching on-body targets on the torso outperformed on-body targets on the dominant arm and legs which is consistent with the reported comfort perceived by users. This is a surprising result giving that recent literature [HTM10; Lin+11] where particular focused on the technical realization of detecting on-body touch on the dominant arm.

5.4 Discussion and Conclusion

Combining on-body touch and mid-air pointing creates an interference between two seemingly independent input motor assemblies because one input motor assembly shares body parts that are affected by the other input motor assembly. It was suggested by the BodyScape representation of the technique in Figure 5.1, that illustrates this dual relationship between the two input motor assemblies IMA_1 and IMA_2.

This is explained by two types of trade-offs users need to make between movements: shifting the body's balance point in order to reach for lower body-targets destabilizes the position control of all body limbs (see Fig. 5.12a); and applying force to motors contained in a second input motor assembly, performing input with respect to a fixed target in the world, requires to stabilize the motor's position (see Fig. 5.12b).

The results suggest to further investigate on-body touch on stable body parts such as the torso rather than the dominant arm. Touching on-body targets, even as single atomic interaction technique was more performant on the torso than on the dominant arm. This is surprising, since on-body touch had been

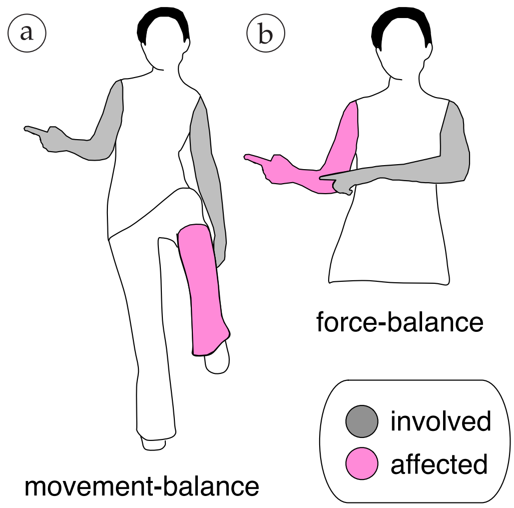


Figure 5.12: Users have to account for two types of balances: a) movement-balance and b) force-balance.

until now studied exclusively on the dominant arm [HTM10; Lin+11].

5.5 Contributions

1. Implemented high- and low- fidelity prototype for an on-body touch interaction technique on the entire body.
2. Applied BodyScape to systematically look at different types of body-centric interaction techniques, both in their atomic form and when combined to a compound technique.
3. Found that two seemingly separated input motor assemblies interfere with each other if one input motor assembly is affecting body parts of another input motor assemblies.
4. Found that one input motor assembly (IMA_1) affects the performance of a world-fixed input motor assembly (IMA_2) if IMA_1 affects body parts necessary for keeping body balance (e.g. legs).

6

Interaction Effects between Two Input Motor Assemblies

*This chapter investigates twelve pan-and-zoom technique for large displays involving two input motor assemblies, IMA_1 for mid-air pointing and IMA_2 for controlling the zoom direction. Input motor assembly IMA_2 varies across techniques in size and in the type of composition with IMA_1 (separated or overlapping). The techniques are based on three key dimensions: handedness (uni-manual vs. bi-manual), gestures type (circular vs. linear), and level of guidance (*1D path, 2D surface, *and* 3D free*). We found that (1) separated input motor assemblies perform better than overlapping, (2) smaller IMA_2 perform faster than larger IMA_2, and (3) guidance can significantly enhance input performance.*

6.1 Introduction

In the two previous chapters, I applied the BodyScape framework to identify potential interaction effects between input motor assemblies and affected body parts. While BodyScape itself does not predict the performance or comfort of specific combinations of input motor assemblies, I have shown that it gives precious insights into the systematic study of potential issues or benefits. In this chapter, I apply the same methodology to investigate the interaction effects between two combined input motor assemblies when they *overlap* (see Fig. 6.1).

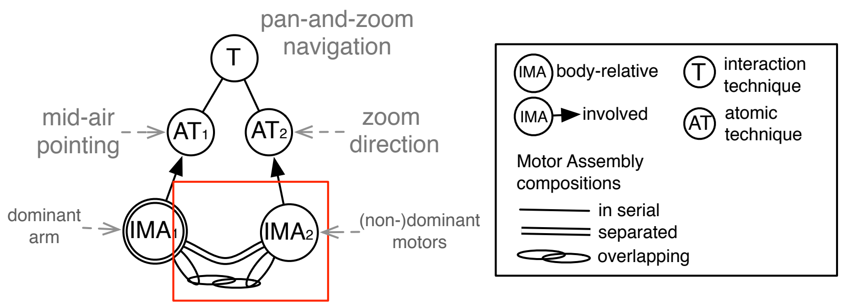


Figure 6.1: Pan-and-Zoom navigation technique T involves two atomic interaction techniques for mid-air pointing (AT_1) and zooming (AT_2). AT_1 involves the dominant arm; the assignment of body parts to IMA_2 is varying across techniques in size, hand or entire arm, and composition type between IMA_1 and IMA_2, separated or overlapping, and the size IMA_2.

My colleagues and I[1] developped twelve pan-and-zoom navigation techniques T for large displays with either overlapping or separate input motor assemblies. The panning technique remains consistent across all techniques. Zoom navigation consists of two atomic interaction techniques that perform two control tasks: (i) AT_1 controls the center of zoom and (ii) AT_2 controls the zoom direction (see Fig. 6.1). Input motor assembly IMA_1 controls the center of zoom using mid-air pointing with the dominant hand and coordinates its movements fixed in the world. The second input motor assembly, IMA_2, controls the *zoom direction* (zoom in/out) using movements coordinated relative to the body. Across techniques, IMA_2 varies in size, i.e. the number of motors, and the types of motor assembly composition established between IMA_1 and IMA_2, either a separated or overlapping composition (red frame in Fig. 6.1):

Overlapping MAs All uni-manual techniques involve the dominant arm for the mid-air pointing task (IMA_1) and either a part or all of the motors of the *dominant arm* for the zoom task (IMA_2).

Separated MAs All bi-manual techniques involve the dominant arm for the mid-air pointing task (IMA_1) and either a part or all of the motors of the *non-dominant arm* for the zoom task (IMA_2).

[1] This is a collaborative work together with Mathieu Nancel (Ph.D student at INSITU), Emmanuel Pietriga, Olivier Chapuis and Wendy Mackay. We published this paper at CHI 2011. It has received a *best paper* award (top 1%). This chapter illustrates how BodyScape pertains to interaction techniques that compose two separated or overlapping input motor assemblies. The full paper appears in appendix C—"Selected Publications" (page 169).

6.2 Motivation and Related Work

The WILD room (described in chapter 2) offers new opportunities for interacting with large data sets. Very-high-resolution wall-sized displays can accommodate several hundred megapixels and make it possible to visualize very large, heterogeneous datasets in many domains [AEN10; BNB07; YHN07]. Astronomers can use them to display telescope images constructed from hundreds of thousands of frames stitched together, such as Spitzer's 4.7 billion pixels images of the inner part of our galaxy (Figure 6.2). Biologists can explore the docking of complex molecules.

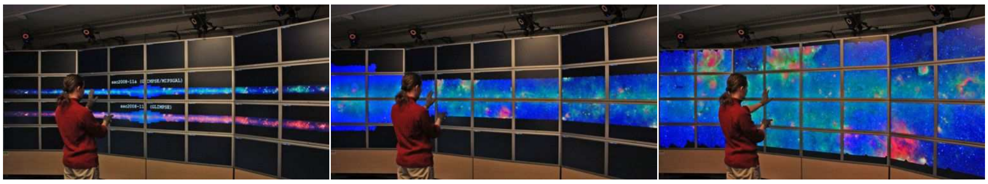


Figure 6.2: Astrophysicist zooming into a large telescope image using a free-hand gesture technique.

While pointing on this type of display has been studied extensively [VB05; MJ01; Pec01], higher-level, more complex tasks such as pan-zoom navigation have received little attention. It thus remains unclear which techniques are best suited to perform multiscale navigation in these environments. High-resolution wall-sized displays pose different sets of trade-offs. It is critical to their success that interaction techniques account for both the physical characteristics of the environment and the context of use, including cooperative work aspects. Input should be location-independent and should require neither a hard surface such as a desk nor clumsy equipment: users should have the ability to move freely in front of the display and interact at a distance [BNB07; YHN07].

This chapter applies the BodyScape framework to compare a set of interaction techniques for pan-zoom navigation on wall-sized displays. Building upon empirical data gathered from studies on bi-manual input [BM86; Gui87; LZB98], the influence of limb segments on input performance [BM97; ZMB96], on types of gestures [MH04; Whe03] and on the integral nature, in terms of perceptual structure, of the pan-zoom task [JS92], we identified three key factors for the design of mid-air pan-and-zoom techniques: *uni- vs. bi-manual* interaction, *linear vs. circular* movements, and *level of guidance* to accomplish the gestures in mid-air (see Table 6.1).

Dimensions		Advantages	Disadvantages
Hands	One	• One hand available for other actions	• Pan and zoom are performed sequentially
	Two	• Pan and zoom can be performed in parallel	• No hand available for other actions
Gesture	Linear	• Direct, natural mapping to zoom actions	• Potentially requires clutching
	Circular	• No clutching (continuous gesture)	• Less natural mapping to zoom actions
Degree of Guidance	1D path	• Input guided by strong haptic feedback • Mainly involves fingers	• Only 1 degree of freedom
	2D surface	• Many degrees of freedom • Mainly involves fingers	• Input guided by limited haptic feedback
	3D free hand	• Many degrees of freedom • No device	• No haptic feedback • Mainly involves whole hand and arms

Table 6.1: Key Dimensions of the Design Space

Uni-manual vs. Bi-manual Input In their paper on the perceptual structure of multidimensional input, Jacob and Sibert claim that panning and zooming are integrally related [JS92]: the user does not think of them as separate operations, but rather as a single, integral task like "*focus on that area over there*". Buxton and Myers [BM86] and later Bourgeois and Guiard [BG02] observed high levels of parallelism for pan-zoom operations, further supporting this argument. The level of parallelism correlates with task performance and is typically well afforded by the use of bi-manual input techniques [Gui87; LZB98]. While we expect bi-manual techniques to outperform uni-manual ones, we are still interested in comparing their performance, as the latter might still be of interest in more complex, real-world tasks that require input channels for other actions.

Linear vs. Circular Gestures Navigating in the scale dimension (zooming in and out) is a task typically performed through vertical scroll gestures on, e.g., a mouse wheel or a touchpad. The mapping from input to command is natural, but often entails clutching as the course of mouse wheels and touchpads is very limited. An alternative consists of mapping continuous circular gestures to zooming. Clockwise gestures zoom in; counter-clockwise gestures zoom out. Despite the less natural mapping from input to commands, such continuous, clutch-free gestures have been successfully applied to vertical scrolling in documents [MH04; Whe03], and to pan and zoom on large, touch-sensitive surfaces in CycloStar [MLG10]. Circular gestures potentially benefit from an automatic Vernier effect [ETW81]: as zooming is mapped to angular movements, the larger the circular gesture's radius, the greater the distance that has to be covered to make a full circle, and consequently the more precise the input.

Guidance through Passive Haptic Feedback Mid-air interaction on wall-sized displays consists primary of freehand techniques based on motion tracking [VB05; Zig+10] and techniques that require the user to hold an input device [BSW06; CB03; LSH99; MI09]. Input devices provide some guidance to the user in terms of what gesture to execute, as all of them provide some sort of passive haptic feedback: A finger operating a knob or a mouse wheel follows a specific path; gestures on touch-enabled devices are made on planar surfaces. Freehand techniques, on the contrary, provide essentially no feedback to the user who can only rely on ***proprioception*** [MBS97] to execute the gesture. We call this dimension the *degree of guidance*. Gestures can be guided to follow a particular path in space (*1D path*); they can be guided on a touch-sensitive surface (*2D surface*) ; or they can be totally free (*3D free*). These three values correspond to decreasing amounts of passive haptic feedback for the performance of input gestures.

6.3 Design Choice

6.3.1 Task Analysis

Figure 6.3 provides an analyzes of the requirements for high-level pan-and-zoom navigation tasks. It consists of two subtasks, "panning" and "zooming" that are performed in sequence. Panning consists of two control tasks, "position control" to drag data to a specific position and "pan command" to switch into the pan mode. Zooming also consists of two control tasks, "position control" to indicate the "focus of expansion" and "zoom direction" (zoom in/out).

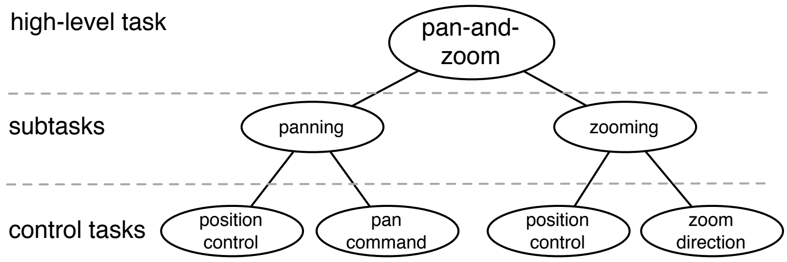


Figure 6.3: A high-level pan-and-zoom navigation task is decomposed into three control tasks: position control, pan command, and zoom direction.

Position Control All conditions using a device achieve the cursor's position by ray-casting from the device to the wall display. All free-hand conditions, achieve the position by ray-casting from the dominant hand to the display (dashed arrows in Figure 6.4). Pointing plays an important role when zooming, as it specifies the focus of expansion (zoom in)/contraction (zoom out). Letting users specify this focus point is important on very large displays, as users rarely stand right in the center. Placing the focus of expansion to the center would make zooming operations tedious and hard to control: every zoom operation would require multiple panning actions to compensate for drifts induced by the offset focus.

Pan command Panning is achieved by dragging, as in applications such as Adobe Illustrator™ or Google Maps™ with their typical hand-shaped cursor. In all conditions users pushed a device button to trigger the "pan" command.

Zoom direction *Linear* techniques, e.g. Google Maps, zoom in by moving forward towards the display and zoom out by moving backwards. *Circular* techniques zoom in by turning clockwise and zoom out by turning counter-clockwise (solid arrows in Figure 6.4).

6.3.2 Choice of Input Motor Assembly

The "panning" subtask is performed consistently for all techniques using the dominant arm for "position control" and a hand-held mouse button to switch into pan mode. Figure 6.5 shows the four different assignments of input motor assemblies for the subtasks "zooming" (see Fig. 6.3). The position control task is accomplished using a mid-air pointing technique (AT_1)

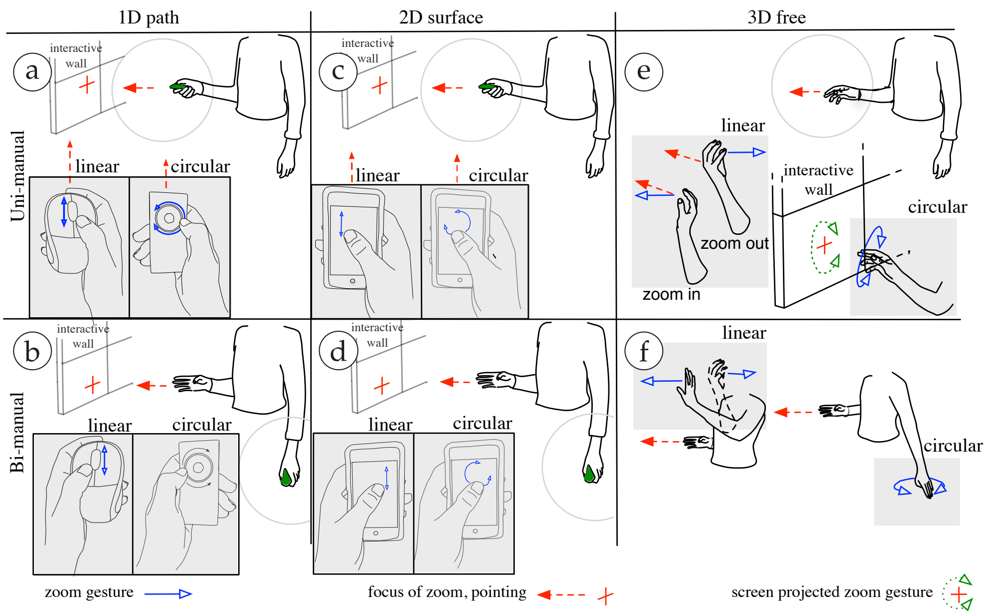


Figure 6.4: Matrix of the 12 techniques organized according to key design dimensions: uni- vs. bi-manual, degree of guidance, linear vs. circular gestures. 1D path involves guiding gestures along a particular path in space; in 2D surface gestures are made on a touch-sensitive surface; while in 3D free gestures are totally free.

which involves the dominant arm (IMA_1) in all conditions to perform input according to a fixed target in the world. The "zoom direction" control task is accomplished using an atomic technique AT_2 that involves IMA_2 in the interaction.

Bi-manual vs. Uni-manual Interaction For all uni-manual techniques, the second input motor assembly (IMA_2) involves motors of the dominant arm and thus forms an overlapping composition between the two input motor assemblies. For all bi-manual techniques, both input motor assemblies are separated. In accordance with Guiard's Kinematic chain model [Gui87], we assign (Figure 6.4, bottom row) pointing and panning to motors of the dominant arm (IMA_1), while motors of the non-dominant hand control zoom, as is typically the case for bi-manual pan-zoom techniques on the desktop [BG02; BM86].

Hand-held Devices All atomic interaction techniques AT_2 that require a phone-sized input device (see Fig. 6.4a-d) involve the thumb and, because

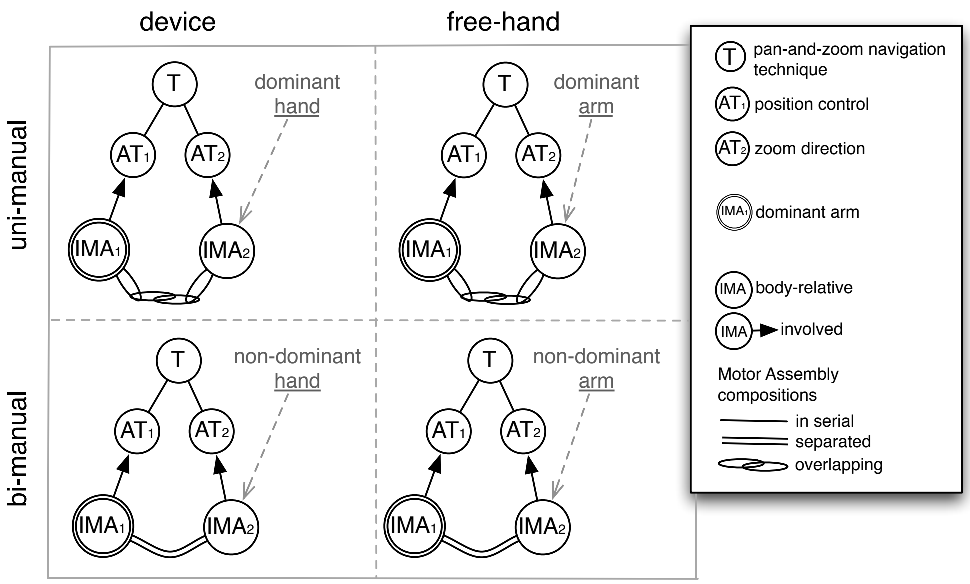


Figure 6.5: The twelve pan-and-zoom techniques are either composed by two overlapping or separated input motor assemblies (uni-manual or bi-manual). All techniques using a device (left column) involve the hand in the input motor assembly. All free-hand techniques (right column) involve the arm in the input motor assembly.

the user must also support the device, the rest of the hand to peform the interaction. Column *1D path* (see Fig. 6.4a-b) illustrates techniques that provide a high degree of guidance for executing the zooming gestures. Users can perform *Linear* gestures using, e.g., a wireless handheld mouse featuring a scroll wheel or *circular* gestures using, e.g., any type of handheld knob. Depressing a button on the device activates drag mode for panning.

Column *2D surface* illustrates techniques that use a touch-sensitive surface for input, providing a lesser degree of guidance. The surface is divided horizontally in two areas. Users zoom in the upper area either by moving the thumb up and down (*linear* case), or by drawing approximate circles (*circular* case). Touching the lower area activates drag mode for panning. Users just rely on proprioceptive information to switch between both areas and do not have to look at the device. These techniques can be implemented with a touch-sensitive handheld device such as a PDA or smartphone.

1D path techniques employing *circular* gestures will provide more guidance, but will not benefit from the earlier-mentioned Vernier effect, as input is constrained to one specific trajectory. However, the range of amplitudes that can be covered with the thumb is limited [RLG09]. This should minimize the trade-off between *1D path* and *2D surface* in that respect. For *2D surface* techniques, rubbing gestures [OFH08] were considered to avoid clutching when performing linear gestures, but were found to be impractical when performed with the thumb on a handheld touch-sensitive surface. As a technique designed specifically for thumb input, we were also interested in MicroRolls [RLG09]. However, these were originally designed for discrete input and circular MicroRolls are not precise enough for zoom control.

Free-hand Interaction All techniques using free-hand gestures (see Fig. 6.4, column *3D free*) involve the entire arm into the input motor assembly performing either *linear* or *circular* gestures. The technique using *circular* gestures is actually very close to the CycloStar zooming gesture [MLG10], but performed in mid-air, without touching any surface. Users perform circular gestures with the dominant hand and forearm oriented toward the display. As in CycloStar, the focus of expansion is the centroid of the round shape corresponding to the cursor's circular path, here projected on the display surface (dotted arrow in Figure 6.4-e). The technique using *linear* gestures consists in pushing the dominant hand forward to zoom in, as if reaching for something, with the palm towards the target. Turning the hand and pulling backward (away from the display) zooms out. Users point orthogonally to the palm of the same hand (red arrows in Figure 6.4-e, left side), with the arm slightly tilted for greater comfort. All other postures and movements being ignored by the system for the non-dominant hand, the user can easily clutch.

6.3.3 Investigating Twelve Pan-and-zoom Navigation Techniques

We conducted an experiment using a [2×2×3] within-subjects design with three primary factors: HANDEDNESS $\in$ {*OneHanded*, *TwoHanded*}, GESTURE $\in$ {*Circular*, *Linear*}, and GUIDANCE $\in$ {*1DPath*, *2DSurface*, *3DFree*} to evaluate the 12 unique interaction techniques. We controlled for potential distance effects by introducing the DISTANCE between two consecutive targets as a secondary within-subjects factor. We systematically varied these factors in the context of a multiscale navigation task within a wall-sized display environment.

Measures include performance time and number of overshoots, treated as errors. Overshoots occur when participants zooms beyond the target zoom level, and indicate situations in which the participant has less precision of control over the level of zoom. For instance, from an overview of Canada, zooming down to street level in Google Maps when what the user actually wanted was to get an overview of Vancouver.

Hypotheses Based on the research literature and our own experience with the above techniques, we made the following 7 hypotheses.

Handedness: prior work [BG02; BM86; GB04; LZB98] suggests that two-handed gestures will be faster than one-handed gestures (*H1*) because panning and zooming are complementary actions, integrated into a single task [JS92]. Two-handed gestures should also be more accurate and easier to use (*H2*).

Gesture: Linear gestures should map better to the zooming component of the task, but should eventually be slower because of clutching, the limited action space compared to zoom range requiring participants to repeatedly reposition their hand/finger (*H3*). Prior work of Moscovich and Hughes [MH04] and Wherry [Whe03] suggests that users will prefer clutch-free circular gestures (*H4*).

Device vs. Free Space: Zhai et al. [ZMB96] suggest that techniques using the smaller muscle groups of fingers should be more efficient than those using upper limb segments. Balakrishnan and MacKenzie [BM97] moderate this observation with findings suggesting that the fingers are not performing better than forearm or wrist for a reciprocal pointing task. Nevertheless, they acknowledge that differences exist in the motor system's ability to control the different limb segments. Based on the gestures to be performed and taking into account the physical size and mass of the segments involved, we predicted that techniques using fingers (*1DPath* and *2DSurface* conditions), should be faster than those requiring larger muscle groups (hands and arms, *3DFree* conditions) (*H5*).

We also predicted that *1DPath* gestures would be faster, with fewer overshoots than techniques with lesser haptic feedback, i.e., *2DSurface* and *3DFree* (*H6*). Finally, we predicted that *3DFree* gestures would be more tiring (*H7*).

Participants We recruited 12 participants (1 female), ranging in age from 20 to 30 years old (average 24.75, median = 25). All are right-handed daily computer users. None are color-blind.

Apparatus Participants performed the experiment using the wall-sized display of the WILD room (see appendix A—"WILD: Technical Equipment"). Our goal is to identify the performance characteristics of each technique from the user's perspective. It is thus essential that each technique operates equally well from a purely technological perspective. We therefore use the high-precision VICON motion-capture system to track passive IR retroreflective markers attached to the input devices or the user's body (although gesture recognition technologies are constantly improving, such a system is still necessary to get reliable and precise 3D position/orientation information).

The *Linear 1DPath* condition uses the wheel of a wireless Logitech M305 mouse (Fig. 6.4-a,b). The *Circular 1DPath* condition uses a wireless Samsung SM30P pointing device, normally used for presentations (Fig. 6.4-a,b). All *2DSurface* conditions use an iPod Touch. So as to avoid failures from gesture segmentation algorithms that would impact task performance in an uncontrolled manner, we use an explicit mode switch to unambiguously engage drag mode (panning). As mentioned earlier, we use the device's main button for *1DPath* conditions, and the lower area of the touch-sensitive surface for *2DSurface* conditions. While in real-world applications we would use specific hand postures such as pinching in *3DFree* conditions, for the sake of robustness we use a wireless mouse button whose activation is seamlessly integrated with the gesture.

The experiment was written in Java 1.5 running on Mac OS X and was implemented with the open source jBricks framework [Pie+11] for display walls. Touchstone [Mac+07] was used to design and manage the experiment.

Pan-Zoom Task The task is a variation of Guiard et al.'s multiscale pointing task [GB04], adapted to take overshoots into account. Participants navigate through an abstract information space made of two groups of concentric circles: the *start group* and the *target group*. Each group consists of seven series of 10 concentric circles symbolizing different zoom levels, each designated by a different color (Fig. 6.6.2). The target group features two additional green circles (dashed in Fig. 6.6.4) and a disc, referred to as C_1, C_2 and C_3 from smallest to largest.

Participants start at a high zoom level in the start group (Fig. 6.6.1). They zoom out until the neighboring target group appears (Fig. 6.6.2). It may appear either on the left or right side of the start group. Then they pan and zoom into the target group until they reach the correct zoom level and the target is correctly centered. A stationary gray ring symbolizes the correct zoom level and position (Fig. 6.6-(1-4)). Its radii are $r_1 = 4400$ and $r_2 = 12480$ pixels. All three criteria must be met for the trial to end: $\mathcal{A}$) C_1 is fully contained within the stationary ring's hole (radius = r_1), $\mathcal{B}$) $radius(C_2) < r_2$, $\mathcal{C}$) $radius(C_3) > r_2$. Overshoots occur when the zoom level is higher than the maximum level required to meet criteria $\mathcal{B}$ and $\mathcal{C}$, in which case participants have to zoom out again (C_1 becomes white instead of green in that situation).

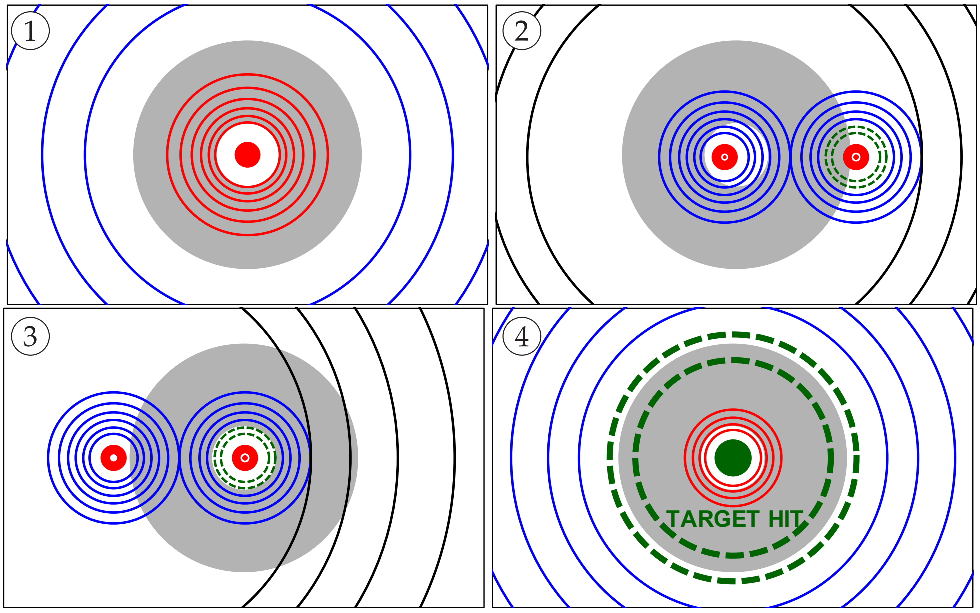


Figure 6.6: Task (schematic representation using altered colors): (1) Groups of concentric circles represent a given position and zoom level. (2) Zooming out until the neighboring set of circles appears. (3-4) Pan and zoom until the target (green inner disc and circles, dashed for illustration purposes only) is positioned correctly with respect to the stationary gray ring.

When all conditions are met, the message TARGET HIT appears and the thickness of C_2 and C_3 is increased (Fig. 6.6.4). The trial ends when the position and zoom level have stabilized for at least 1.2 seconds (all trials must be successfully completed).

Procedure The experiment presents each subject with six replications of each of the 12 techniques at three DISTANCES. The experiment is organized into four sessions that each present three techniques: One combination of the GESTURE and HANDEDNESS factors and all three degrees of GUIDANCE. Each session lasts between 30 and 90 minutes, depending on techniques and participant. Participants are required to wait at least one hour between two consecutive sessions, and to complete the whole experiment within four days or fewer, with a maximum of two sessions per day to avoid too much fatigue and boredom. Participants stand 1.7m from the wall and are asked to find a comfortable position so they can perform gestures quickly, but in a relaxed way.

Practice Condition: Participants are given a brief introduction at the beginning of the first session. Each technique begins with a practice condition, with trials at three different DISTANCES: (49 920, 798 720 and 12 779 520 pixels). Measures for the experimental condition start as soon as 1) participants feel comfortable and 2) task performance time variation for the last four trials is less than 30% of the task time average in that window.

Experimental Condition: Each technique is presented in a block of 18 trials consisting of 6 replications at each DISTANCE. Trials, blocks and sessions are fully counter-balanced within and across subjects, using a Latin square design.

Measures: We measure movement time *MT* and number of overshoots for each of 2592 trials: 2 GESTURE × 2 HANDEDNESS × 3 GUIDANCE × 3 DISTANCE × 12 participants × 6 replications. Participants also answer questions, based on a 5-point Likert scale, about their perceived *performance*, *accuracy*, *ease of learning*, *ease of use*, and *fatigue*. They rank the techniques with respect to the GUIDANCE factor after each session. When they have been exposed to both conditions of HANDEDNESS or GESTURE, they rank those as well. After the last session, they rank the techniques individually and by factor. Participants are encouraged to make additional observations and comments about any of the above.

6.4 Results and Discussion

6.4.1 Movement Time

Prior to our analysis, we checked the performance for unwanted effects from secondary factors. We checked for individual performance differences across subjects and found that, for all 12 participants, movement time and number of overshoots were perfectly correlated with the overall performance measures. As expected, movement time data are skewed positively; replications of unique experimental conditions are thus handled by taking the median (note that taking the mean yields similar results). In all remaining analysis, we handled participant as a random variable, using the standard repeated measures REML technique. We found no significant fatigue effect although we did find a significant learning effect across sessions. Participants performed about $1.4\,s$ more slowly in the first session and then became slightly faster over the next three sessions. However, we found no significant interaction between session orders and main factors. As the factors were counter-balanced, this created no adverse effects in the analysis.

Factor	DF, DFDen	F Ratio	p
Hands	1,11	24.65	0.0004 *
Gesture	1,11	42.87	$<$ 0.0001 *
Guidance	2,22	58.80	$<$ 0.0001 *
Dist	2,22	228.8	$<$ 0.0001 *
Hands×Gesture	1,11	2.060	0.1790
Hands×Guidance	2,22	4.914	0.0172 *
Gesture×Guidance	2,22	10.38	0.0007 *
Gesture×Dist	2,22	17.27	$<$ 0.0001 *
Hands×Dist	2,22	11.57	0.0004 *
Guidance×Dist	4,44	3.828	0.0094 *
Hands×Gesture×Guidance	2,22	1.127	0.3420
Hands×Gesture×Dist	2,22	0.790	0.4661
Hands×Guidance×Dist	4,44	0.650	0.6301
Gesture×Guidance×Dist	4,44	3.750	0.0104 *
Hands×Gesture×Guidance×Dist	4,44	1.049	0.3929

Table 6.2: Results of the full factorial ANOVA for MT.

Table 6.2 details results of the full factorial ANOVA for the model *MT* ∼ Hands × Guidance × Gesture × Dist × Rand(Participant). We observe that Hands has a significant effect on *MT* (Figure 6.7-a[2]). A post-hoc Tukey test shows that *TwoHanded* gestures are significantly faster than *OneHanded* gestures (avg. $9690\,ms$ vs. $11869\,ms$). We found a significant interaction effect of Hands × Guidance (Figure 6.7-a). The interaction does not change the significance of the post-hoc test, but indicates that the magnitude of the difference is greater for *3DFree* than for *2DSurface* and greater for *2DSurface* than for *1DPath* techniques.

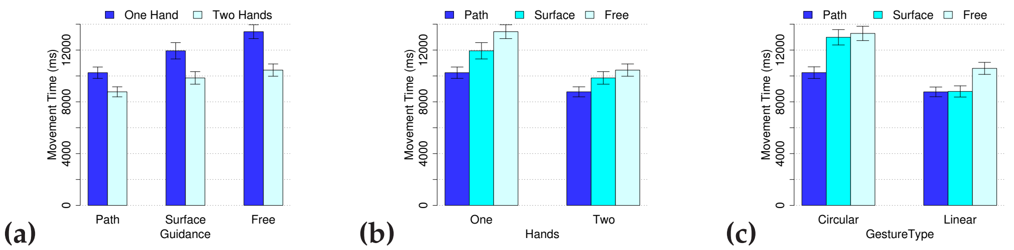


Figure 6.7: (a): *MT* per Hands × Guidance. (b) *MT* per Guidance × Hands. (c) *MT* per Guidance × Gesture.

Unsurprisingly, performance data strongly support (*H1*): all other conditions being equal, two-handed techniques are consistently faster than one-handed techniques. An interesting observation is that using two hands is more ad-

[2]Error bars in all the figures represent the 95% confidence limit of the mean of the medians per participants ($\pm StdErr \times 1.96$).

vantageous when the degree of guidance for achieving gestures is low.

GUIDANCE has a significant effect on *MT* (Figure 6.7-b). A post-hoc Tukey test shows that *1DPath* (avg. $9511\,ms$) is significantly faster than *2DSurface* ($10894\,ms$), which in turn is significantly faster than *3DFree* ($11934\,ms$). This time the HANDS × GUIDANCE interaction changes the significance of the test (Figure 6.7-b). The difference is that a post-hoc Tukey test shows no significant difference between *2DSurface* and *3DFree* for *TwoHanded*.

Both hypotheses (*H5*) and (*H6*) are supported: involving smaller muscle groups improves performance; providing higher guidance further contributes to this. However, this effect is less pronounced in *TwoHanded* conditions. This confirms the previous observation that a higher degree of guidance is especially useful when a single hand is involved.

GESTURE also has a significant effect on *MT*. A post-hoc Tukey test shows that *Linear* movements (avg. $9384\,ms$) performed significantly faster than *Circular* gestures ($12175\,ms$). However, we have a strong significant interaction of GESTURE × GUIDANCE (Figure 6.7-c). A post-hoc Tukey test shows that (i) for *Circular* gestures: *1DPath* guidance is faster than both *2DSurface* and *3DFree* with no significant difference between *2DSurface* and *3DFree*; (ii) for *Linear* gestures, there is no significant difference between *1DPath* and *2DSurface*, but a significant difference between *2DSurface* and *3DFree*; (iii) for *1DPath* guidance there is no significant difference between *Circular* and *Linear* gestures, but there is a significant difference between *Circular* and *Linear* for *2DSurface* and *3DFree* guidance.

Surprisingly, *Linear* gestures are generally faster than *Circular* ones. (*H3*), that claimed that *Linear* gestures should be slower because of clutching, is not supported. Performance differences between gesture types are however affected by the degree of guidance: *Circular* gestures with *1DPath* guidance (e.g., a knob) are comparable to *Linear* gestures with low guidance. We tentatively explain the lower performance of *Circular* gestures with *2DSurface* guidance by the difficulty of performing circular gestures with the thumb [RLG09], also observed here.

Another interesting observation is that our analogue of CycloStar in mid-air (*Circular* gestures with *3DFree* guidance) performs poorly. It seems that the lack of a surface to guide the gesture significantly degrades this technique's usability. Another factor contributing to its poor performance in our study is likely related to overshoots, as discussed below.

As expected, distance to target (DIST) has a significant effect on *MT*. A post-hoc Tukey test shows that *MT* increases significantly with *distance*. There are several significant interactions between DIST and the main factors (Fig. 6.8), but none of these change the relative performance ordering for the main factors. These interactions are due to a change in the magnitude of the difference across conditions, confirming that the choice of an efficient technique is of increasing importance as the task becomes harder.

6.4.2 Overshoots

As detailed earlier in the description of task design, overshoots correspond to zooming beyond the target zoom level and are treated as errors. We consider the model *Overshoots* $\sim$ HANDS $\times$ GUIDANCE $\times$ GESTURE $\times$ DIST $\times$ Rand(Participant).

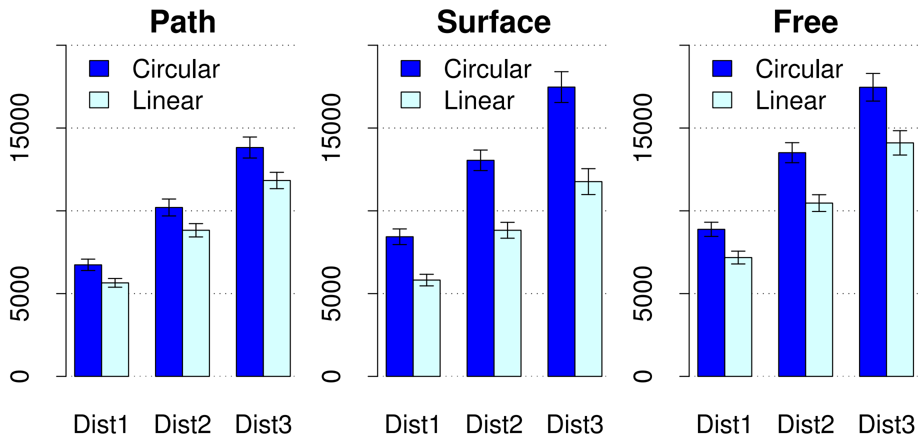


Figure 6.8: *MT* per DIST $\times$ GESTURE, for each GUIDANCE

We observe significant simple effects on *Overshoots* for GESTURE ($F_{1,11} = 21.04$, $p = 0.0008$) and GUIDANCE ($F_{2,22} = 53.80$, $p < 0.0001$), and one significant interaction effect for GESTURE $\times$ GUIDANCE ($F_{2,22} = 8.63$, $p = 0.0017$). *Circular* gestures exhibit more overshoots than *Linear* gestures (1.65 vs. 2.71). *2DSurface* gestures exhibit more overshoots than *1DPath* and *3DFree* gestures (3.75 for *2DSurface* vs. 1.52 for *1DPath*, and 1.26 for *3DFree*). There is a significant difference between *Linear* and *Circular* gestures for *2DSurface* and *3DFree*, but not *1DPath*. Moreover, overshoots exhibit the same interaction effect for *2DSurface* gestures: *Circular 2DSurface* result in significantly more overshoots than *Linear 2DSurface* (4.68 vs. 2.82).

The observed higher number of overshoots for *Circular* techniques helps explain the generally lower *MT* performance measured for this type of gestures.

The best-fitting ellipse algorithm involved in the recognition of *Circular* gestures has an inherently higher cost of recovery, introducing a delay when reversing course. The poor performance of our analogue of CycloStar is at least partially due to this, knowing that there was a major difference between the zooming experiment reported in [MLG10] and the present one: we included overshoots in our task design, whereas the CycloStar experiment apparently did not (there is no report of such a measure in task design or results analysis), thus ignoring this issue.

6.4.3 Qualitative Results

Qualitative data confirms our results. Participants generally preferred *TwoHanded* to *OneHanded* techniques (8/12) and *Linear* to *Circular* gestures (10/12). Subjective preferences about degree of guidance were mixed, with 4 participants preferring the high degree of guidance provided by *1DPath* techniques, only 1 for both of *2DSurface* and *3DFree* techniques, and all others expressing no particular preferences. Looking at the details of answers to our 5-point Likert scale questions about perceived speed, accuracy, ease of use and fatigue, significant results ($p < 0.002$) were obtained only for degree of GUIDANCE, with *1DPath* being consistently rated higher than *2DSurface* and *3DFree*; and for HANDS, *TwoHanded* techniques being considered less tiring than *OneHanded* techniques ($p < 0.03$).

Comments from participants suggest that in the *OneHanded* condition, zoom gestures interfere with pointing as they introduce additional hand jitter and consequently lower accuracy. Some participants also commented that pointing and zooming were confounded in the *OneHanded* conditions, making the techniques difficult to use (*H2*). However, two participants strongly preferred one-handed gestures, arguing that they were less complex and less tiring. They assumed their performance was better (even though it was not), probably because they experienced more overshoots in the *two handed* condition, which may have led to their conclusions. One of them mentioned that for the *one handed* condition there was *"no need for coordination"*; techniques were *"more relaxed"* and made it *"easier to pan and zoom"*.

Linear gestures were preferred to *Circular* ones, participants commenting that circular gestures were difficult to perform without guidance, that circular gestures for zooming interfered with linear gestures for panning, and that circular gestures were hard to map to zoom factor. All but one participants preferred linear gestures overall although one commented that he liked *"the continuity of circular gestures"*. Others commented that *"making good circles*

GROUP	HANDS	GESTURE	GUIDANCE	Figure	*MT* (ms)
Gr1	*TwoHanded*	*Linear*	*2DSurface*	6.4-d	8 100
	TwoHanded	*Linear*	*1DPath*	6.4-b	8 377
Gr2	*OneHanded*	*Linear*	*1DPath*	6.4-a	9 160
	TwoHanded	*Circular*	*1DPath*	6.4-b	9 168
	TwoHanded	*Linear*	*3DFree*	6.4-f	9 185
	OneHanded	*Linear*	*2DSurface*	6.4-c	9 504
Gr3	*OneHanded*	*Circular*	*1DPath*	6.4-a	11 340
	TwoHanded	*Circular*	*2DSurface*	6.4-d	11 591
	TwoHanded	*Circular*	*3DFree*	6.4-f	11 718
	OneHanded	*Linear*	*3DFree*	6.4-e	11 981
Gr4	*OneHanded*	*Circular*	*2DSurface*	6.4-c	14 380
	OneHanded	*Circular*	*3DFree*	6.4-e	14 851

Table 6.3: Groups of techniques according to *MT*

without a guide is hard" and did not like having to turn their hands. These findings contradict our hypothesis that users would prefer clutch-free circular gestures (*H4*). This hypothesis was based on observations made for techniques operated on a desktop, not in mid-air, and involved different limb segments. In many of our conditions, the gestures had to be performed with the thumb, and were thus more complex to achieve than when using, e.g., the index finger in conjunction with hand or forearm movements. Several participants commented on this interaction effect: *"[It is] too hard to do circle gestures without a guide"*, *"Linear movements are easier on the iPod"* and *"[Is it] impossible to do circular movements on a surface, maybe with some oil?"*.

Finally, as hypothesized (*H7*), participants found *1DPath* guidance least tiring while *3DFree* caused the most fatigue.

6.4.4 Individual Techniques

The analysis of variance for the model *MT* $\sim$ HANDS $\times$ GUIDANCE $\times$ GESTURE $\times$ DIST $\times$ Rand(Participant) does not show a significant triple interaction between the three main factors (Table 6.2). Formally, we cannot say more than the above about the ranking of the twelve techniques. However, based on the results about *MT* above, we can observe four distinct groups of techniques, shown in Table 6.3. As a side note, if we consider the model *MT* $\sim$ GROUP $\times$ Rand(Participant), the ANOVA shows a significant effect of GROUP ($F_{3,33} = 65.35$, $p < 0.0001$) and a post-hoc Tukey test shows a significant difference between each groups.

Gr1 contains the two fastest techniques with similar *MT* : *TwoHanded*, *Linear* gestures with either *2DSurface* or *1DPath* degrees of guidance. Optimal performance in terms of movement time implies the use of two hands and a device to guide gestural input.

Gr2 contains the four techniques that come next and also have close *MT* : the *OneHanded* version of the two fastest techniques, the *TwoHanded Circular 1DPath* and the *TwoHanded Linear 3DFree* techniques. Techniques in this group are of interest as they exhibit a relatively good level of performance while broadening possible choices for interaction designers. For instance, the uni-manual techniques in this group make one hand available to perform other actions. The *3DFree* technique is also of interest as it does not require the user to hold any equipment and is generally appealing to users.

Gr3 contains techniques that again have very close *MT* but about $2.3\,s$ slower than the techniques of *Gr2*. This group consists of *OneHanded Circular 1DPath*, *TwoHanded Circular 2DSurface* and *3DFree*, and *OneHanded Linear 3DFree*. Techniques in this group are of lesser interest, except maybe for the *OneHanded Linear 3DFree* technique, which is the fastest uni-manual technique using gestures performed in free space.

Gr4 contains the 2 techniques performing worst, *OneHanded Circular 2DSurface* and *3DFree*. These are significantly slower than all others, about $3\,s$ slower than the techniques of *Gr3* and about $6\,s$ slower than the techniques of *Gr1*. Our data suggest that these techniques should be rejected.

6.5 Discussion and Conclusion

Figure 6.1 illustrates the alternatives we explored in the BodyScape formalism. Recall that, IMA_1 involves the dominant arm in all techniques. Interaction techniques vary (1) in the types of composition between IMA_1 and IMA_2, separated (bi-manual) or overlapping (uni-manual) and (2) in the size of IMA_2, hand or entire arm. Figure 6.9 shows an overview of the results with respect to the presented performance groups presented in table 6.3.

Input motor assembly composition Composing two separated input motor assemblies (bimanual) performs faster than composing two overlapping input motor assemblies (uni-manual), since overlapping motor assemblies lead to interaction effects between two concurrent produced movements.

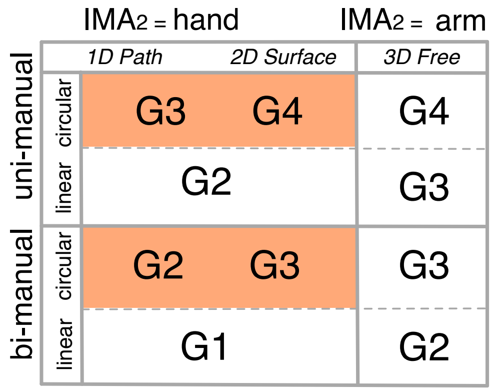


Figure 6.9: Results according to the significantly different groups presented in table 6.3 with respect to input motor assembly composition, separated (bi-manual) or overlapping (uni-manual) and the size of input motor assembly IMA_2.

Size of IMA_2 For linear gestures, we found that input motor assemblies that involve a smaller number of motors results in increased performance. For circular gestures, we found that a smaller body involvement into the interaction does not necessarily result in increased performance. However, guidance of input motor assemblies (*1D path* vs. *2D surface*) can significantly increase performance.

6.6 Contributions

1. Implementation of twelve pan-and-zoom navigation techniques in collaboration with Mathieu Nancel.
2. Two separated input motor assemblies perform faster than two overlapping input motor assemblies.
3. A smaller number of motors involved in the input motor assembly results in increased performance for simple linear gestures.
4. For more complex gestures, the number of involved motors does not change the interaction performance.
5. Complex gestures benefit from the guidance of motion.

7 Conclusion and Future Work

Multi-surface environments are complex environments where multiple people interact with multiple interactive surface devices. Interaction and the graphical interface are distributed across devices. Previous work has explored several point designs to explore the technological aspects of multi-surface environments. I explored interaction using a body-centric approach. From a body-centric perspective technological implementation of a technique does not matter. The body-centric analysis enables to reason why some interaction techniques are performed faster than others and why some techniques are more preferred.

BodyScape is a body-centric framework that takes into account how the body is involved into the interaction and how input movements are performed with respect to the interactive environment. BodyScape introduces a notation that can *describe* interaction techniques in terms of (1) motor assemblies responsible for performing a control task (input motor assembly) or bringing the body into a position to visually perceive output (output motor assembly), and (2) the movement coordination of motor assemblies, body-relative or world-fixed, with respect to the interactive environment.

Through three practical examples, I demonstrated that BodyScape can be used to *describe*, *generate*, and *compare* interaction techniques:

Describing interaction techniques The body-centric approach allowed me to describe differences in performance and perceived comfort of the BiPad interaction techniques.

Generating interaction techniques BodyScape allowed me to create a novel compound interaction technique by combining two existing atomic interaction techniques.

Comparing interaction techniques Using BodyScape, I compared twelve interaction techniques for pan-and-zoom navigation on large displays.

A body-centric perspective can help to analyze the effect of the spatial body-device relationship on input performance: when interaction and device support are shared across the input motor assembly's motors, the support affects input performance. Holding a device limits the motion of motors within the input motor assembly: the more motors are involved into the support, the more comfortable the users perceive the interaction; when less motors are involved into the support, interacting limbs are less restricted in movement, resulting in increased performance (see chapter 4).

BodyScape can predict interaction effects that can occur between concurrent input movements: Two input motor assemblies that are seemingly independent can interfere with each other if one input motor assembly affects body parts contained in the second input motor assembly (see chapter 5). When input motor assemblies share some body limbs (overlapping), movements of both input motor assemblies can interfere with each other, resulting in a drop of performance (see chapter 6).

Promising future research can be inspired by Billinghurst et al.'s work with wearable computers that provide spatialized 3D graphics and audio cues to aid communication with remote collaborators that virtually surround the user [Bil+98]. They introduce visual output which remains at a constant position relative to a specific body part, e.g. the head. Future research can further explore the body-centric aspects of body-relative input and output motor assemblies when they coordinate their movements with respect to a specific body part.

While the studies that I have conducted have demonstrated that most of the BodyScape insights are relevant, the framework does not yet permit an a priori valuation of the performances of a body-centric technique. This has to be investigated in future work, and I believe that the studies that I have presented to validate my body-centric approach on practical examples could be an inspiring starting point for more general experiments focusing on performance prediction with BodyScape. In this thesis, as the first step of a body-centric approach for analyzing interaction techniques, I focused only on the study of input motor assemblies and their combination. To complete the approach, future work is required to investigate the combination of output and input motor assemblies and to analyze possible interaction effects.

7.1 Interaction techniques from a new perspective

The performance of an interaction technique depends on several aspects: physical, cognitive and perceptual actions [CNM00], context of use within an interaction sequence [ALM04]. In my thesis, I have also shown that the spatial relationship between the user's body and interaction devices matters. I argue that a body-centric approach can change our point of view on interaction techniques, beyond multi-surface environments. For instance, interaction techniques such as small circular movements with the thumb on a mobile phone [RLG09] depend upon the spatial relationship between the thumb and the interactive surface. A technology oriented approach will consider this gesture to be similar when it is performed on any other tactile surface, thus presupposing that performance remains similar. However, from a body-centric point of view, the same gesture performed on a horizontal surface will be a different interaction technique, implying different body involvement and restrictions, and consequently potential differences in performances and comfort of use. This is a new perspective on the description, design and evaluation of interaction techniques. It augments existing approaches with the study of body-device relationships and should be investigated deeper in order to be generalized beyond multi-surface environments.

7.2 Exploring compositions of input and output motor assemblies

In my thesis, I primarily studied input motor assemblies in the BodyScape framework. For future work, I am interested in studying output motor assemblies, and in particular the effects of divided visual attention on interaction performance. For instance, figure 7.1 illustrates a user applying a tool from a palette on a mobile phone to a target on a large display. She performs both mid-air pointing and direct-touch at the same time, which requires her to shift her head back and forth to perceive feedback, thus involving output motor assembly differently from one technique to another.

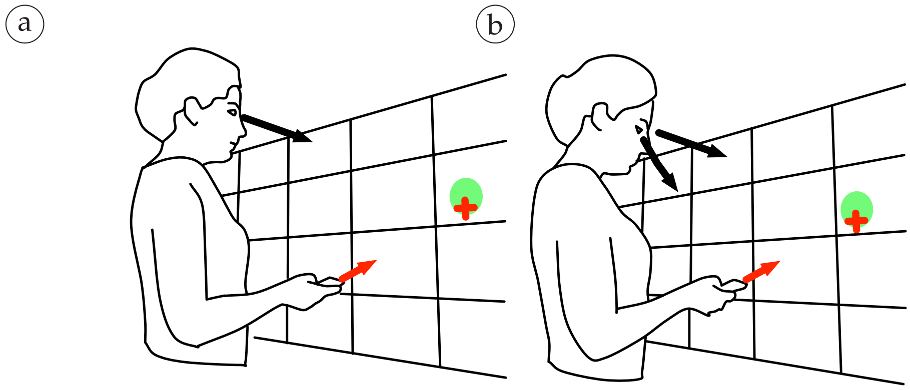


Figure 7.1: An interaction sequence with divided visual attention: a) acquire virtual object on the wall, b) remain with cursor inside the object and apply tool to it displayed on the hand-held device.

I hypothesize that such techniques, that require the output motor assembly to move back and forth between several locations, can be improved by designing relevant interaction sequences where the output motor assembly remains focused on an unique display for each interaction step. On a more general point of view, in the BodyScape framework, it consists in studying the relationships between combinations of input motor assemblies and output motor assemblies, wether they are fixed in the world or relative to the body.

In summary, BodyScape characterizes interaction techniques from a body-centric perspective. It clarifies the role of device support on the user's balance and subsequent comfort and performance. Using BodyScape, designers can identify situations in which multiple body movements interfere with each other, with a corresponding decrease in performance. Finally, it highlights the trade-offs among different combinations of techniques, enabling the analysis and generation of a variety of multi-surface interaction techniques. I argue that including a body-centric perspective when defining interaction techniques is essential for addressing the combinatorial explosion of interactive devices in new interactive environments, i.e., multi-surface and beyond.

Publications

Refereed International Journals and Magazines

[Bea+12] Michel Beaudouin-Lafon, Olivier Chapuis, James Eagan, Tony Gjerlufsen, Stéphane Huot, Clemens Klokmose, Wendy Mackay, Mathieu Nancel, Emmanuel Pietriga, Clément Pillias, Romain Primet, and **Julie Wagner**. "Multi-surface Interaction in the WILD Room". In: *Computer* 45.4 (2012). Pp. 48–56. http://hal.inria.fr/hal-00687825/PDF/WILD-IEEEComputer-authorversion.pdf. (Cit. on pp. 14, 140).

Refereed International Conferences

[Wag+13] **Julie Wagner**, Mathieu Nancel, Sean Gustafson, Stéphane Huot, and Wendy E. Mackay. "A Body-centric Design Space for Multi-surface Interaction". In: *Proceedings of the 31th international conference on Human factors in computing systems*. CHI '13. New York, NY, USA: ACM, 2013, to appear. (Cit. on p. 147).

[WHM12] **Julie Wagner**, Stéphane Huot, and Wendy E. Mackay. "BiTouch and BiPad: Designing Bimanual Interaction for Hand-held Tablets". In: *Proceedings of the 30th international conference on Human factors in computing systems*. CHI '12. Austin, Texas, USA: ACM, 2012, pp. 2317–2326. http://hal.inria.fr/hal-00663972/PDF/bipadA.pdf. (Cit. on pp. 44, 147).

[Nan+11] Mathieu Nancel, **Julie Wagner**, Emmanuel Pietriga, Olivier Chapuis, and Wendy Mackay. "Mid-air pan-and-zoom on wall-sized displays". In: *Proceedings of the 2011 annual conference on Human factors in computing systems*. CHI '11. Vancouver, BC, Canada: ACM, 2011, pp. 177–186. *Best Paper Award*. http://hal.

archives-ouvertes.fr/docs/00/59/12/94/PDF/CHI11-pzwild-avf.pdf. (Cit. on p. 147).

[WM10] **Julie Wagner** and Wendy Mackay. “Exploring Sustainable Design with Reusable Paper”. In: *CHI ’10: Proceedings of the SIGCHI conference on Human factors in computing systems*. Atlanta, Georgia, USA: ACM, 2010, pp. 1871–1874. http://doi.acm.org/10.1145/1753326.1753606.

[Wei+09b] Malte Weiss, **Julie Wagner**, Roger Jennings, Yvonne Jansen, Ramsin Khoshabeh, James D. Hollan, and Jan Borchers. “SLAPbook: tangible widgets on multi-touch tables in groupware environments”. In: *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction*. TEI ’09. Cambridge, United Kingdom: ACM, 2009, pp. 297–300. http://doi.acm.org/10.1145/1517664.1517725.

[Wei+08] Malte Weiss, Roger Jennings, **Julie Wagner**, James D. Hollan, Ramsin Khoshabeh, and Jan Borchers. “SLAP: Silicone Illuminated Active Peripherals”. In: *Extended Abstracts of Tabletop ’08*. 2008, pp. 37–38. http://hci.rwth-aachen.de/materials/publications/weiss2008a.pdf.

Refereed Domestic Conferences

[WMH12] **Julie Wagner**, Wendy E. Mackay, and Stéphane Huot. “Leftover Windows Cause Window Clutter... But What Causes Leftover Windows?” In: *Ergo IHM 2012: Proceedings of the 24th French Speaking Conference on Human-Computer Interaction*. Biarritz, France: ACM, 2012. 4 pages, to appear.

Demonstrations

[Wei+09a] Malte Weiss, Roger Jennings, Ramsin Khoshabeh, Jan Borchers, **Julie Wagner**, Yvonne Jansen, and James D. Hollan. “SLAP widgets: bridging the gap between virtual and physical controls on tabletops”. In: *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*. CHI EA ’09. Boston, MA, USA: ACM, 2009, pp. 3229–3234.

Bibliography

[AEN10] Christopher Andrews, Alex Endert, and Chris North. “Space to think: large high-resolution displays for sensemaking”. In: *Proceedings of the 28th international conference on Human factors in computing systems*. CHI ’10. Atlanta, Georgia, USA: ACM, 2010, pp. 55–64. (Cit. on p. 91).

[Ale+12] Jason Alexander, Teng Han, William Judd, Pourang Irani, and Sriram Subramanian. “Putting your best foot forward: investigating real-world mappings for foot-based gestures”. In: *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*. CHI ’12. Austin, Texas, USA: ACM, 2012, pp. 1229–1238. (Cit. on pp. 23, 35).

[ALM04] Caroline Appert, Michel B. Lafon, and Wendy E. Mackay. “Context matters: Evaluating Interaction Techniques with the CIS Model”. In: *Proceedings of the The 18th British HCI Group Annual Conference*. 2004. (Cit. on pp. 17, 111).

[Ang+03] J. Ängeslevä, I. Oakley, S. Hughes, and S. O’Modhrain. “Body mnemonics: portable device interaction design concept”. In: *Adjunct Proc. UIST*. 2003. (Cit. on p. 71).

[Aug+10] Thomas Augsten, Konstantin Kaefer, René Meusel, Caroline Fetzer, Dorian Kanitz, Thomas Stoff, Torsten Becker, Christian Holz, and Patrick Baudisch. “Multitoe: high-precision interaction with back-projected floors based on high-resolution multi-touch input”. In: *Proc. UIST*. 2010, pp. 209–218. (Cit. on pp. 31, 33, 35).

[Bai+12] 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 2012 ACM annual conference on Human Factors in Computing Systems*. CHI ’12. Austin, Texas, USA: ACM, 2012, pp. 1239–1248. (Cit. on p. 23).

[BB93] Thomas Baudel and Michel Beaudouin-Lafon. “Charade: remote control of objects using free-hand gestures”. In: *Commun. ACM* 36.7 (July 1993). Pp. 28–35. (Cit. on p. 12).

[BG02] Frédréric Bourgeois and Yves Guiard. "Multiscale pointing: facilitating pan-zoom coordination". In: *CHI '02 EA*. ACM, 2002, pp. 758–759. (Cit. on pp. 92, 95, 98).

[BH99] Ravin Balakrishnan and Ken Hinckley. "The role of kinesthetic reference frames in two-handed input performance". In: *UIST '99*. Asheville, North Carolina, United States: ACM, 1999, pp. 171–178. (Cit. on p. 45).

[Bie+93] Eric A. Bier, Maureen C. Stone, Ken Pier, William Buxton, and Tony D. DeRose. "Toolglass and magic lenses: the see-through interface". In: *SIGGRAPH '93*. Anaheim, CA: ACM, 1993, pp. 73–80. (Cit. on pp. 16, 45).

[Bil+98] M. Billinghurst, J. Bowskill, M. Jessop, and J. Morphett. "A Wearable Spatial Conferencing Space". In: *Proceedings of the 2nd IEEE International Symposium on Wearable Computers*. ISWC '98. Washington, DC, USA: IEEE Computer Society, 1998, pp. 76–. (Cit. on pp. 27, 110).

[BM86] W. Buxton and B. Myers. "A study in two-handed input". In: *SIGCHI Bull.* 17.4 (1986). Pp. 321–326. (Cit. on pp. 91, 92, 95, 98).

[BM97] Ravin Balakrishnan and I. Scott MacKenzie. "Performance differences in the fingers, wrist, and forearm in computer input control". In: *Proc. CHI '97*. ACM, 1997, pp. 303–310. (Cit. on pp. 91, 98).

[BNB07] Robert Ball, Chris North, and Doug A. Bowman. "Move to improve: promoting physical navigation to increase user performance with large displays". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. CHI '07. San Jose, California, USA: ACM, 2007, pp. 191–200. (Cit. on pp. 71, 91).

[Bor+10] Sebastian Boring, Dominikus Baur, Andreas Butz, Sean Gustafson, and Patrick Baudisch. "Touch projector: mobile interaction through video". In: *Proceedings of the 28th international conference on Human factors in computing systems*. CHI '10. Atlanta, Georgia, USA: ACM, 2010, pp. 2287–2296. (Cit. on pp. 16, 36).

[Bra+08] Peter Brandl, Clifton Forlines, Daniel Wigdor, Michael Haller, and Chia Shen. "Combining and measuring the benefits of bimanual pen and direct-touch interaction on horizontal interfaces". In: *AVI '08*. Napoli, Italy: ACM, 2008, pp. 154–161. (Cit. on p. 45).

[Bro+97] Rodney A. Brooks, Darren Dang, Jeremy De Bonet, Joshua Kramer, Tomas Lozano-perez, John Mellor, and Polly Pook. "The Intelligent Room Project". In: *In Proceedings of the Second International Cognitive Technology Conference (CT'97), Aizu*. 1997. (Cit. on p. 2).

[Bry92] Wm. L. Bryan. "On the Development of Voluntary Motor Ability". English. In: *The American Journal of Psychology* 5.2 (1892). pages. (Cit. on p. 3).

[BSW06] Patrick Baudisch, Mike Sinclair, and Andrew Wilson. "Soap: a pointing device that works in mid-air". In: *Proc. UIST '06*. ACM, 2006, pp. 43–46. (Cit. on p. 93).

[BWB06] Hrvoje Benko, Andrew D. Wilson, and Patrick Baudisch. "Precise selection techniques for multi-touch screens". In: *CHI '06*. Montréal, Québec, Canada: ACM, 2006, pp. 1263–1272. (Cit. on p. 45).

[CB03] Xiang Cao and Ravin Balakrishnan. "VisionWand: interaction techniques for large displays using a passive wand tracked in 3D". In: *Proc. UIST '03*. ACM, 2003, pp. 173–182. (Cit. on p. 93).

[CMR91] Stuart K. Card, Jock D. Mackinlay, and George G. Robertson. "A morphological analysis of the design space of input devices". In: *ACM Trans. Inf. Syst.* 9.2 (Apr. 1991). Pp. 99–122. (Cit. on pp. 3, 23).

[CNM00] Stuart K. Card, Allen Newell, and Thomas P. Moran. *The Psychology of Human-Computer Interaction*. Hillsdale, NJ, USA: L. Erlbaum Associates Inc., 2000. (Cit. on p. 111).

[Col+05] Maxime Collomb, Mountaz Hascoët, Patrick Baudisch, and Brian Lee. "Improving drag-and-drop on wall-size displays". In: *Proceedings of Graphics Interface 2005*. GI '05. Victoria, British Columbia: Canadian Human-Computer Communications Society, 2005, pp. 25–32. (Cit. on p. 15).

[CR11] Rajal G. Cohen and David A. Rosenbaum. "Prospective and retrospective effects in human motor control: planning grasps for object rotation and translation". English. In: *Psychological Research* 75 (4 2011). Pp. 341–349. (Cit. on p. 26).

[Dez+12] Niloofar Dezfuli, Mohammadreza Khalilbeigi, Jochen Huber, Florian Müller, and Max Mühlhäuser. "PalmRC: imaginary palm-based remote control for eyes-free television interaction". In: *Proc. EuroiTV*. 2012, pp. 27–34. (Cit. on pp. 31, 32).

[DL04] R. Dickstein and Y. Laufer. "Light touch and center of mass stability during treadmill locomotion". In: *Gait & Posture* 20.1 (2004). Pp. 41–47. (Cit. on p. 29).

[DSK09] Florian Daiber, Johannes Schöning, and Antonio Krüger. “Whole Body Interaction with Geospatial Data”. In: *Proceedings of the 10th International Symposium on Smart Graphics*. SG ’09. Salamanca, Spain: Springer-Verlag, 2009, pp. 81–92. (Cit. on pp. 22, 23).

[ETW81] Kenneth Evans, Peter Tanner, and Marceli Wein. “Tablet-based valuators that provide one, two, or three degrees of freedom”. In: *Proc. SIGGRAPH ’81*. ACM, 1981, pp. 91–97. (Cit. on p. 93).

[Fei+93] Steven Feiner, Blair MacIntyre, Marcus Haupt, and Eliot Solomon. “Windows on the world: 2D windows for 3D augmented reality”. In: *Proceedings of the 6th annual ACM symposium on User interface software and technology*. UIST ’93. Atlanta, Georgia, United States: ACM, 1993, pp. 145–155. (Cit. on p. 27).

[FIB95] George W. Fitzmaurice, Hiroshi Ishii, and William A. S. Buxton. “Bricks: laying the foundations for graspable user interfaces”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. CHI ’95. Denver, Colorado, United States: ACM Press/Addison-Wesley Publishing Co., 1995, pp. 442–449. (Cit. on p. 2).

[For+07] Clifton Forlines, Daniel Wigdor, Chia Shen, and Ravin Balakrishnan. “Direct-touch vs. mouse input for tabletop displays”. In: *Proc. CHI*. San Jose, California, USA, 2007, pp. 647–656. (Cit. on p. 71).

[FWC90] James D. Foley, Victor L. Wallace, and Peggy Chan. “Human-computer interaction”. In: ed. by Jenny Preece. Upper Saddle River, NJ, USA: Prentice Hall Press, 1990. Chap. The human factors of computer graphics interaction techniques, pp. 67–121. (Cit. on p. 39).

[FWK07] Jon Froehlich, Jacob O. Wobbrock, and Shaun K. Kane. “Barrier pointing: using physical edges to assist target acquisition on mobile device touch screens”. In: *Assets ’07*. Assets ’07. Tempe, Arizona, USA: ACM, 2007, pp. 19–26. (Cit. on p. 45).

[GB04] Yves Guiard and Michel Beaudouin-Lafon. “Target acquisition in multiscale electronic worlds”. In: *IJHCS* 61.6 (2004). Pp. 875–905. (Cit. on pp. 98, 99).

[GB10] Susan Goldin-Meadow and Sian L. Beilock. “Action’s Influence on Thought: The Case of Gesture”. In: *Perspectives on Psychological Science* 5.6 (Nov. 1, 2010). Pp. 664–674. (Cit. on p. 12).

[GHB11] S. Gustafson, C. Holz, and P. Baudisch. “Imaginary Phone: learning imaginary interfaces by transferring spatial memory from a familiar dvice”. In: *Proc. UIST*. 2011, pp. 283–292. (Cit. on pp. 32, 71).

[GIB62] J. J. GIBSON. “Observations on active touch.” In: *Psychol Rev* 69 (Nov. 1962). Pp. 477–491. (Cit. on p. 29).

[Gib86] James J. Gibson. *The Ecological Approach To Visual Perception*. New edition. Lawrence Erlbaum Associates, Sept. 1, 1986. (Cit. on p. 26).

[Gui87] Yves Guiard. *Asymmetric Division of Labor in Human Skilled Bimanual Action: The Kinematic Chain as a Model*. 1987. (Cit. on pp. 24, 52, 53, 66, 70, 71, 91, 92, 95).

[Han05] Jefferson Y. Han. “Low-cost multi-touch sensing through frustrated total internal reflection”. In: *Proceedings of the 18th annual ACM symposium on User interface software and technology*. UIST ’05. Seattle, WA, USA: ACM, 2005, pp. 115–118. (Cit. on p. 140).

[Hin+98] Ken Hinckley, Randy Pausch, Dennis Proffitt, and Neal F. Kassell. “Two-handed virtual manipulation”. In: *ACM TOCHI* 5 (3 1998). Pp. 260–302. (Cit. on p. 45).

[HL07] Stéphane Huot and Eric Lecolinet. “Focus+context visualization techniques for displaying large lists with multiple points of interest on small tactile screens”. In: *INTERACT’07*. Rio de Janeiro, Brazil: Springer-Verlag, 2007, pp. 219–233. (Cit. on p. 45).

[HRH12] C. Harrison, S. Ramamurthy, and S Hudson. “On-body interaction: armed and dangerous”. In: *Proc. TEI*. 2012, pp. 69–76. (Cit. on pp. 22, 23, 31).

[HTM10] C. Harrison, D. Tan, and D. Morris. “Skinput: Appropriating the Body as an Input Surface”. In: *Proc. CHI*. 2010, pp. 453–462. (Cit. on pp. 22, 28, 31, 35, 87, 88).

[Iza+03] Shahram Izadi, Harry Brignull, Tom Rodden, Yvonne Rogers, and Mia Underwood. “Dynamo: a public interactive surface supporting the cooperative sharing and exchange of media”. In: *Proceedings of the 16th annual ACM symposium on User interface software and technology*. UIST ’03. Vancouver, Canada: ACM, 2003, pp. 159–168. (Cit. on p. 11).

[JFW02] Brad Johanson, Armando Fox, and Terry Winograd. “The Interactive Workspaces Project: Experiences with Ubiquitous Computing Rooms”. In: *IEEE Pervasive Computing* 1.2 (Apr. 2002). Pp. 67–74. (Cit. on p. 11).

[Joh+02] Brad Johanson, Greg Hutchins, Terry Winograd, and Maureen Stone. “PointRight: experience with flexible input redirection in interactive workspaces”. In: *Proceedings of the 15th annual ACM symposium on User interface software and technology*. UIST ’02. Paris, France: ACM, 2002, pp. 227–234. (Cit. on p. 11).

[JS92] Robert Jacob and Linda Sibert. "The perceptual structure of multidimensional input device selection". In: *Proc. CHI '92*. ACM, 1992, pp. 211–218. (Cit. on pp. 91, 92, 98).

[KAD09] Kenrick Kin, Maneesh Agrawala, and Tony DeRose. "Determining the benefits of direct-touch, bimanual, and multifinger input on a multitouch workstation". In: *GI '09*. Kelowna, British Columbia, Canada: Canadian Information Processing Society, 2009, pp. 119–124. (Cit. on p. 45).

[Kar+11] T. Karrer, M. Wittenhagen, L. Lichtschlag, F. Heller, and J. Borchers. "Pinstripe: eyes-free continuous input on interactive clothing". In: *Proc. CHI*. 2011, pp. 1313–1322. (Cit. on pp. 29, 72, 75, 81).

[KB06] Amy K. Karlson and Benjamin B. Bederson. *Understanding Single-Handed Mobile Device Interaction*. Tech. rep. 2006. (Cit. on p. 45).

[KB08] Amy K. Karlson and Benjamin B. Bederson. "One-handed touchscreen input for legacy applications". In: *CHI '08*. Florence, Italy: ACM, 2008, pp. 1399–1408. (Cit. on p. 45).

[KBS05] Amy K. Karlson, Benjamin B. Bederson, and John SanGiovanni. "AppLens and launchTile: two designs for one-handed thumb use on small devices". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '05. Portland, Oregon, USA: ACM, 2005, pp. 201–210. (Cit. on p. 45).

[KBS94] Paul Kabbash, William Buxton, and Abigail Sellen. "Two-handed input in a compound task". In: *CHI '94*. Boston, Massachusetts, United States: ACM, 1994, pp. 417–423. (Cit. on p. 45).

[KGH85] M. W. Krueger, T. Gionfriddo, and K. Hinrichsen. "VIDEOPLACE—an artificial reality". In: *Proc. CHI*. San Francisco, California, United States, 1985, pp. 35–40. (Cit. on p. 32).

[KHT06] Scott R. Klemmer, Björn Hartmann, and Leila Takayama. "How bodies matter: five themes for interaction design". In: *Proceedings of the 6th conference on Designing Interactive systems*. DIS '06. University Park, PA, USA: ACM, 2006, pp. 140–149. (Cit. on p. 71).

[Lat+10] C. Latulipe, D. Wilson, S. Huskey, M. Word, A. Carroll, E. Carroll, B. Gonzalez, V. Singh, M. Wirth, and D. Lottridge. "Exploring the design space in technology-augmented dance". In: *CHI Extended Abstracts*. 2010, pp. 2995–3000. (Cit. on pp. 36, 71).

[LDT09] F. Li, D. Dearman, and K. Truong. "Virtual shelves: interactions with orientation aware devices". In: *Proc. UIST*. 2009, pp. 125–128. (Cit. on p. 31).

[LDT10] Frank Chun Yat Li, David Dearman, and Khai N. Truong. "Leveraging proprioception to make mobile phones more accessible to users with visual impairments". In: *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility*. ASSETS '10. Orlando, Florida, USA: ACM, 2010, pp. 187–194. (Cit. on p. 71).

[Lin+11] S. Lin, Z. Su, K. Cheng, R. Liang, T. Kuo, and B. Chen. "PUB - Point Upon Body: exploring eyes-free interactions and methods on an arm". In: *Proc. UIST*. 2011, pp. 481–488. (Cit. on pp. 22, 29, 35, 71, 72, 87, 88).

[LRE06] Astrid Twenebowa Larssen, Toni Robertson, and Jenny Edwards. "How it feels, not just how it looks: when bodies interact with technology". In: *Proceedings of the 18th Australia conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*. OZCHI '06. Sydney, Australia: ACM, 2006, pp. 329–332. (Cit. on p. 26).

[LSH99] Robert Lindeman, John Sibert, and James Hahn. "Hand-Held Windows: Towards Effective 2D Interaction in Immersive Virtual Environments". In: *Proc. VR '99*. IEEE, 1999, pp. 205–212. (Cit. on p. 93).

[LZB98] Andrea Leganchuk, Shumin Zhai, and William Buxton. "Manual and cognitive benefits of two-handed input: an experimental study". In: *ACM TOCHI* 5 (4 1998). Pp. 326–359. (Cit. on pp. 45, 91, 92, 98).

[Mac+07] Wendy E. Mackay, Caroline Appert, Michel Beaudouin-Lafon, Olivier Chapuis, Yangzhou Du, Jean-Daniel Fekete, and Yves Guiard. "Touchstone: exploratory design of experiments". In: *Proc. CHI '07*. ACM, 2007, pp. 1425–1434. (Cit. on p. 99).

[MBS97] Mark Mine, Frederick Brooks Jr., and Carlo Sequin. "Moving objects in space: exploiting proprioception in virtual-environment interaction". In: *Proc. SIGGRAPH '97*. ACM Press/Addison-Wesley, 1997, pp. 19–26. (Cit. on pp. 71, 93).

[MCR90] Jock Mackinlay, Stuart K. Card, and George G. Robertson. "A semantic analysis of the design space of input devices". In: *Hum.-Comput. Interact.* 5.2 (June 1990). Pp. 145–190. (Cit. on p. 23).

[MH04] Tomer Moscovich and John F. Hughes. "Navigating documents with the virtual scroll ring". In: *Proc. UIST '04*. ACM, 2004, pp. 57–60. (Cit. on pp. 91, 93, 98).

[MH08] Tomer Moscovich and John F. Hughes. "Indirect mappings of multi-touch input using one and two hands". In: *CHI '08*. Florence, Italy: ACM, 2008, pp. 1275–1284. (Cit. on pp. 44, 45).

[MI09] David McCallum and Pourang Irani. “ARC-Pad: absolute+relative cursor positioning for large displays with a mobile touchscreen”. In: *Proc. UIST ’09*. ACM, 2009, pp. 153–156. (Cit. on p. 93).

[Mic+09] Mark Micire, Munjal Desai, Amanda Courtemanche, Katherine M. Tsui, and Holly A. Yanco. “Analysis of natural gestures for controlling robot teams on multi-touch tabletop surfaces”. In: *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*. ITS ’09. Banff, Alberta, Canada: ACM, 2009, pp. 41–48. (Cit. on p. 22).

[MJ01] I. Scott MacKenzie and Shaidah Jusoh. “An Evaluation of Two Input Devices for Remote Pointing”. In: *Proceedings of the 8th IFIP International Conference on Engineering for Human-Computer Interaction*. EHCI ’01. London, UK, UK: Springer-Verlag, 2001, pp. 235–250. (Cit. on pp. 72, 91).

[MLG10] Sylvain Malacria, Eric Lecolinet, and Yves Guiard. “Clutch-free panning and integrated pan-zoom control on touch-sensitive surfaces: the cyclostar approach”. In: *Proc. CHI ’10*. ACM, 2010, pp. 2615–2624. (Cit. on pp. 22, 93, 97, 105).

[NHB09] Mathieu Nancel, Stéphane Huot, and Michel Beaudouin-Lafon. “Un espace de conception fondé sur une analyse morphologique des techniques de menus”. In: *Proceedings of the 21st International Conference on Association Francophone d’Interaction Homme-Machine*. IHM ’09. Grenoble, France: ACM, 2009, pp. 13–22. (Cit. on p. 23).

[OFH08] Alex Olwal, Steven Feiner, and Susanna Heyman. “Rubbing and tapping for precise and rapid selection on touch-screen displays”. In: *Proc. CHI ’08*. ACM, 2008, pp. 295–304. (Cit. on p. 97).

[Pec01] Choon Hong Peck. “Useful parameters for the design of laser pointer interaction techniques”. In: *CHI ’01 extended abstracts on Human factors in computing systems*. CHI EA ’01. Seattle, Washington: ACM, 2001, pp. 461–462. (Cit. on p. 91).

[Pie+11] Emmanuel Pietriga, Stéphane Huot, Mathieu Nancel, and Romain Primet. “Rapid Development of User Interfaces on Cluster-Driven Wall Displays with jBricks”. In: *Proceedings of the 3rd ACM SIGCHI symposium on Engineering interactive computing systems*. EICS ’11. Pisa, Italy: ACM, 2011, pp. 185–190. http://hal.inria.fr/docs/00/58/54/79/PDF/jbricks-eics11.pdf. (Cit. on p. 99).

[PRM00] Jason Pascoe, Nick Ryan, and David Morse. “Using while moving: HCI issues in fieldwork environments”. In: *ACM TOCHI* 7 (3 2000). Pp. 417–437. (Cit. on p. 45).

[RB10] Julie Rico and Stephen Brewster. "Usable gestures for mobile interfaces: evaluating social acceptability". In: *Proceedings of the 28th international conference on Human factors in computing systems*. CHI '10. Atlanta, Georgia, USA: ACM, 2010, pp. 887–896. (Cit. on p. 22).

[Rek97] Jun Rekimoto. "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. Banff, Alberta, Canada: ACM, 1997, pp. 31–39. (Cit. on pp. 15, 31, 32).

[RHL08] Anne Roudaut, Stéphane Huot, and Eric Lecolinet. "TapTap and MagStick: improving one-handed target acquisition on small touch-screens". In: *AVI '08*. Napoli, Italy: ACM, 2008, pp. 146–153. (Cit. on p. 45).

[RLG09] Anne Roudaut, Eric Lecolinet, and Yves Guiard. "MicroRolls: expanding touch-screen input vocabulary by distinguishing rolls vs. slides of the thumb". In: *Proc. CHI '09*. ACM, 2009, pp. 927–936. (Cit. on pp. 3, 96, 97, 103, 111).

[RS99] Jun Rekimoto and Masanori Saitoh. "Augmented surfaces: a spatially continuous work space for hybrid computing environments". In: *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*. CHI '99. Pittsburgh, Pennsylvania, United States: ACM, 1999, pp. 378–385. (Cit. on p. 11).

[RT09] Volker Roth and Thea Turner. "Bezel swipe: conflict-free scrolling and multiple selection on mobile touch screen devices". In: *CHI '09*. Boston, MA, USA: ACM, 2009, pp. 1523–1526. (Cit. on pp. 45, 51).

[SB05] Ramona E. Su and Brian P. Bailey. "Put them where? towards guidelines for positioning large displays in interactive workspaces". In: *Proceedings of the 2005 IFIP TC13 international conference on Human-Computer Interaction*. INTERACT'05. Rome, Italy: Springer-Verlag, 2005, pp. 337–349. (Cit. on p. 29).

[SBC11] Adriano Scoditti, Renaud Blanch, and Joëlle Coutaz. "A novel taxonomy for gestural interaction techniques based on accelerometers". In: *Proceedings of the 16th international conference on Intelligent user interfaces*. IUI '11. Palo Alto, CA, USA: ACM, 2011, pp. 63–72. (Cit. on p. 23).

[SCI04] Stacey Scott, Sheelag Carpendale, and Kori Inkpen. "Territoriality in collaborative tabletop workspaces". In: *Proc. CSCW*. 2004, pp. 294–303. (Cit. on p. 30).

[Sco+10] James Scott, Shahram Izadi, Leila Sadat Rezai, Dominika Ruszkowski, Xiaojun Bi, and Ravin Balakrishnan. “RearType: text entry using keys on the back of a device”. In: *MobileHCI '10*. Lisbon, Portugal: ACM, 2010, pp. 171–180. (Cit. on p. 46).

[SER03] Chia Shen, Katherine Everitt, and Kathleen Ryall. “UbiTable: Impromptu Face-to-Face Collaboration on Horizontal Interactive Surfaces”. In: *In Proc. UbiComp 2003*. 2003, pp. 281–288. (Cit. on p. 11).

[SH06] Masanori Sugimoto and Keiichi Hiroki. “HybridTouch: an intuitive manipulation technique for PDAs using their front and rear surfaces”. In: *MobileHCI '06*. Helsinki, Finland: ACM, 2006, pp. 137–140. (Cit. on p. 45).

[Sho+10] G. Shoemaker, T. Tsukitani, Y. Kitamura, and K. S Booth. “Body-centric interaction techniques for very large wall displays”. In: *Proc. NordiCHI*. 2010, pp. 463–472. (Cit. on pp. 23, 29, 31, 32, 71, 141).

[Sho+11] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. “Real-time human pose recognition in parts from single depth images”. In: *Proc. CVPR*. 2011, pp. 1297–1304. (Cit. on p. 71).

[SPM04] Carsten Schwesig, Ivan Poupyrev, and Eijiro Mori. “Gummi: a bendable computer”. In: *CHI '04*. Vienna, Austria: ACM, 2004, pp. 263–270. (Cit. on p. 22).

[STB07] G. Shoemaker, A. Tang, and K. S Booth. “Shadow reaching: a new perspective on interaction for large displays”. In: *Proc. UIST*. 2007, pp. 53–56. (Cit. on p. 71).

[Ste+87] Mark Stefik, Gregg Foster, Daniel G. Bobrow, Kenneth Kahn, Stan Lanning, and Lucy Suchman. “Beyond the chalkboard: computer support for collaboration and problem solving in meetings”. In: *Commun. ACM* 30.1 (Jan. 1987). Pp. 32–47. (Cit. on p. 11).

[Str+99] Norbert A. Streitz, Jörg Geißler, Torsten Holmer, Shin'ichi Konomi, Christian Müller-Tomfelde, Wolfgang Reischl, Petra Rexroth, Peter Seitz, and Ralf Steinmetz. “i-LAND: an interactive landscape for creativity and innovation”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*. CHI '99. Pittsburgh, Pennsylvania, United States: ACM, 1999, pp. 120–127. (Cit. on p. 12).

[TC03] Desney S. Tan and Mary Czerwinski. "Effects of Visual Separation and Physical Discontinuities when Distributing Information across Multiple Displays". In: *In Proceedings of Interact 2003*. 2003, pp. 252–255. (Cit. on p. 29).

[Tsa12] Theophanis Tsandilas. "Interpreting Strokes on Paper with a Mobile Assistant". In: *UIST '12: Proceedings of the 25th annual ACM symposium on User interface software and technology*. UIST '12. Cambridge, MA, USA: ACM, 2012, 10 pages, to appear. (Cit. on p. 9).

[VB05] Daniel Vogel and Ravin Balakrishnan. "Distant freehand pointing and clicking on very large, high resolution displays". In: *Proceedings of the 18th annual ACM symposium on User interface software and technology*. UIST '05. Seattle, WA, USA: ACM, 2005, pp. 33–42. (Cit. on pp. 23, 71, 72, 91, 93).

[Wac+08] J.P. Wachs, H.I. Stern, Y. Edan, M. Gillam, J. Handler, C. Feied, and M. Smith. "A gesture-based tool for sterile browsing of radiology images". In: *Journal of the American Medical Informatics Association* 15.3 (2008). Pp. 321–323. (Cit. on p. 71).

[WB03] Mike Wu and Ravin Balakrishnan. "Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays". In: *UIST '03*. Vancouver, Canada: ACM, 2003, pp. 193–202. (Cit. on p. 45).

[WB09] Raphael Wimmer and Sebastian Boring. "HandSense: discriminating different ways of grasping and holding a tangible user interface". In: *TEI '09*. Cambridge, United Kingdom: ACM, 2009, pp. 359–362. (Cit. on p. 45).

[WB10] Andrew D. Wilson and Hrvoje Benko. "Combining multiple depth cameras and projectors for interactions on, above and between surfaces". In: *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. UIST '10. New York, New York, USA: ACM, 2010, pp. 273–282. (Cit. on pp. 22, 34).

[Wei99] Mark Weiser. "The computer for the 21st century". In: *SIGMOBILE Mob. Comput. Commun. Rev.* 3.3 (July 1999). Pp. 3–11. (Cit. on p. 2).

[Wel93] Pierre Wellner. "Interacting with paper on the DigitalDesk". In: *Commun. ACM* 36.7 (July 1993). Pp. 87–96. (Cit. on p. 2).

[Whe03] Elaine Wherry. "Scroll ring performance evaluation". In: *CHI '03 EA*. ACM, 2003, pp. 758–759. (Cit. on pp. 91, 93, 98).

[Wig+06] Daniel Wigdor, Chia Shen, Clifton Forlines, and Ravin Balakrishnan. "Table-centric interactive spaces for real-time collaboration". In: *Proceedings of the working conference on Advanced visual interfaces*. AVI '06. Venezia, Italy: ACM, 2006, pp. 103–107. (Cit. on p. 11).

[Wig+07] Daniel Wigdor, Clifton Forlines, Patrick Baudisch, John Barnwell, and Chia Shen. "Lucid touch: a see-through mobile device". In: *UIST '07*. Newport, Rhode Island, USA: ACM, 2007, pp. 269–278. (Cit. on p. 46).

[Wig+09] Daniel Wigdor, Hao Jiang, Clifton Forlines, Michelle Borkin, and Chia Shen. "WeSpace: the design development and deployment of a walk-up and share multi-surface visual collaboration system". In: *Proceedings of the 27th international conference on Human factors in computing systems*. CHI '09. Boston, MA, USA: ACM, 2009, pp. 1237–1246. (Cit. on p. 11).

[WMA08] Jacob O. Wobbrock, Brad A. Myers, and Htet Htet Aung. "The performance of hand postures in front- and back-of-device interaction for mobile computing". In: *Int. J. Hum.-Comput. Stud.* 66 (12 2008). Pp. 857–875. (Cit. on p. 45).

[WMH07] John Williamson, Roderick Murray-Smith, and Stephen Hughes. "Shoogle: excitatory multimodal interaction on mobile devices". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. CHI '07. San Jose, California, USA: ACM, 2007, pp. 121–124. (Cit. on p. 22).

[YHN07] Beth Yost, Yonca Haciahmetoglu, and Chris North. "Beyond visual acuity: the perceptual scalability of information visualizations for large displays". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. CHI '07. San Jose, California, USA: ACM, 2007, pp. 101–110. (Cit. on p. 91).

[Zig+10] Jamie Zigelbaum, Alan Browning, Daniel Leithinger, Olivier Bau, and Hiroshi Ishii. "g-stalt: a chirocentric, spatiotemporal, and telekinetic gestural interface". In: *Proc. TEI '10*. ACM, 2010, pp. 261–264. (Cit. on pp. 22, 93).

[ZMB96] Shumin Zhai, Paul Milgram, and William Buxton. "The influence of muscle groups on performance of multiple degree-of-freedom input". In: *Proceedings of the SIGCHI conference on Human factors in computing systems: common ground*. CHI '96. Vancouver, British Columbia, Canada: ACM, 1996, pp. 308–315. (Cit. on pp. 23, 91, 98).

List of Figures

List of Tables

Appendices

A

WILD: Technical Equipment

The WILD room (**w**all-sized **i**nteraction with **l**arge **d**atasets) had been inaugurated in 2009 and is physically located at the *Laboratoire de Recherche en Informatique (LRI)* on the Orsay campus of Université Paris Sud. It features many interactive surfaces: a large *wall-sized display* (1.8 m ×5.5 m), an *interactive table*, and several portable hand-held devices ranging from paper sheet sized Apple iPad tablets [1] to smart phone sized Apple iPod touch [2] devices.

The wall-sized display (see Fig.A.1b) is powered by a cluster of 16 computers and two front-end computers and consists of 32 off-the-shelf 30-inch monitors organized in an 8×4 grid and supports the total resolution of 131 million pixels (20.480×6.400 pixels) with a high pixel density of about 100 dpi. Each group of 8 displays is mounted on a movable cart (four carts in total) that can be rearranged in various configurations, such as a complete flat surface or a triptych layout.

The interactive table (see Fig.A.1c) consists of an infra-red camera and projector inside a physical table covered with an acrylic surface. Stripes of LEDs around the frame of the acrylic fed infra-red light sideways into the acrylic.

[1] http://www.apple.com/ipad/
[2] http://www.apple.com/ipod/

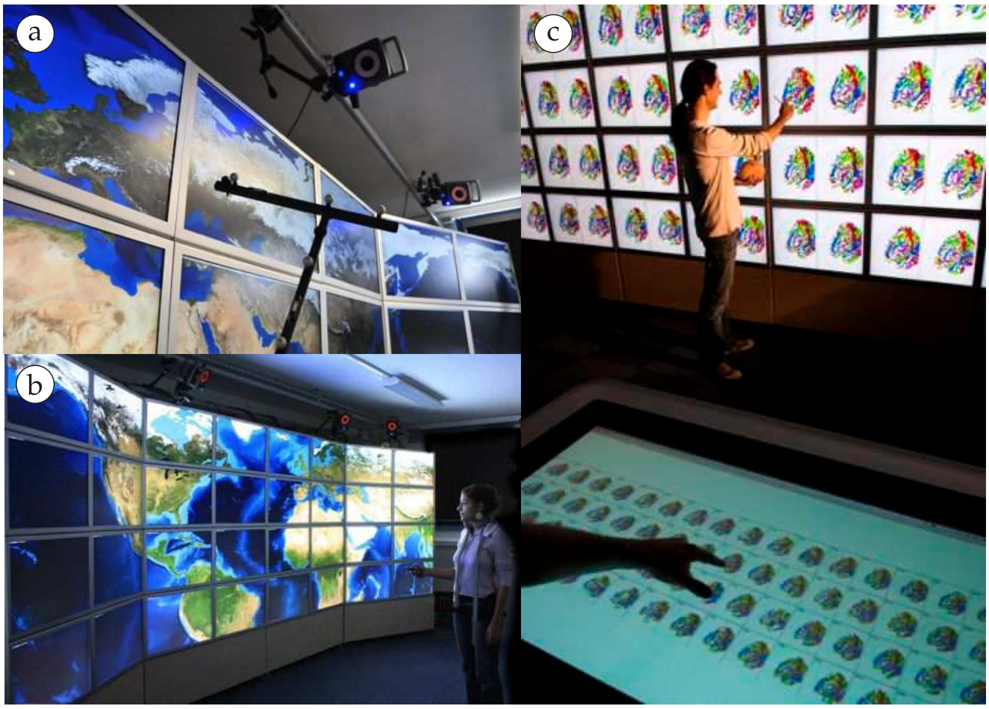


Figure A.1: The wild platform: a) VICON tracking system, b) user pointing to the wall with a hand-held device, c) a large wall display used in combination with an interactive table.

Depending on the type of acrylic, the interactive table supports either the FTIR (frustrated internal reflection) ([Han05]) or the DSI (diffused surface illumination) technique[3]. The FTIR technique can detect the user's touch on the table surface, the DSI technique can detect touch as well as objects with mounted marker tags on top of the table's surface.

Ten VICON motion tracking cameras (see Fig.A.1a) track passive IR retroreflective markers and provide 3D object coordinates with sub-millimeter accuracy at 200Hz.

Research strategies, software toolkits, and interaction technique development is further described in an article I published with colleges from the INSITU research group in the *Computer* magazine [Bea+12].

[3]http://iad.projects.zhdk.ch/multitouch/?p=90

B

On-body Touch Prototype

The on-body touch prototype includes three main components, as suggested by Shoemaker et al. [Sho+10]:

1. *Sensing:* the hardware that captures body data may be based on vision, magnetism, markers or some other system;
2. *Modeling:* the software that interprets the raw data and constructs a model of the user's body, e.g. a skeleton or 3D volumetric model, depending upon the spatial and temporal resolution, accuracy and other capacities of the sensor; and
3. *Interaction:* the logic of the interaction techniques with respect to the above model, e.g. interaction related to detected postures. Our prototype defines on-body targets, detects the positions of on-body touches by the user's hand and identifies potentially touched targets.

This cascading modular approach keeps each component relatively independent and offers different options when designing on-body touch interaction techniques, depending upon the specific capabilities of each module. However, the modules do involve some interdependencies, when combined. In particular, the 3D model of the user depends upon the precision of the sensor: the greater the accuracy of the body data and the skeleton reconstruction, the more targets that can be specified and accurately differentiated by the system. In addition, the quality of the interaction depends upon the spatial and tem-

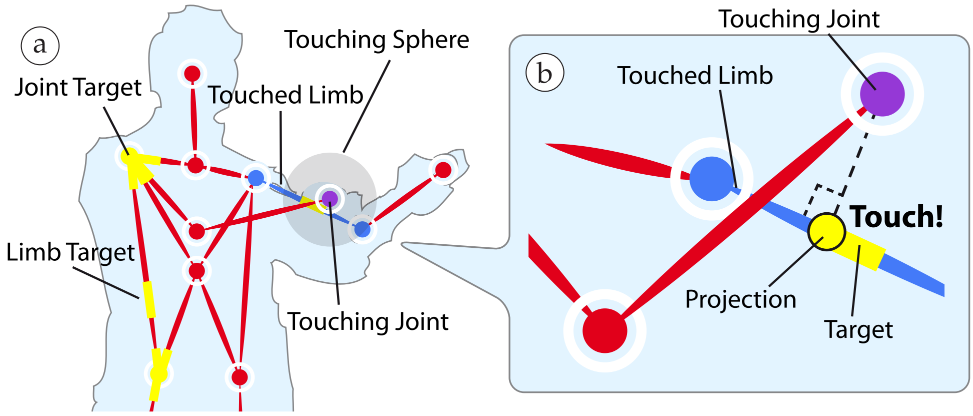


Figure B.1: Detecting a touch on the user's body. a) A "touched limb" intersects the "touching sphere". b) A "touched target" is detected on the touched limb if the orthogonal projection of the "touching joint" on the touched limb is inside a body target.

poral reliability and robustness of model; e.g., to avoid missing or triggering unintentional touches after detecting an incorrect position.

B.1 The Interaction Module

Our prototype is based on a common skeleton reconstruction of the user's body, made of joint locations and the corresponding limbs between them (see Fig. B.1a). A full volumetric body reconstruction might provide more possibilities for defining on-body targets and other "body touching interactions", such as gestures. However, we found that the skeleton-based approach is not only simpler to implement, but is applicable in more varied contexts and has proven effective for our studies of on-body interaction. If we assume that the model provides the necessary, accurate body data, implementing On-body touch is straightforward and consists of three functions:

On-body Target Specification. We identify on-body targets with respect to their location on a limb (Fig. B.1a). This location is defined relative to the length of the limb, making it independent of the user. On-body targets are essentially sub-segments of a limb and can be specified with a size or as a percentage of the limb segment. Targets may also be located on a specific skeleton joint and thus slightly extend to all connected limbs (see "Joint Target" in Fig B.1a).

Body-Touch Detection. We can define a "touching joint" on the skeleton model, such as the right index finger or the left hand, and can detect on-body touches in two ways: The *Position-Based* approach computes the 3D intersection of a "touching sphere" centered around the "touching joint" and a possible segment of the skeleton (Fig. B.1b). This approach is simple and does not require hardware other than the tracking system, but is sensitive to the resolution of the skeleton reconstruction: Whether or not the user's hand is fully reconstructed, the radius of the "touching sphere" must be adapted to estimate a touch from the finger and the spatial proximity of some limbs relative to the radius of the sphere. The *Mechanical* approach uses an additional sensor on the user's finger to trigger more robust touch events. We implemented both approaches in our prototype.

Touched-Target Identification. Once a touch has been detected, our prototype calculates whether or not it lies on a body target. It computes the orthogonal projection of the touching joint on the touched limb and then checks whether it lies within the segment associated with the target (Fig. B.1b).

B.2 Implementation

We implemented two prototypes based on these principles: (i) *On-body Touch Hi* uses *VICON*'s high-fidelity motion capture system and provided highly accurate data for our whole-body interaction experiments; (ii) *On-body Touch Lo* uses the less accurate but more affordable low-fidelity *Kinect* sensor that offers a portable, low-cost alternative.

On-body Touch Hi The sensing and modeling modules of *On-body Touch Hi* both use the *VICON* to detect the user's body segments and to update a skeleton model of the user's body (Fig. B.2a). Ten cameras track passive IR reflective markers in three dimensions, with sub-millimeter accuracy up to 200 Hz. We chose not to use the high-fidelity feature that precisely tracks the position of the whole body, including the volume of each limb, because users had to wear a skin-tight bodysuit and it required a cumbersome calibration process. Instead, we chose a less elegant but simpler approach, accurate and reliable enough for our needs. We mounted IR markers on protective sport gear, including a belt, a forearm protector for the dominant arm, and protectors for both shoulders and legs (Fig. B.2a). These could be adjusted to fit

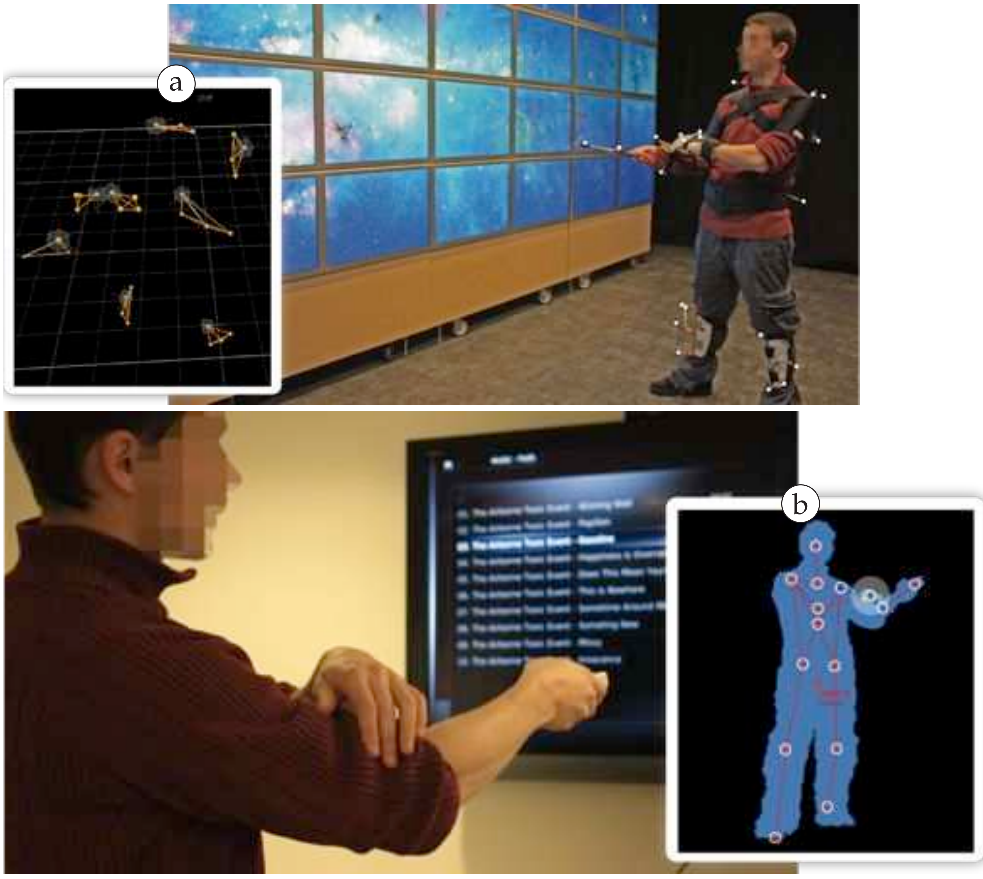


Figure B.2: The on-body touch prototype running (a) with a high-end tracking system in front of a wall-sized display or (b) with an affordable depth sensing camera in front of a TV screen. The insets show the data from the tracking systems.

users of any height or body shape and did not require them to change clothing.

The modeling component interprets the 3D location and orientation of these six tracked objects to reconstruct the user's skeleton into segments, joined by the major articulations: left and right upper arms, forearms, thighs and tibias, left and right sides of the torso. Since the IR markers are not placed directly onto these body parts, *On-body Touch Hi* requires a short calibration phase to ensure accurate computation of the relative positions of the 3D objects on the user's body, in which the user touches a series of body parts and their 3D locations are recorded relative to the tracked objects.

The interaction module implements the on-body target specification and touch-identification functions. To enhance accuracy, this prototype uses the "mechanical approach" to detect touches: users wear a glove on their non-dominant hand that has a force sensor attached to the index finger and VICON IR markers on top. The combination of the two provides extremely accurate detection of touches onto the on-body targets, which was required for our experiment that compared 18 on-body targets. The error rate is approximately 8% and affects a few specific situations, which are explained in the results section of the experiment.

On-body Touch Lo The *On-body Touch Lo* prototype demonstrates the feasibility of our approach with simpler, less expensive hardware (Fig. B.2a). The sensing module consists of a *Kinect* connected to a standard computer. We use the free *OpenNI* library to acquire body data; skeleton reconstruction relies upon the *PrimeSense* middleware for *OpenNI*[1].

The interaction components are the same for the high- and low-fidelity prototypes, except that the latter detects touch based only on the software "touching sphere" principle. Accuracy is lower, both because of the lower tracking resolution and also because the user's hands are not fully reconstructed. We conducted informal tests with 12 targets and the error rate was approximately 20%. This could be improved with better-tuned algorithms (e.g., adaptive radius of the touching sphere with respect to the tested limb), improved tracking sensors or even with a glove similar to the *On-body Touch Hi* prototype. However, with appropriate design (around 8 targets distributed evenly across the body), the actual low-fidelity prototype is sufficient for home use, for example, to control interaction with a television in the living room.

[1] `http://www.openni.org/`

C
Selected Publications

This chapter contains three publications that give a complete related work review, and details on the design and evaluation of the interaction techniques described in:

Chapter 4—"The Effect of Support on Input Motor Assemblies", p. 43.
Julie Wagner, Stéphane Huot, and Wendy E. Mackay. "BiTouch and BiPad: Designing Bimanual Interaction for Hand-held Tablets". In: *Proceedings of the 30th international conference on Human factors in computing systems*. CHI '12. Austin, Texas, USA: ACM, 2012, pp. 2317–2326. `http://hal.inria.fr/hal-00663972/PDF/bipadA.pdf`, see p. 149.

Chapter 5—"Interaction Effects between Input Motor Assembly and Affected Body Parts", p. 69.
Julie Wagner, Mathieu Nancel, Sean Gustafson, Stéphane Huot, and Wendy E. Mackay. "A Body-centric Design Space for Multi-surface Interaction". In: *Proceedings of the 31th international conference on Human factors in computing systems*. CHI '13. New York, NY, USA: ACM, 2013, to appear, see p. 159.

Chapter 6—"Interaction Effects between Two Input Motor Assemblies", p. 89
Mathieu Nancel, **Julie Wagner**, Emmanuel Pietriga, Olivier Chapuis, and Wendy Mackay. "Mid-air pan-and-zoom on wall-sized displays". In: *Proceedings of the 2011 annual conference on Human factors in computing systems*. CHI '11. Vancouver, BC, Canada: ACM, 2011, pp. 177–186. *Best*

Paper Award. http://hal.archives-ouvertes.fr/docs/00/59/12/94/PDF/CHI11-pzwild-avf.pdf, see p. 169.

BiTouch and BiPad: Designing Bimanual Interaction for Hand-held Tablets

Julie Wagner[1,2,3]
wagner@lri.fr
[1] INRIA
F-91405 Orsay, France

Stéphane Huot[2,1,3]
huot@lri.fr
[2] Univ Paris-Sud (LRI)
F-91405 Orsay, France

Wendy E. Mackay[1,2,3]
mackay@lri.fr
[3]CNRS (LRI)
F-91405 Orsay, France

Figure 1. Bimanual interaction with BiPad: a) navigating a PDF, b) shifting to uppercase, c) zooming on a map. The non-dominant support hand can tap, make gestures or perform chords, thus modifying interaction by the dominant hand.

ABSTRACT

Despite the demonstrated benefits of bimanual interaction, most tablets use just one hand for interaction, to free the other for support. In a preliminary study, we identified five holds that permit simultaneous support and interaction, and noted that users frequently change position to combat fatigue. We then designed the BiTouch design space, which introduces a support function in the kinematic chain model for interacting with hand-held tablets, and developed BiPad, a toolkit for creating bimanual tablet interaction with the thumb or the fingers of the supporting hand. We ran a controlled experiment to explore how tablet orientation and hand position affect three novel techniques: bimanual taps, gestures and chords. Bimanual taps outperformed our one-handed control condition in both landscape and portrait orientations; bimanual chords and gestures in portrait mode only; and thumbs outperformed fingers, but were more tiring and less stable. Together, BiTouch and BiPad offer new opportunities for designing bimanual interaction on hand-held tablets.

Author Keywords

Bimanual Interaction; Hand-held tablets; Multi-touch tablets; BiTouch design space; BiPad.

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User Interfaces—*Interaction styles; Input devices and strategies*

J. Wagner, S. Huot, W. E. Mackay.
BiTouch and BiPad:
Designing Bimanual Interaction for Hand-held Tablets.
In CHI'12: Proceedings of the 30th International Conference on Human Factors in Computing Systems, ACM, May 2012.
Authors Version

INTRODUCTION

Multi-touch tablets have become increasingly popular over the past few years, combining relatively large screens with portability. Their form factor encourages uses in situations in which the user stands or walks, for example teachers can control simulations in class and nurses can track patients on interactive clipboards [7]. Although commercial tablets offer intuitive interaction techniques such as a *swipe* to displace an object or a *tap* to select an item, they do not fully exploit the range of interaction possibilities found in the research literature. In particular, tablets are not designed to support bimanual input, despite the demonstrated ability to increase performance [18] and precision [4], as well as to enhance the user experience [16, 29].

Existing bimanual interaction techniques were designed for independently supported displays or tabletops. Portable devices pose an additional challenge: how to account for the need to hold the device while interacting with it. Very small devices, such as PDAs and smart phones, offer limited possibilities for bimanual interaction, usually just typing with both thumbs. Multi-touch tablets, with their larger screens, offer as-yet unexplored opportunities for true bimanual interaction. Our goal is to better understand the design space for bimanual, multi-touch interaction on hand-held tablets and to demonstrate how designers can obtain the benefits of bimanual techniques, taking into account the challenge of supporting the device while interacting with it.

We begin by analyzing the related literature and describe a preliminary study that investigates how users hold tablets as they interact. Next, we present the *BiTouch* design space which identifies the key dimensions for designing bimanual multi-touch interaction. We next present BiPad, a toolkit that helps designers add various bimanual interaction to off-the-shelf multi-touch tablets, illustrated with three sample applications. We also report the results of an experiment that compares one- and two-handed interaction performance with respect to tablet orientation, finger placement and interaction technique. We conclude with implications for design and directions for future research.

RELATED RESEARCH

Desktop-based bimanual interaction techniques increase both performance and accuracy [1, 5, 12] and are more convenient when performing highly demanding cognitive tasks [16, 10]. Some techniques provide symmetric control [2]. For example, *Symspline* gives both hands equal roles when manipulating curves [15]. However, most bimanual interaction techniques build upon Guiard's *kinematic chain model* [9], based on his observations about the asymmetric relationship between the two hands [1]. For example, toolglasses, magic lenses and bimanual palettes [5, 3, 17] each use the non-dominant hand to control the position of an interactive palette while the dominant hand selects specific functions.

Bimanual Interaction: Stationary Multi-touch Surfaces

Multi-touch tables and graphics tablets are inherently well-adapted to bimanual interaction, since the user can use multiple fingers from either or both hands. Studies have shown that bimanual interaction techniques can improve performance [6, 14] and selection accuracy [4]. However, these studies assume that both hands are free to interact, e.g. on a stationary multi-touch surface or a small multi-touch device placed on a table. We are interested in hand-held tablets which require at least one hand to support the device, thus restricting the ability to interact.

Bimanual Interaction: Small Portable Devices

Commercially available PDAs and smart phones are designed primarily for one-handed interaction [20] due to their small size. Most interaction is accomplished with the index finger, although some techniques use the thumb, since it can reach the screen from most carrying positions [11, 13, 22]. Other approaches use the outer frame of the phone to improve pointing accuracy [8] or to disambiguate among actions and enrich the interaction vocabulary [21].

Several research prototypes offer the potential for bimanual interaction by adding hardware. For example, *HandSense* [27] uses capacitive sensors to distinguish among six different grasping styles. One could create simple bimanual tasks by allowing these grasps to modify the actions of the dominant interaction hand. An alternative is *HybridTouch* [25], which adds a rear touchpad to a PDA to enable simultaneous front and back interaction.

Wobbrock et al. [28] investigated how different hand positions on the front or back of a handheld device affect interaction performance with the index finger or the thumb. They found that the index finger performed best in all conditions, front or back, and that horizontal movements were faster and more accurate. Although useful for comparing thumb and finger performance on small devices, additional research is needed to understand bimanual interaction on larger portable devices, such as multi-touch tablets.

Bimanual Interaction: Multi-touch Tablets

Hand-held tablets offer new possibilities for bimanual interaction. Although their larger screen size and bezels make two-handed thumb typing less convenient, they also afford various support positions and can accommodate interaction with the thumbs and multiple fingers from both hands.

To date, most bimanual interaction techniques require additional hardware, e.g. to detect touches on the back or sides of the device. For example, *RearType* [24] includes a physical keyboard on the back of a tablet PC. Users hold it with both hands while entering text, thus avoiding an on-screen keyboard and graphical occlusion by the fingers. *Lucid Touch* [26] is a proof-of-concept see-through tablet that supports simultaneous touch input on the front and on the back of the device. Users hold the device with both hands, with thumbs on the front and remaining fingers on the back. The device is small enough that users can reach the entire screen, allowing multi-touch interaction with both support hands without graphical occlusion. However, the arm-mounted camera currently makes this approach impractical.

Another intriguing possibility is *Gummi* [23], a prototype "bendable" tablet that enables limited bimanual interaction by deforming the device. For example, a user could scroll through a list via a 2D position sensor on the back and then select an item by bending the device. Such dual-surface approaches are well suited for simple selection and navigation tasks [30], but are less appropriate for complex tasks that require additional input from the back or when users adjust how they hold the tablet.

Our goal is to incorporate bimanual interaction on tablets, using only the multi-touch surface without additional hardware. The next section describes a preliminary study that investigates how users unconsciously hold tablets while interacting with them, as they sit, stand and walk.

PRELIMINARY STUDY: HOLDING TABLETS

Studying how people 'naturally' hold tablets is tricky. Rather than asking directly, we asked users to perform a distractor task while observing how they held the tablet.

Participants. Six men and two women, average age 30. Four owned iPads, four had never used a tablet.

Apparatus. Apple iPad1 (display: 9.7", weight: 680 g, dimensions: $19 \times 24.3 \times 1.3$ cm).

Procedure. We told participants that we were interested in how pointing and scrolling performance varies as people sit, stand and walk, given different tablet orientations. This was intentionally misleading, since we were really studying how they unconsciously held the tablet while interacting with it. The true experiment was a [2x3] within-subjects design with two factors: tablet *orientation* (landscape, portrait) and *stance* (sit, stand, walk), with tablet *hold* as the dependent measure. The distractor tasks were *pointing* (tapping five successive on-screen targets) and *scrolling* (moving a slider's thumbwheel from one end to the other). Pointing targets were distributed across six equal squares on the screen; slider

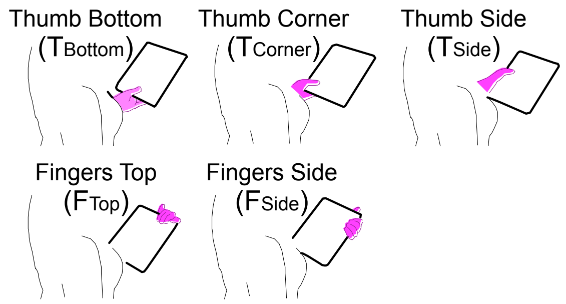


Figure 2. Five spontaneous holds (portrait orientation).

positions included the four screen borders and horizontally and vertically in the screen center.

Participants were asked to hold the iPad comfortably and perform each task as quickly as possible. They were allowed to adopt a new hold only when beginning a new block. Sessions lasted approximately 45 minutes. At the end, we debriefed each participant as to the true goal of the study to learn how they chose to hold the tablets. We first asked them to reproduce the holds they had used and then to adapt them so that the fingers or thumb of the support hand could reach the touch screen. We asked them to rate comfort and ease of interaction when using the support hand to interact and whether they had suggestions for other holding positions.

Data collection. We videotaped each trial and coded how participants supported the tablet with the non-dominant hand, wrist or forearm. We collected touch events, including those that occurred outside experiment trials and while reading instructions. We also measured completion time per trial.

Results

We did not find a single, optimal hold and found significant differences according to experience. All four novices used the same uncomfortable position: the fingers, thumb and palm of their non-dominant hand supported the center of the tablet, like a waiter holding a tray. Novices found this tiring but worried that the tablet would slip if they held it by the border. None found other holds. In contrast, the four experts easily found a variety of secure, comfortable holds. We identified ten unique holds, five per orientation, all of which involved grasping the border of the tablet with the thumb and fingers. Fig. 2 shows these five holds in portrait mode, with the thumb on the bottom, corner or side, or the fingers on the top or side.

Table 1 shows how these holds were distributed across the six conditions: most common was F-side (41%), least common was T-side (9%). The latter was deemed least comfortable, especially in landscape mode, but participants felt that they could use it for a short time. Experts tried nine of ten possible holds in the sitting and walking conditions, but only six when standing, omitting F-top or T-side in both orientations. Individuals varied as to how many unique holds they tried, from three to eight of ten possible. All switched holds at least once and two switched positions often (50% and 66%) across different blocks of the same condition.

Table 1. Total holds per condition (expert users)

	F_{side}	T_{bottom}	F_{top}	T_{corner}	T_{side}
Landscape	3	4	4	4	1
	8	4	0	4	0
	4	4	7	0	1
Portrait	8	3	1	0	4
	8	4	0	4	0
	8	1	3	1	3
	41%	21%	16%	14%	9%

We were also interested in whether *accidental touches*, defined as touches located more than 80 pixels from the target or slider, during or outside of experiment trials, interfered with intentional touches by the dominant hand. Experts who carried the tablet by the border made very few accidental touches (3%). All were with the dominant hand, far from the screen border, suggesting that they unconsciously prevented the support hand from touching the screen.

Design Implications

First, tablets can feel heavy and users are more comfortable when they can change orientation or swap the thumb and fingers. We should thus seek a small set of roughly equivalent bimanual interactive holds that are easy to shift between, rather than designing a single, 'optimal' hold. Second, users can use the thumb and fingers of the support hand for interaction. We can thus create interactive zones on the edges of the tablet, corresponding to the holds in Fig. 2, which were not vulnerable to accidental touches. Fig. 3 shows these zones in portrait and landscape mode. Although changes in the form factor of a tablet, such as its size, shape or weight, may affect these holds, users are still likely to shift between holds for comfort reasons, just as when reading a book or holding a notebook.

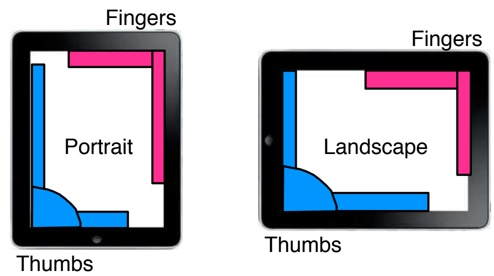


Figure 3. Five support-hand interaction zones.

The next section describes BiTouch, a design space for exploring how to incorporate bimanual interaction on hand-held multitouch tablets.

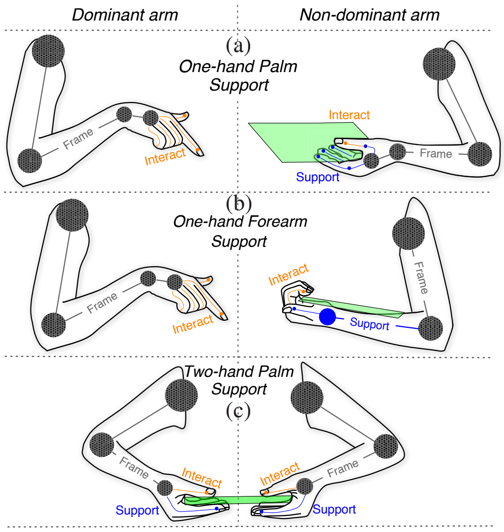


Figure 4. The user creates a spatial *frame*, *supports* the device, and *interacts* with it. Different holds offer different trade-offs with respect to interactive power and comfort.

BiTouch DESIGN SPACE

Unlike desktop PCs or multi-touch tables, bimanual interaction on hand-held tablets must account for the dual role of the non-dominant hand as it simultaneously carries the tablet and interacts with it. Although we designed the BiTouch design space to explore bimanual interaction on hand-held tablets, the reasoning applies to a wider range of human-body interaction with objects [19] and devices ranging from small, mobile devices to large, fixed interactive tables or walls.

Kinematic Chain: Frame, Support, Interact

The first step is to understand the complementary roles of support and interaction. Guiard's [9] analysis of bimanual interaction emphasizes the asymmetric relationship commonly observed between the two hands. He proposes the kinematic chain as a general model, in which the shoulder, elbow, wrist and fingers work together as a series of abstract motors. Each consists of a proximal element, e.g. the elbow, and a distal element, e.g. the wrist, which together make up a specific link, e.g. the forearm. In this case, the distal wrist must organize its movement relative to the output of the proximal elbow, since the two are physically attached.

Guiard argues that the relationships between the non-dominant and dominant hands are similar to those between proximal and distal elements: the former provides the spatial frame of reference for the detailed action of the latter. In addition, the movements of the proximal element or non-dominant hand are generally less frequent and less precise and usually precede the movements of the higher frequency, more detailed actions of the distal element or dominant hand.

We see the kinematic chain in action when users interact with hand-held tablets: the non-dominant hand usually supports the tablet, leaving the fingers and thumb of the dominant hand free to interact. Fig. 4 shows three bimanual alternatives,

Table 2. Trading off *framing*, *support* and *interaction* functions of the kinematic chain with respect to the body and the device.

Framing	
Location:	proximal link in the kinematic chain
Distribution:	1 – n body parts
Support	
Location:	none or middle link in the kinematic chain
Distribution:	0 – n body parts
Independence:	0% – 100% body support
Interaction	
Location:	distal link in the kinematic chain
Distribution:	1 – n body parts
Degrees of freedom:	0% – 100% body movement
Technique:	touch, deformation,...

based on the location of tablet support within the kinematic chain: the palm or forearm of the non-dominant arm (Fig. 4a, 4b); shared equally between the palms of both hands (Fig. 4c). In each case, the most proximal links control the spatial frame of reference; support links are always intermediate between framing and interaction links; and the most distal links use whatever remains of the thumb and fingers to interact.

The preliminary study highlighted ten user-generated support holds that permit the thumb or fingers to reach the interactive area. Each poses trade-offs between comfort and degrees of freedom available for interaction. For example, supporting the tablet with the forearm (Fig. 4b) provides a secure, stable hold but forces the fingers to curl around the tablet, leaving little room for movement. In contrast, holding the tablet in the palm (Fig. 4a) gives the thumb its full range of movement, but is tiring and less stable.

Note that comfort is subjective, influenced not only by the physical details of the device, such as its weight, thickness and size of the bezels, but also by how the tablet is held. For example, shifting between landscape and portrait orientations changes the relative distance between the tablet's central balance point and the most distal part of the support link. The tablet acts as a lever: users perceive it as heavier as support moves further from the fulcrum. The next step is to formalize these observations into a design space that describes existing and new bimanual holds and interaction techniques.

BiTouch Design Space

Table 2 summarizes the key dimensions of the BiTouch design space, according to framing, support and interaction functions of the kinematic chain. Each is affected by the relationship between specific characteristics of the human body, the physical device and the interaction between them.

Framing is handled at the most proximal *locations* within the kinematic chain and may be *distributed* over multiple parts of the body. ***Support*** always occurs in *locations* within the kinematic chain, distal to the frame. Support may be completely *distributed* over one or more body parts, symmetrically or not; shared with an *independent* support, e.g. a table or lap; or omitted, e.g. interacting on a freestanding interactive table.

Interaction is always handled at the most distal *location* in the kinematic chain, immediately after the support link. Interaction may be *distributed* across one or more body parts, often incorporating the thumbs or sets of fingers. The *degrees of freedom* available for interaction depend upon what remains after framing and support functions have been allocated, e.g. a finger tip, and the inherent movement capabilities of the body part, e.g. the pinky has little independent movement compared to the index finger. Possible *interaction techniques* are affected by all of the above, as well as the technical capabilities of the device. For example, touch sensors might appear on the front, side or back of the device, or the device itself might be deformable.

Hands that interact as well as support the device have fewer degrees of freedom available for movement. We thus expect the support hand to be non-dominant, capable of limited interaction, e.g. mode switches or menu choices, that frame the interaction of the freer dominant hand.

The BiTouch design space allows us to describe all of the user-generated holds from the preliminary study, as well as many from the literature, e.g. bimanual interaction on free-standing interactive tabletops. It also suggests directions for designing new bimanual interaction techniques. For example, although the hold in Fig. 4c did not appear in the preliminary study, it becomes an obvious possibility if we examine ways to share support across hands. Similarly, once we understand which thumbs or fingers are available for interaction and what constrains their potential movement, we can design novel interaction techniques.

The five basic holds in Fig. 2 can each support an interactive area on the edge of the tablet, reachable by either the thumb or fingers of the support hand. The BiTouch design space helps us create a set of novel bimanual interaction techniques that take into account the potential of the thumbs and fingers at the end of the kinematic chain. For example, all thumbs and fingers have at least a small amount of mobility available to perform *Taps*. The thumb in the T_{corner} hold is fully mobile and can perform *Gestures*. The presence of multiple fingers in the F_{side} hold makes it possible to perform *Chords*. The non-dominant role of the support hand suggests that these *Taps*, *Gestures* and *Chords* can be used to frame more elaborate interaction by the dominant hand, e.g. to select a menu item or to shift color while drawing a line.

BiPad TOOLKIT AND APPLICATIONS

Based on our preliminary study and the BiTouch design space, we designed the BiPad toolkit to help developers add bimanual interaction to off-the-shelf multi-touch tablets. BiPad creates five interactive zones, corresponding to those in Fig. 2, where the fingers or the thumb of the supporting hand can interact.

Software Prototype

The BiPad toolkit, written in Objective-C on Apple's iOS operating system, supports the development of bimanual applications as follows:

BiPad applications consist of one or more views, widgets and controllers, similar to standard iOS applications. The framework lays out the interface in the main view to control overlay feedback and advanced input management required to enable BiTouch interaction. The application defines BiPad-enabled functions that can be mapped to interactions with the support hand. For example, a text editing application could define `shift` and `num` functions equivalent to pressing the shift or number keys of a virtual keyboard.

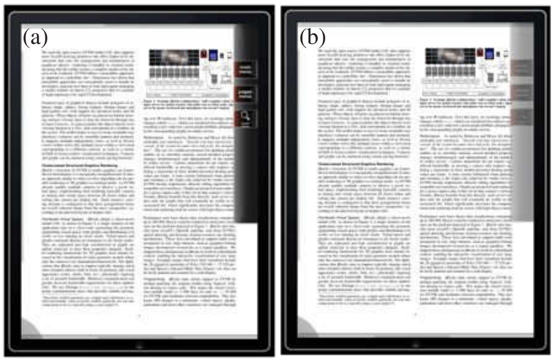


Figure 5. BiPad. a) F_{side} zone is active; other zones are shrunken. b) Unused zones remain partially visible if commands were assigned.

BiPad zones appear on the sides and corners of the screen (Fig. 5). Applications can define various interactions for the support hand and modify the default visual representation, e.g., buttons for taps and guides for chords. Zones are displayed as 80-pixel strips, of which the 40 outermost are semi-transparent, on top of the edges of the application view. Zones may be permanently or temporarily visible and the user's hand position determines which is active. Temporarily visible areas shrink automatically when not in use, displaying only a narrow semi-transparent strip of pixels on the appropriate side. Touching once on the outer part of a shrunken BiPad zone causes it to slide out and enables interaction. If a zone contains interaction widgets and is configured to be temporarily visible, it does not shrink completely but remains semi-transparent (Fig. 5b).

BiPad Interaction Techniques

BiPad introduces three predefined interaction techniques for the support hand: bimanual *Taps*, *Chords* and *Gestures*. Bimanual *Taps* involve a press-and-release action on a button within a BiPad zone, using a finger or the thumb (Fig. 6a). Bimanual *Chords* involve multiple fingers pressing down simultaneously within a BiPad zone, and are not possible with thumbs. Fig. 6b shows how pressing the 'stroke' button with the index finger adds additional finger positions below. The user can adjust the stroke size by holding down a second finger on the appropriate button.

Bimanual *Gestures* involve sliding the thumb or finger, starting from a BiPad zone or from an edge related to a BiPad zone, as in Bezel Swipe [21]. In the border zones, *Gestures* are limited to orthogonal movements from the edge, but offer additional degrees of freedom for the thumb in the corner (up-to-down, right-to-left and diagonal). Small stroke shapes indicate the direction of the gesture and its function (Fig. 6c).

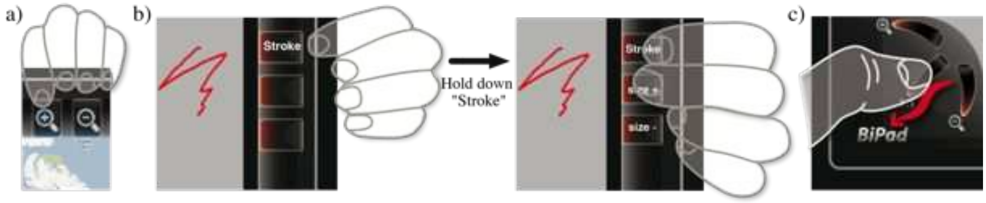


Figure 6. BiPad interaction techniques: a) Taps on buttons. b) Chords with multiple fingers. c) Gestures in multiple directions.

The application defines which BiPad interaction(s) will trigger which function in which zone(s). Applications can specify several interaction techniques for the same function depending upon which BiPad zone (and therefore *Hold*) the user registers. For example, an application might specify that a *Tap* with a finger on the F_{side} zone and a downward *Gesture* with the thumb in the T_{corner} zone will both shift modes for the dominant hand, triggering a pop-up menu rather than selecting an on-screen object.

BiPad Applications

We used BiPad to implement three applications that illustrate how to add bimanual interaction to handheld tablets (Fig. 1).

Quasi-modes and Shortcuts

BiPDF (Fig. 1a) is a PDF reader that uses standard touch gestures to navigate through pages, scroll or zoom the document. A pie menu contains additional commands, e.g. first/last page. As with many tablet applications, the user must touch and dwell to activate the menu instead of executing a gesture. We added a bimanual tap that speeds up interaction: while the user is touching the screen with the dominant hand, a tap on a BiPad button activates the menu immediately.

BiText (Fig. 1b) lets users create custom bimanual shortcuts for text entry, e.g. a button for the '`space`' key and a quasi-mode button for the soft keyboard's '`keypad`' key. Although the dominant hand can also reach these keys, it requires extra movement. The user can also assign any key from the keyboard to a BiPad button by simultaneously pressing the two. Modifier keys, such as the '`keypad`' key become quasi-modes: they activate the mode as long as they are being pressed. Two other BiPad buttons accept or reject the suggestions from the standard text completion engine, reducing movements by the dominant hand.

Menu navigation

BiSketch uses BiPad *Chords* to navigate a tool menu. First-level items, e.g. `color` or `stroke`, appear in the BiPad zone. The user chooses a tool and holds down the corresponding finger in the BiPad zone to trigger the next menu level. The user can then use another finger to select the desired option, e.g., `color` then `red`. Chords can trigger frequently used tools or options while drawing with the dominant hand.

Spatial multiplexing

The previous example refers to two-handed interactions based on temporal multiplexing. BiPad can also handle spatially multiplexed tasks. BiMap (Fig. 1c) lets users zoom in and out by pressing buttons with the support hand. They can select part of the map larger than the view port by (i) selecting with the dominant hand; (ii) simultaneously controlling the zoom factor with the non-dominant hand; and (iii) continuing to change the selection with the dominant hand.

EXPERIMENT

We ran a controlled experiment to determine whether BiPad bimanual interaction techniques outperform a common one-handed technique. We also wanted to see if the BiTouch kinematic chain analysis successfully identifies which bimanual holds are most comfortable and efficient.

We asked participants to stand while holding a multi-touch tablet, using one of the holds identified in the preliminary study. We then asked them to perform a series of bimanual *Taps*, *Gestures* or *Chords*, using the thumb or fingers of the non-dominant support hand to modify the actions of the dominant hand. The key research questions were:

Q1 Are two-handed BiPad techniques faster than a similar one-handed technique?

Q2 What are the trade-offs among the different bimanual holds, orientations and interaction techniques?

Participants. Nine men, three women, all right-handed, aged 22-35. Six own a touch-screen phone, one owns a tablet PC.

Apparatus. iPad1 (display: 9.7" , weight: 680g, dimensions: $190 \times 243 \times 13$ mm), running BiPad.

Procedure. We conducted a $[2 \times 5 \times 3]$ within-subjects design with three factors: ORIENTATION (portrait, landscape), HOLD ($F_{side}, F_{top}, T_{bottom}, T_{corner}, T_{side}$), corresponding to the five BiPad interaction zones, and TECHNIQUE (tap, chord, gesture), i.e. 30 unique conditions, plus the no-BiPad control, a standard one-handed task. We discarded eight conditions as impossible or impractical:

Chords can only be performed with the F_{side} and F_{top} HOLD (both *Orientations*) since a single thumb cannot perform multi-finger interactions.

Gestures were omitted from the F_{side} and F_{top} landscape conditions, since the short edge of the tablet cannot be held steadily on the forearm.

Trials were organized into blocks of 6 trials according to TECHNIQUE, ORIENTATION, and HOLD. Participants were asked to stand and support the tablet with a specified hold. In each trial, the participant touched four successive 80-pixel circular targets with the index finger of the dominant hand while holding the tablet with the non-dominant hand. Targets were arranged randomly around the center of the screen. The first target of a series was always green and one randomly chosen target of the following three targets was red. When the red target appeared, the participant was instructed to use the specified technique to turn the target from red back to green before touching it with the dominant hand.

The four techniques for changing red targets to green include the three BiPad techniques: *Tap*, *Chord*, *Gesture*, and the no-BiPad control condition. The three chords use the index finger and one or both of the remaining fingers of the support hand (middle or ring finger). Gestures slide toward the center of the screen, except for T_{corner}, where the thumb slides up-down, down-up or diagonally. In the no-BiPad control condition, the user touches a button at the bottom of the screen with the dominant hand. The task was chosen to support both pointing and bimanual interaction, including mode switches and quasi-modes.

Participants began with the unimanual no-BiPad control condition, followed by the bimanual BiPad conditions (ORIENTATION, HOLD, TECHNIQUE) counter-balanced across subjects using a Latin square. Although this simplifies the experimental design, it does not account for potential order effects between unimanual and bimanual conditions. On the other hand, all of today's tablets are one-handed and it is unlikely that performing a bimanual task prior to a unimanual one would improve performance on the latter. Indeed, the more likely effect would be a drop in performance due to fatigue. To ensure that participants were familiar with the basic task and both conditions, we asked them to perform a three-trial practice block in portrait mode prior to each no-BiPad condition and to each TECHNIQUE×HOLD condition. They were also allowed to perform a one-trial recall prior to each TECHNIQUE×ORIENTATIONS×HOLD so they the could find a comfortable position for the assigned hold.

To begin an experimental BiPad block, participants touched the specified BiPad zone to register the support hand. Participants were asked to maintain this hold throughout the block and perform each task as quickly as possible. At the end of each condition, they evaluated how comfortable it was to interact with the support hand using that hold. Each session lasted approximately 45 minutes.

In summary, we presented two orientations for no-BiPad, all 10 holds for bimanual taps, eight for bimanual gestures (no landscape thumb holds) and four for bimanual chords (fingers only). We thus collected 216 trials per participant:

- 6 replications of the *no-BiPad* control condition in both ORIENTATIONS (landscape, portrait): 12 trials;
- 6 replications of the *Tap* technique in all HOLD and ORIENTATION conditions: 60 trials;
- 6 replications of the three *Chord* techniques in both ORIENTATIONS for finger-based HOLDS (F_{side}, F_{top}): 72 trials;
- 6 replications of each of the three *Gesture* techniques:
  - two-finger-based HOLDS (F_{side}, F_{top}) in portrait ORIENTATION: 12 trials;
  - two thumb-based HOLDS (T_{bottom}, T_{side}) in both ORIENTATIONS: 24 trials;
  - one thumb-based HOLD (T_{corner}) in both ORIENTATIONS: 36 trials.

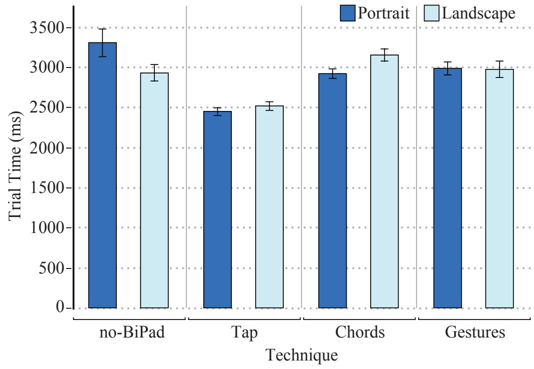


Figure 7. Mean *Trial Time* for each TECHNIQUE by ORIENTATION.

Data Collection. We videotaped each trial and recorded three temporal measures: (i) trial time: from the appearance of the first target to final target selection; (ii) BiPad reaction time: from the appearance of the red target to the first touch in the BiPad area; and (iii) BiPad completion time: from the appearance of the red target to the successful execution of the BiPad interaction. Comfort ratings used a 5-point Likert scale (1 = very uncomfortable; 5 = very comfortable).

RESULTS

We conducted a full factorial ANOVA and handled 'participant' as a random variable, using the standard repeated measures REML technique from the JMP statistical package.

Q1: Bimanual BiPad vs. one-handed interaction

We compared the mean trial time of BiPad techniques to the no-BiPad control condition, using the TECHNIQUE×ORIENTATION ×Random(PARTICIPANT) ANOVA model. We found a significant effect for TECHNIQUE ($F_{3,33} = 16.16$, $p < 0.0001$) but no effect for ORIENTATION ($F_{1,11} = 0.30$, $p = 0.60$). However, we did find a significant interaction effect between TECHNIQUE and ORIENTATION ($F_{3,33} = 8.23$, $p = 0.0003$).

This can be explained by the faster performance in landscape mode for the one-handed no-BiPad condition (Fig. 7): participants performed 11.4% faster ($F_{1,11} = 4.6$, $p = 0.04$) because the distance to reach the button is shorter. Thus, while bimanual taps are significantly faster than the control condition for both orientations (25.9% in portrait and 14% in landscape), bimanual gestures and chords are only significantly faster than no-BiPad in portrait mode (10.4% and 11.7% resp.). In landscape mode, the differences between no-BiPad and bimanual gestures and chords are not significant.

Bimanual taps are significantly faster than bimanual gestures and chords in both device orientations (17.3% and 16.1% in portrait, 14.7% and 19.7% in landscape). Participants significantly preferred bimanual taps (3.5) over bimanual chords (3.3) and gestures (2.7) ($F_{2,22} = 17.5$, $p < 0.0001$). Overall, BiPad techniques were more efficient than the one-handed technique we compared them with.

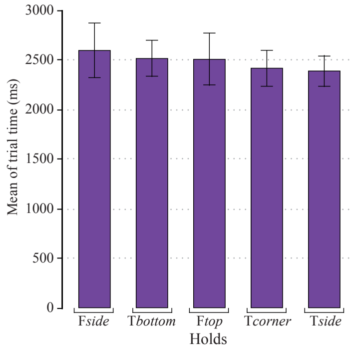


Figure 8. *Tap* performance according to HOLD.

Q2: BiPad tradeoffs: HOLD×ORIENTATION by TECHNIQUE

BiPad Taps

We ran an ANOVA with the model HOLD×ORIENTATION× Random(PARTICIPANT) on trial time for BiPad taps. We found significant effects for HOLD and ORIENTATION ($F_{4,44} = 3.10$, $p = 0.02$ and $F_{1,11} = 5.37$, $p = 0.04$) and no interaction effect ($F_{4,44} = 0.65$, $p = 0.63$).

For HOLD, Tukey post-hoc tests revealed only one significant result: placing the fingers on the right is slower than placing the thumb on the left side of the tablet for right-handed participants (see Fig. 8). For ORIENTATION, a Student's t-test reveals that portrait is significantly faster (LSM $= 2447.31\,ms$) than landscape (LSM $= 2515.99\,ms$).

Performance among bimanual taps is very similar across conditions, making them suitable for all ten holds. The only significant difference is between fingers and thumbs with a side hold. However, although the F_{side} hold is slightly slower, participants preferred it to the T_{side} hold: fingers are more stable than thumbs and cause less fatigue.

BiPad Gestures

As we discarded the two bimanual holds with fingers placed on the right and top of the device in landscape mode, we examined trial Time for each ORIENTATION condition separately for the remaining eight holds. HOLD has a significant effect on the performance time in both portrait ($F_{4,44} = 4.14$, $p = 0.01$) and landscape ($F_{2,22} = 4.75$, $p = 0.02$).

In *Portrait*, post-hoc Tukey HSD tests show that, for a right-handed user, performing gestures with the fingers on the right side of the device is significantly slower than with the thumb on the left side (Fig. 9a). Participants preferred performing gestures with the fingers or with the thumb on the side of the device. In fact, gestures are most difficult to perform when the support hand is placed on the top or bottom of the device when held in portrait mode.

In landscape, where only the *Thumb* placements were tested, performing gestures while supporting the tablet with the thumb on the bottom of the device is significantly faster than in the corner (Fig. 9b). However, since gestures were performed in both ORIENTATION conditions with the thumb, we also compared performance according to thumb holds in both orientation conditions (HOLD×ORIENTATION×Random(PARTICIPANT)).

We found no significant effect of HOLD and ORIENTATION but a significant interaction effect for HOLD×ORIENTATION ($F_{2,22} = 15.08$, $p < 0.0001$). This is because performing gestures with the thumb is significantly faster in portrait, when the support hand is on the side, but significantly slower when the thumb is on the bottom, in which case landscape is faster. The difference between orientations is not significant when the thumb is placed in the corner (Fig. 9c).

The latter effect is interesting and can be explained by the principle of a lever. The greater the distance between the balance point and the most distal support link, the heavier the tablet is perceived. This is considered less comfortable and users find it more difficult to perform gestures. The exception is when the thumb is in the corner: the distal point of the support is equally close to the tablet's balance point in both orientations, thus the two holds are not significantly different. This explanation correlates with the participants' comfort ratings and comments. They preferred to perform gestures with the thumb on the side in portrait and on the bottom in landscape but had no preference for orientation when the thumb is in the corner. Compared to other BiPad techniques, however, gestures were perceived as relatively uncomfortable and practical only for rapid or occasional use.

BiPad Chords

We ran an ANOVA with the model HOLD×ORIENTATION×CHORD TYPE×Random(PARTICIPANT) on *Trial Time*. We found no significant effects of HOLD and ORIENTATION and no interaction effects. For CHORD TYPE, we found a significant effect ($F_{2,22} = 9.09$, $p = 0.01$): holding the index finger down together with the middle finger is significantly faster ($2875\,ms$) than holding down three fingers ($3095\,ms$) or the index and ring finger together ($3131\,ms$).

Participants did not express any significant comfort preferences with respect to chords. However, some participants reported that chords are difficult to perform at the top of the device, especially in landscape mode, due to tension in the arm. Two users could only perform two-finger chords since their third finger could not easily reach the screen.

DISCUSSION

Our results demonstrate not only that hand-held touch tablets can support bimanual interaction, but that it outperforms all tested uni-manual interactions in almost all of our experimental conditions. We created a set of 22 bimanual interaction techniques that combine the ten holds identified in the preliminary study with bimanual taps (10), chords (4) and gestures (8). These offer users trade-offs in performance, comfort and expressive power; BiPad lets users transition smoothly among them.

In the future, we hope to develop the predictive power of the BiTouch design space, building upon our existing understanding of the physical characteristics of the human body and exploring its relationship to hand-held interactive devices. For example, we observed that bimanual taps (in both orientations) and bimanual gestures (in Portrait mode) are significantly faster in holds with thumbs on the side (T_{side}) compared to holds with fingers on the side (F_{side}).

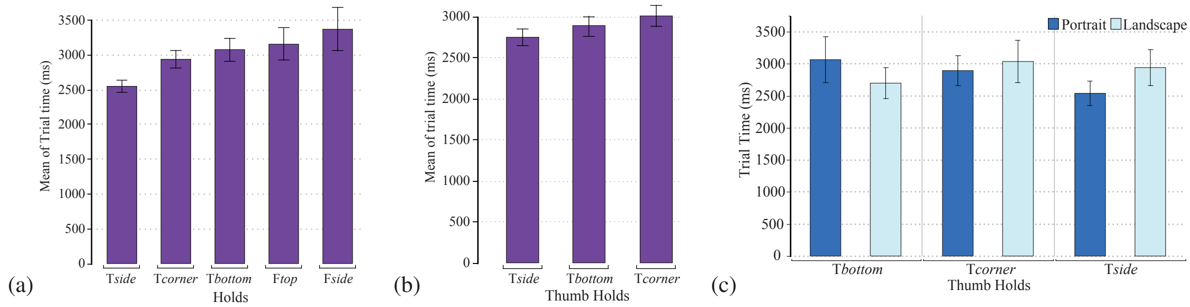


Figure 9. *Gesture* performance according to HOLD (a) in *Portrait*, (b) in landscape, and (c) for the *Thumb* according to HOLD and ORIENTATION.

In contrast, T_{side} is perceived as less comfortable than F_{side}. If we examine thumbs and fingers, we see that the T_{side} hold leaves only two joints available for interaction, whereas the F_{side} hold has three. This suggests that, all other things being equal, performance will be better with interaction techniques that offer a wider range of movement. Additional research is necessary to verify if this prediction obtains for other holds.

We can also use the BiTouch design space to help us understand differences in perceived comfort. One hypothesis is that comfort is correlated with perceived weight, which is determined by both the location of support in the kinematic chain and the orientation of the tablet. If we examine the two holds, we see that the support link for the F_{side} hold, the forearm, is longer than that for the T_{side}, the palm. On the other hand, the former hold restricts movement more than the latter. This suggests two open research questions:

1. Does performance decrease and comfort increase with longer support links?
2. Does performance decrease and comfort increase with increased support link mobility?

We also observed a major effect of tablet orientation in some conditions, such as bimanual gestures. The previously mentioned lever effect plays a role here. If we view the tablet as an extension of the support link, we can estimate its perceived weight based on the distance from the most distal element of the support link to the balance point of the tablet. This raises the question:

3. Do performance and comfort increase as the distance to the balance point decreases?

Finally, multitouch tablets exist in a variety of different shapes, sizes, and weights. We used the popular iPad1 for the first experiment. However, when the iPad2 was released, we replicated the experiment with six participants, and found no significant differences despite the 30% reduction in weight. Of course different tablet designs might affect the performance and comfort of BiPad bimanual interaction. In the future, we plan to extend the BiTouch design space to include device-specific characteristics to increase its predictive power.

SUMMARY AND CONCLUSIONS

We investigated how to introduce effective bimanual interaction into hand-held tablets. We began with a preliminary study that identified support positions while sitting, standing and walking. We found that, although novices found it difficult to come up with effective holds, more experienced users produced ten unique holds that can be adapted to support bimanual interaction. We also found that users do not seek a single, optimal hold, but instead prefer to modify their holds over time, to reduce fatigue and increase comfort. We concluded that the design challenge was not to create a single bimanual technique but rather to create a set of equally comfortable and effective techniques.

We next examined the theoretical basis of the ten observed holds and presented the BiTouch design space, based on Guiard's kinematic chain model. We argue that we can understand bimanual interaction with hand-held devices by examining how three functions – framing, support and interaction – are distributed along the kinematic chain. Our goal is to offer *descriptive*, *predictive* and *generative* power, and BiTouch offers a good start: we can describe all of the unimanual and bimanual interaction techniques observed in the preliminary study; we can make informal predictions about which factors affect performance, comfort and expressive power; and we have generated a set of bimanual interaction techniques that offer different trade-offs with respect to the above:

- *Bimanual Taps*: one finger or thumb taps the screen,
- *Bimanual Chords*: several fingers touch the screen,
- *Bimanual Gestures*: a finger or thumb slides on the screen.

We implemented these techniques in BiPad, a user interface toolkit we made for designing bimanual interaction with off-the-shelf hand-held tablets[1], and developed three working applications in which the non-dominant hand can modify the dominant hand's interaction using taps, chords or gestures.

We tested these interaction techniques in a controlled experiment for each of the five holds and two orientations found in the preliminary study. Bimanual taps are faster than reaching on-screen buttons with the dominant hand only, regardless of tablet orientation or hold. However, they can handle at most three buttons, since the pinky cannot reach the screen and the range of thumb movement is limited. Bimanual chords and gestures offer a richer vocabulary for shortcuts to off-screen functions, but have their own limitations. Chords require multiple fingers and gestures are restricted in landscape to thumb

[1]The BiPad toolkit is freely available at http://insitu.lri.fr/bipad

holds. The BiTouch analysis helps explain why bimanual chords and gestures are faster only in portrait orientation: the position of the support link in the kinematic chain directly affects which fingers or thumbs are available for interaction and the number of available degrees of freedom.

Together, the BiTouch design space and the BiPad toolkit offer developers a richer understanding of bimanual interaction and a practical approach for adding bimanual interaction to hand-held tablets. Future work will explore how we can generate new possibilities for bimanual interaction on a range of devices in different mobile settings.

ACKNOWLEDGMENTS

We wish to thank Michel Beaudouin-Lafon for many fruitful discussions and suggestions for improving this article. We also thank the anonymous reviewers for their helpful comments, as well as the participants for their time and effort.

REFERENCES

1. Balakrishnan, R., and Hinckley, K. The role of kinesthetic reference frames in two-handed input performance. In *UIST '99*, ACM (1999), 171–178.
2. Balakrishnan, R., and Hinckley, K. Symmetric bimanual interaction. In *CHI '00*, ACM (2000), 33–40.
3. Beaudouin-Lafon, M., Mackay, W. E., Andersen, P., Janecek, P., Jensen, M., Lassen, M., Lund, K., Mortensen, K., Munck, S., Ravn, K., Ratzer, A., Christensen, S., and Jensen, K. Cpn/tools: revisiting the desktop metaphor with post-wimp interaction techniques. In *CHI '01*, ACM (2001), 11–12.
4. Benko, H., Wilson, A. D., and Baudisch, P. Precise selection techniques for multi-touch screens. In *CHI '06*, ACM (2006), 1263–1272.
5. Bier, E. A., Stone, M. C., Pier, K., Buxton, W., and DeRose, T. D. Toolglass and magic lenses: the see-through interface. In *SIGGRAPH '93*, ACM (1993), 73–80.
6. Brandl, P., Forlines, C., Wigdor, D., Haller, M., and Shen, C. Combining and measuring the benefits of bimanual pen and direct-touch interaction on horizontal interfaces. In *AVI '08*, ACM (2008), 154–161.
7. Fonville, A., Choe, E. K., Oldham, S., and Kientz, J. A. Exploring the use of technology in healthcare spaces and its impact on empathic communication. In *IHI '10*, ACM (2010), 497–501.
8. Froehlich, J., Wobbrock, J. O., and Kane, S. K. Barrier pointing: using physical edges to assist target acquisition on mobile device touch screens. In *Assets '07*, Assets '07, ACM (2007), 19–26.
9. Guiard, Y. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *J. Motor. Behav. 19* (1987), 486–517.
10. Hinckley, K., Pausch, R., Proffitt, D., and Kassell, N. F. Two-handed virtual manipulation. *ACM TOCHI 5* (1998), 260–302.
11. Huot, S., and Lecolinet, E. Focus+context visualization techniques for displaying large lists with multiple points of interest on small tactile screens. In *INTERACT'07*, Springer-Verlag (2007), 219–233.
12. Kabbash, P., Buxton, W., and Sellen, A. Two-handed input in a compound task. In *CHI '94*, ACM (1994), 417–423.
13. Karlson, A. K., and Bederson, B. B. One-handed touchscreen input for legacy applications. In *CHI '08*, ACM (2008), 1399–1408.
14. Kin, K., Agrawala, M., and DeRose, T. Determining the benefits of direct-touch, bimanual, and multifinger input on a multitouch workstation. In *GI '09*, Canadian Information Processing Society (2009), 119–124.
15. Latulipe, C., Mann, S., Kaplan, C. S., and Clarke, C. L. A. symspline: symmetric two-handed spline manipulation. In *CHI '06*, ACM (2006), 349–358.
16. Leganchuk, A., Zhai, S., and Buxton, W. Manual and cognitive benefits of two-handed input: an experimental study. *ACM TOCHI 5* (1998), 326–359.
17. Mackay, W. E. Which interaction technique works when?: floating palettes, marking menus and toolglasses support different task strategies. In *AVI '02*, ACM (2002), 203–208.
18. Moscovich, T., and Hughes, J. F. Indirect mappings of multi-touch input using one and two hands. In *CHI '08*, ACM (2008), 1275–1284.
19. Oulasvirta, A., and Bergstrom-Lehtovirta, J. Ease of juggling: studying the effects of manual multitasking. In *Proc.*, CHI '11, ACM (2011), 3103–3112.
20. Pascoe, J., Ryan, N., and Morse, D. Using while moving: Hci issues in fieldwork environments. *ACM TOCHI 7* (2000), 417–437.
21. Roth, V., and Turner, T. Bezel swipe: conflict-free scrolling and multiple selection on mobile touch screen devices. In *CHI '09*, ACM (2009), 1523–1526.
22. Roudaut, A., Huot, S., and Lecolinet, E. Taptap and magstick: improving one-handed target acquisition on small touch-screens. In *AVI '08*, ACM (2008), 146–153.
23. Schwesig, C., Poupyrev, I., and Mori, E. Gummi: a bendable computer. In *CHI '04*, ACM (2004), 263–270.
24. Scott, J., Izadi, S., Rezai, L. S., Ruszkowski, D., Bi, X., and Balakrishnan, R. Reartype: text entry using keys on the back of a device. In *MobileHCI '10*, ACM (2010), 171–180.
25. Sugimoto, M., and Hiroki, K. Hybridtouch: an intuitive manipulation technique for pdas using their front and rear surfaces. In *MobileHCI '06*, ACM (2006), 137–140.
26. Wigdor, D., Forlines, C., Baudisch, P., Barnwell, J., and Shen, C. Lucid touch: a see-through mobile device. In *UIST '07*, ACM (2007), 269–278.
27. Wimmer, R., and Boring, S. Handsense: discriminating different ways of grasping and holding a tangible user interface. In *TEI '09*, ACM (2009), 359–362.
28. Wobbrock, J. O., Myers, B. A., and Aung, H. H. The performance of hand postures in front- and back-of-device interaction for mobile computing. *Int. J. Hum.-Comput. Stud. 66* (2008), 857–875.
29. Wu, M., and Balakrishnan, R. Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. In *UIST '03*, ACM (2003), 193–202.
30. Yang, X.-D., Mak, E., Irani, P., and Bischof, W. F. Dual-surface input: augmenting one-handed interaction with coordinated front and behind-the-screen input. In *MobileHCI '09*, ACM (2009), 5:1–5:10.

A Body-centric Design Space for Multi-surface Interaction

ABSTRACT

We introduce *BodyScape*, a body-centric design space for both analyzing existing multi-surface interaction techniques and suggesting new ones. We examine the relationship between users and their environment, specifically how different body parts enhance or restrict movement in particular interaction techniques. We illustrate the use of BodyScape by comparing two free-hand techniques, *on-body touch* and *mid-air pointing*, separately and in combination. We found that touching the torso is faster than touching the lower legs, since it affects the user's balance; individual techniques outperform compound ones; and touching the dominant arm is slower than other body parts because the user must compensate for the applied force. The latter is surprising, given that most recent on-body touch techniques focus on touching the dominant arm.

Author Keywords

Multi-surface interaction, Body-centric design space

ACM Classification Keywords

H.5.2. Information Interfaces and Presentation: User Interfaces.: Graphical user interfaces

INTRODUCTION

Multi-surface environments encourage users to interact while standing or walking, using their hands to manipulate objects on multiple displays. Klemmer et al. [16] argue that using the body enhances both learning and reasoning and this interaction paradigm has proven effective for gaming [27], in immersive environments [21], when controlling multimedia dance performances [18] and even for skilled, hands-free tasks such as surgery [30].

Smartphones and devices such as Nintendo's Wii permit such interaction via a hand-held device, allowing sophisticated control. However, holding a device is tiring [22] and limits the range of gestures for communicating with co-located users, with a corresponding negative impact on thought, understanding, and creativity [10]. Krueger's VIDEOPLACE [17] pioneered a new form of whole-body interaction in which users stand or walk while pointing to a wall-sized display. Today, off-the-shelf devices like Sony's Eyetoy and Microsoft's Kinect let users interact by pointing or moving their bodies, although most interaction involves basic pointing or drawing. Most research in complex device-free interaction focuses on hand gestures, e.g. Charade's [2] vocabulary of hand-shapes that distinguish between "natural" and explicitly learned hand positions, or touching the fore-arm, e.g. Skinput's [13] use of bio-acoustic signals or PUB's [20] ultrasonic signals.

However, the human body offers a number of potential targets that vary in size, access, physical comfort, and social acceptance. We are interested in exploring these targets to create more sophisticated body-centric techniques, sometimes in conjunction with hand-held devices, to interact with complex data in multi-surface environments. Advances in sensor and actuator technologies have produced a combinatorial explosion of options, yet, with few exceptions [26, 22], we lack clear guidelines on how to combine them in a coherent, powerful way. We argue that taking a *body-centric* approach, with a focus on the sensory and motor capabilities of human beings, will help restrict the range of possibilities in a form manageable for an interaction designer.

This paper introduces *BodyScape*, a design space that classifies body-centric interaction techniques with respect to multiple surfaces according to input and output location relative to the user. We describe an experiment that illustrates how to use the design space to investigate atomic and compound body-centric interaction techniques, in this case, compound mid-air interaction techniques that involve pointing on large displays to designate the focus or target(s) of a command. Combining on-body touch with the non-dominant hand and mid-air pointing with the dominant hand is appealing for interacting with large displays: both inputs are always available without requiring hand-held devices. However, combining them into a single, compound action may result in unwanted interaction effects. We report the results of our experiment and conclude with a set of design guidelines for placing targets on the human body depending on simultaneous body movements.

BODYSCAPE DESIGN SPACE & RELATED WORK

Multi-surface environments (MSEs) require users to be "physically" engaged in the interaction and afford physical actions like pointing to a distant object with the hand or walking towards a large display to see more details [3]. The body-centric paradigm is well-adapted to device- or eyes-free interaction techniques because they account for the role of the body in the interactive environment. However, few studies and designs take this approach, and most of those focus on large displays [19, 26, 22].

Today's off-the-shelf technology is capable of tracking both the human body and its environment [14]. Recent research prototypes also permit direct interaction on the user's body [13, 20] or clothes [15]. These technologies and interaction techniques suggest new types of body-centric interaction, but it remains unclear how they will work in conjunction with more conventional techniques; particularly from the user's perspective.

Although the field includes a number of isolated point designs, we lack a higher-level framework that characterizes how users coordinate their movements with, around and among multiple devices in a multi-surface environment. Our goal is to define a more general approach to body-centric interaction, and propose a design space that: (i) assesses their adequacy to a new environment or context of use; (ii) informs further design of novel body-centric interaction techniques.

We are aware of only three design spaces that explicitly account for the body during interaction. One focuses on the interaction space of mobile devices [6] and another offers a task-oriented analysis of mixed-reality systems [23]. Both consider of *proximity* to the user's body but neither fully captures the distributed nature of multi-surface environments. We are most influenced by Shoemaker et al.'s [26] pioneering work, which introduced high-level design principles and guidelines for body-centric interaction on large displays.

BodyScape

BodyScape builds upon a morphological analysis [5] focusing on (i) the relationships between the user's body and the interactive environment; (ii) the involvement of the user's body during the interaction, i.e., which of the user's body parts are involved or affected while performing an interaction technique; and (iii) the combination of "atomic" interaction techniques in order to manage the complexity of MSEs. These elements can help us identify appropriate or adverse designs for a given task, as well as the impact they could have on user experience and performance, e.g. because of body movement conflicts or restrictions.

Relationships Between the Body and the Environment

Multi-surface environments distribute user input and system visual output[1] on multiple devices (screens, tactile surfaces, handheld devices, tracking systems, on-body sensors, etc.). The relative location and body positions of the user thus play a central role in the interactions she can perform. For instance, touching a tactile surface while looking at a screen on your back is obviously awkward. This physical separation defines the two first dimensions of BodyScape: *User Input* and *System Visual Output*. Using a body-centric perspective, similar to [6] and [23], we identify two possible cases for input and output: *Relative to the body* and *Fixed in the world*.

System Visual Output – Multi-surface environments are inevitably affected by the separation of the visual outputs [28, 29] which is divided across multiple devices. It requires users to move their gaze and switch their attention onto the output devices that are relevant to their current task by *turning the head* and – if that is not sufficient – *turning the torso*, the *entire body*, or even walking. Visual output *relative to the body* is independent of the user's location in the environment, e.g. the screen of a hand-held device. It does not constrain the user's location and posture (except for the limb that may hold a device). Conversely, visual output *fixed in the world* requires the user's head to be oriented towards its physical location, e.g on a wall projector. When outputs are fixed in the world, users' locations and body configurations are restricted to the positions that allows them to see the visual output effectively.

User Input – Input *relative to the body* can be performed at an arbitrary location in the environment, e.g. on-body touch, whereas input *fixed in the world* requires a specific user location, e.g. next to a tabletop. This impacts the configuration of the body, e.g., the user can carry a handheld device and freely interact with it from anywhere. Some technologies even make an input device unnecessary: PUB [20] enables on-body touch interaction and PinStripe [15] detects pinching and rolling gestures on the users' clothes. However, when a hand-held device or a limb is tracked in 3D, e.g., mid-air pointing on a display, the arm needs to be maintained in a specific position relative to a fixed target.

Our design space also differentiates *Mid-air* and *Touch*-based user input since it can affect performance and possible body restrictions. Body movements and their coordination depends on the physical connection with the environment [9]. For example, previous studies demonstrated that pan-and-zoom techniques for large displays are faster using touch than with mid-air gestures, due to the additional guidance on input movements [22]. In MSEs, having to touch an interactive device could add additional constraints, e.g. walking to an interactive tabletop.

Body Restriction in the Environment – The Input and Visual Output dimensions of BodyScape define a qualitative scale of the restrictions that a given interaction technique will impose on the user's body (see the horizontal axis in Fig. 2). This is an indicative measure of the whole body's degrees of freedom (translation and rotation) remaining for other interactions or body movements. *Body Restriction* is not necessarily negative: e.g., assigning a dedicated fixed display area for each user in a collaborative MSE, thus restricting their operating area, could prevent some issues similar to those encountered on an interactive table [25] (e.g., visual occlusions, collisions, conflicts or privacy concerns).

As illustrated on the horizontal axis of Fig. 2, constraints from the *Input* dimension have more impact on body restriction than those from *Visual Output*, and *Touch* is more constraining than *Mid-air* when the input is fixed in the world: watching a fixed display can still be done at a distance and from different angles while some input devices, e.g. tactile tables, require physical proximity.

Body Involvement The third dimension of BodyScape adresses body restrictions at a finer level of granularity by considering which part of the user's body are *involved* in an interaction technique. In fact, every interaction technique involves the body in varying degrees of freedom, from simple thumb gestures on a handheld device [22], to whole-body movements [18]. We define a group of limbs involved in a technique as *involved body parts*, that could be the *dominant arm*, the *non-dominant arm*, the *dominant leg* or the *non-dominant leg*. For instance, a mid-air pointing technique involves the dominant arm.

[1] We do not consider auditory feedback since sound perception does not depend on the body position in most cases, except in rare environments featuring finely tuned spatial audio.

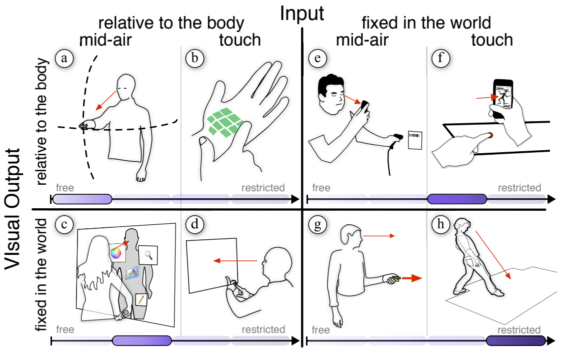


Figure 1. Atomic body-centric interaction techniques in BodyScape according to their User Input and System Output characteristics: a) Virtual Shelves [19] ; b) Skinput [13] ; c) Body-centric interaction techniques for wall-sized displays [26] ; d) PalmRC [8] ; e) Scanning an object with feedback on a mobile device ; f) Pick and Drop [24] ; g) Mid-air pointing ; h) Multitoe [1] .

Some techniques not only involves groups of limbs but also *affect* some others. For instance, on-body touch interaction involves one hand (and the corresponding arm) into the interaction. But the body limbs that are touched with the hand are *affected body parts*. This has implications for further body restrictions, since the affected body parts are unlikely to be involved into the interaction and vice versa (e.g., the dominant arm cannot touch the dominant forearm). These restrictions would be even more critical to account for when we will later consider the combination of interaction techniques in multi-surface environments. We define five groups of *involved body parts*, that could be the *dominant arm*, the *non-dominant arm*, the *dominant leg*, the *non-dominant leg* or the *torso*.

We do not consider the user's head in involved and affected body parts, since its movements are restricted by the constraints of the visual output dimension. While some recent work use the orientation of the head to improve interaction [7] on large displays, it is based on a "passive" input mode where the system is adapting itself to the head orientation, which primary function is still to orient the gaze.

Classification of Body-centric Interaction Techniques

Figures 1 and 2 give graphical representations of BodyScape, and illustrate the properties of several body-centric interaction techniques. Figure 1 presents these techniques along the *Input* and *Visual Output* dimensions, giving insights into their impact on body restrictions in the environment. These are only atomic interaction techniques, that allow to perform elementary actions like moving a cursor or selecting a target. Figure 2 presents the same atomic techniques along the *Body Restriction in the Environment* scale (resulting from their Input and Visual Output characteristics) and the total number of involved and affected body parts, the *Body Involvement* dimension. Some compound interaction techniques that will be discussed later are also presented in this figure.

Relative Input / Relative Output The less restricted combination is obviously when both input and output are relative to

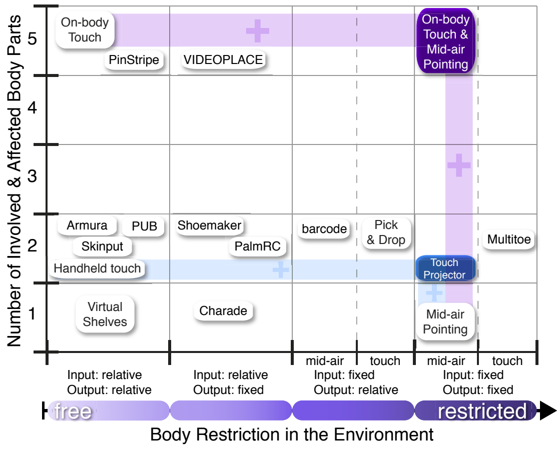


Figure 2. Body-centric interaction techniques in BodyScape, according to their amount of *Body Restriction in the Environment* and number of *Involved and Affected* limbs. Atomic techniques are on a white background. Compound techniques are on a colored background, linked to the corresponding atomic techniques.

the body, since the user can move freely in the environment while still being able to interact and get visual feedback.

VirtualShelf [19] gives access to short-cuts on a mobile phone by orienting the device in mid-air within a spherical space in front of the user (Fig.1a), thus involving the dominant arm. *Armura* [12] extends this approach with wearable hardware that detects mid-air gestures of both arms and projects visual feedback on the user's body. Skinput [13] (Fig. 1b) enables touch input on the users' forearm and provides body-relative visual output with a projector mounted on the user's shoulder. The dominant arm is thus involved into the pointing while the non-dominant arm is only affected by the pointing.

Relative Input / Fixed Output Fixing the output in the environment constrains user's orientation and, if the distance to the display matters, her location. Shoemaker introduced body-centric interaction techniques for large displays [26] where the user select tools by pointing towards body parts, e.g. the stomach, and pressing the button of a hand-held device. Only the pointing arm is involved and the user's shadow is displayed on the wall display, indicating the location of the tools on the body. This requires the user to remain in a restricted body configuration in front of the screen (Fig. 1c). PalmRC [8] (Fig. 1d) allows free-hand operations on a TV set. Users press imaginary buttons on their palm [11] and perceive visual feedback on the fixed TV screen. One arm is involved into the interaction while the other is affected.

Fixed Input / Relative Output An input fixed in the world is more constraining since it requires to stand in a defined perimeter that limits movements. In this case, touch is more constraining than mid-air. For example, while limited, the detection range of a Kinect device is less constraining than having to stand at the edge of an interactive table.

A simple example of a mid-air fixed input with a relative output is when a user is scanning a barcode while watching the feedback on a mobile device (Fig. 1e). With touch interactions, these kind of input/output combinations are common when transferring an object from a fixed surface to a mobile device, like in Fig. 1f (Pick and Drop [24]). Both examples involve the dominant arm into the interaction, and affect the non-dominant arm that caries the handheld device.

Fixed Input / Fixed Output In this situation, the location and visual attention of the user are constrained by both the input and the output. This is the most constraining combination, especially with *Touch* inputs.

Mid-air fixed input and output is one of the most common combination for pointing on todays wall-sized display, using the “laser pointer” metaphor. Even if the interaction is performed at a distance, it is fixed in the world since it requires to stand at an appropriate location in order to be able to directly point toward an object on the surface (Fig. 1g). Conventional touch interaction, with a tabletop or a large display, requires to be in front of the surface. This is even more restrictive with Multitoe [1] since it enables visual output and touch interaction on the floor with the feet (Fig. 1h).

Body Involvement Figure 2 shows that most of body-centric techniques are only involving and affecting one or two group of body parts (in general the arms). In our knowledge, only a very few “whole-body” techniques exists, that are involving or affecting the complete set of body parts: PinStripe [15], that enables gestures on the users’ cloth, and VIDEOPLACE [17] (and similar posture-based approaches for gaming or entertainment).

Compound Techniques in MSEs

Performing more complex tasks in MSEs requires to use several interaction techniques (i) in *series*, e.g., selecting an object on a touch surface and then on another one; (ii) or in *parallel*, e.g., touching an object on a fixed surface while simultaneously touching another one on a handheld device.

Serial Combination We define a serial combination as a temporal sequence of interactions techniques. The combined techniques can be interdependent (sharing the same object, or the output of one being the input of the other), but the first action should be ended before to start the second one. It could consist in selecting an object on a tactile surface (touch and release) and then applying a function onto this object with a menu on a mobile device. In our design space, this kind of compound technique does not change the body restrictions that are imposed by each of the atomic techniques in the sequence, nor the body parts they are involving or affecting as well. However, when designing such sequences, one have to consider their characteristics to avoid awkward situations for the user, e.g., moving back and forth in the environment, constantly switching attention between fixed and relative displays, switching a device from one hand to another.

Parallel Combination Parallel combination consists in performing two techniques at the same time. It could consist in touching two tactile surfaces simultaneously in order to transfer an object from one to the other [31]. This has to be considered differently than serial combinations in BodyScape, in order to determine the impact on user’s body restrictions, and potential conflicts between involved and affected body parts.

The body movement restriction of a parallel combination of techniques depends on the more restrictive of the combined techniques: combining fixed in the world with relative to the body will result in fixed in the world. Touchprojector [4] illustrates this well (see Fig. 2). It consists in using a handheld device as a lens to select objects on a distant display: the user orients the device towards the target (mid-air fixed input and fixed output) and simultaneously touches the tactile screen of the device in order to select the it (touch relative input + relative output). Touchprojector is thus considered as a “touch fixed input and fixed output” technique in our design space. The advantage of minimizing body restrictions with a technique relative to the body is thus waived by the requirement of a fixed input. It is however beneficial in the case of Touchprojector, since it enables direct interaction with a display at a distance, preventing to move to the display or to use another interaction device.

AN EMPIRICAL STUDY WITH BODYSCAPE

Our work with users in complex multi-surface environments highlighted the need for interaction techniques that go beyond simple pointing and navigation. Users need to combine techniques as they interact with complex data spread across multiple surfaces. The BodyScape design space suggests a number of possibilities for both atomic and compound interaction techniques that we can now compare and contrast.

This section illustrates how we can use the BodyScape design space to look systematically at different types of body-centric interaction techniques, both in their atomic form and when combined into compound interaction techniques. We chose two techniques, illustrated in Figure 1d, ON-BODY TOUCH input, and 1g, MID-AIR POINTING input, both with visual output on a wall display, which is where our users typically need to interact with their data. Although the latter has been well-studied in the literature ([22]), we know little of the performance and acceptability trade-offs involved in touching one’s own body to control a multi-surface environment. Because it is indirect, we are particularly interested in on-body touch for secondary tasks such as confirming a selection, triggering an action on a specified object, or changing the scope or interpretation of a gesture. We are also interested in how they compare with each other, since MID-AIR POINTING restricts movement more than ON-BODY TOUCH (Fig. 1g vs. 1d), while ON-BODY TOUCH affects more body parts than MID-AIR POINTING (Fig. 2).

Finally, we want to create *compound* interaction techniques, so as to increase the size of the command vocabulary and offer users more nuanced control. However, because this involves coordinating two controlled movements, we need to understand any potential interaction effects. The following experiment investigates the two atomic techniques above, which also act as baselines for comparison with a compound technique that combines them.

1. *Which on-body targets are most efficient and acceptable?* Users can take advantage of proprioception when touching their own bodies, which enables eyes-free interaction and

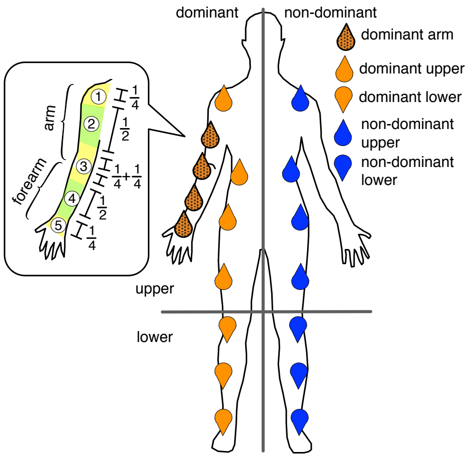


Figure 3. 18 body targets are grouped into five categories.

suggests higher performance. However, body targets differ both in the level of motor control required to reach them, e.g., touching a foot requires more balance than touching a shoulder, and in their social acceptability, e.g., touching below the waist [15].

2. *What performance trade-offs obtain with compound body-centric interaction techniques?* Users must position themselves relative to a target displayed on the wall and stabilize the body to point effectively. Simultaneously selecting on-body targets that force shifts in balance or awkward movements may degrade pointing performance. In addition, smaller targets will decrease pointing performance, but may also decrease On-Body Touch performance.

Method

Participants

We recruited sixteen unpaid right-handed volunteers (13 men, average age 28); five had previous experience using a wall-sized display. All had good to excellent balance (median 4 on a 5-high Likert scale) and practiced at least one activity that requires balance and body control. All wore comfortable, non-restrictive clothing.

Apparatus

Participants stood in front of a wall-sized display consisting of 32 high-resolution 30” LCD displays laid out in an 8×4 matrix (5.5m ×1.8m) with a total of 20480×6400 pixels (100.63 ppi). Participants wore passive infra-red reflective markers that were tracked in three dimensions by ten VICON cameras with sub-millimeter accuracy at a rate of up to 200 Hz. Markers were mounted on a wireless mouse held in the user’s dominant hand[2] to track pointing at a target on the wall, on the index finger of the non-dominant hand to track on-body touches, and on protective sports gear – belt, forearms, shoulders and legs – to track on-body targets. The latter were adjustable to fit over the participants’ clothing.

[2]All subjects were right-handed, so “dominant” refers to the right hand or side.

Based on pilot studies, we defined 18 body target locations distributed across the body (Fig. 3), ranging in size from $9cm$ on the forearm to $16cm$ on the lower limbs, depending upon location and density of nearby targets, grouped as follows:

Dominant Arm: 4 targets on dominant arm (D arm = upper arm, elbow, forearm, wrist)

Dominant Upper Body: 4 targets on dominant side of upper body (D upper = thigh, hip, torso, shoulder)

Non-dominant Upper Body: 4 targets on non-dominant side of upper body (ND upper = thigh, hip, torso, shoulder)

Dominant Lower Leg: 3 targets on dominant side of lower leg (D lower = knee, tibia, foot)

Non-dominant Lower Leg: 3 targets on non-dominant side of lower leg (ND lower = knee, tibia, foot)

In On-Body touch conditions, participants wore an IR tracked glove on the non-dominant hand with a pressure sensor in the index finger. The system made an orthogonal projection from the index finger to the touched limb segment using a skeleton-based model to calculate the closest body target.

Wall pointing tasks varied in difficulty from *easy* (diameter of the circular target was $1200px$ or $30cm$) to *medium* ($850px$ or $21.25cm$) to *hard* ($500px$ or $12.5cm$). Wall targets were randomly placed $4700px$ ($117.5cm$) from the starting target.

Data Collection

We collected timing and error data for each trial, as follows:

Trial Time**:** from trial start to completion.

Pointing reaction time**:** from appearance of on-screen target to cursor displacement of more than $1000px$.

Pointing movement time**:** from initial cursor movement to entry into goal target.

Cursor readjustment time**:** from leaving goal target to relocating cursor onto goal target.

Body reaction time**:** from appearance of trial stimulus to leaving starting position.

Body pointing time**:** from leaving start position to touching on-body target.

Body errors**:** number of incorrect touches detected on body target[3]; includes list of incorrect targets per error.

We debriefed participants at the end of the experiment and asked them to rank on a Likert scale: (i) perceived comfort of each body target according to each Mid-air Pointing condition (‘1=very uncomfortable’ to ‘5=very comfortable’); and (ii) social acceptability of each on-body target:“Would you agree to touch this body target in a work environment with colleagues in the same room?” (‘1=never’ to ‘5=certainly’).

Procedure

Each session lasted about 60 minutes, starting with a training session, followed by blocks of trials of the following conditions, counter-balanced across subjects using a Latin square.

<u>Body only:</u> Non-dominant hand touches one of 18 on-body targets (atomic technique − 18×5 replications = 90 trials)

<u>Pointing only:</u> Dominant hand points to one of three target sizes (atomic technique − 3×5 replications = 15 trials)

[3]Includes both system detection and user errors.

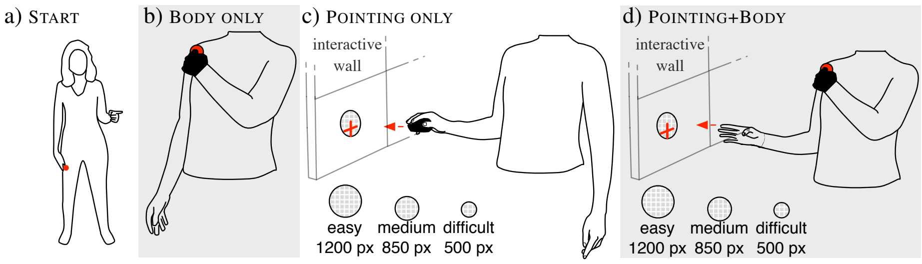


Figure 4. a) Starting position b) BODY ONLY **c)** POINTING ONLY **d)** POINTING+BODY
Starting position: non-dominant hand at the hip and/or dominant hand points to a starting target on the wall display.
BODY ONLY and POINTING ONLY are atomic conditions; POINTING+BODY is compound: a body touch triggers the selected wall target.

POINTING+BODY: Combines touching an on-body target with selecting a wall target (compound technique – $(18\times3)\times5$ replications $= 270$ trials)

Participants were thus exposed to 75 unique conditions, each replicated five times, for a total of 375 trials. BODY ONLY and POINTING+BODY trials were organized into blocks of six, with the location of body targets randomized and no two successive trials involved the same body target group. POINTING ONLY trials were organized into blocks of five and all wall pointing trials were counterbalanced across difficulty. The two atomic interaction techniques, BODY ONLY and POINTING ONLY serve as baseline comparisons for performance with the compound interaction technique, POINTING+BODY. Participants were instructed to perform trials as quickly and accurately as possible.

BODY ONLY (Fig. 4b): The starting position involves standing comfortably facing the wall display, with the non-dominant hand at the thigh (Fig. 4a). The trial begins when a body-target illustration appears on the wall. The participant touches that target with the index finger of the non-dominant hand as quickly and accurately as possible. Participants were asked to avoid crouching or bending their bodies, which forced them to lift their legs to reach lower-leg targets. The trial ends only when the participant successfully selects the correct target; all intermediate incorrect selections are logged.

Figure 5 shows how different body parts interact for different on-body targets. The non-dominant arm is always involved, since it is responsible for pointing at the target. However, some on-body targets also affect other body parts, which may have adverse effects, such as shifting one's balance to touch the foot (Fig. 5c).

POINTING ONLY (Fig. 4c): The starting position involves standing comfortably facing the wall display and using the dominant hand to locate a cursor within a circular target displayed in the center of the wall. The trial begins when the starting target disappears and goal target appears between $0.5s$ and $1s$ later, to reduce anticipatory movements and learning effects. The participant moves the dominant hand to move the cursor to the goal target and selects by pressing the left button of the mouse bearing the optical marker used for pointing. The trial ends only when the participant successfully clicks the mouse button while the cursor is inside the goal target.

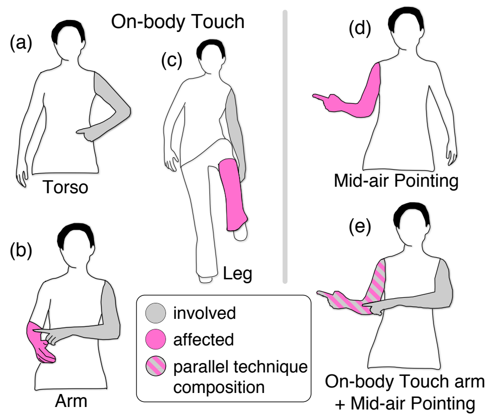


Figure 5. Body parts involved when touching the (a) torso, (b) arm, (c) leg; (d) mid-air pointing; and (e) in parallel, when the dominant hand points in mid-air and non-dominant hand touches the dominant arm.

POINTING+BODY (Fig. 4d): The starting position combines the above, with the non-dominant hand at the thigh and the dominant hand pointing to the starting target on the wall. The trial begins with the appearance of a body-target illustration and the goal target on the wall display. The participant first points the cursor at the goal target, then completes the trial by touching the designated on-body target. The trial ends only when the on-body touch occurs while the cursor is inside the goal target on the wall. As in the BODY ONLY condition, multiple body parts may be involved, sometimes with adverse effects. Fig. 5e shows the interaction between the dominant arm, which is trying to point to a target on the wall and the non-dominant arm, which is pointing at the dominant arm.

Training

Participants began by calibrating the system to their bodies, visually locating, touching and verifying each of the 18 body targets. They were then exposed to three blocks of six BODY ONLY trials, with the requirement that they performed two on-body touches in less than five seconds. They continued with three additional blocks to ensure they could accurately touch each of the targets. Next, they were exposed to all three levels of difficulty for the POINTING ONLY condition: easy, medium and hard, in a single block. Finally, they performed three additional blocks of the compound POINTING+BODY technique.

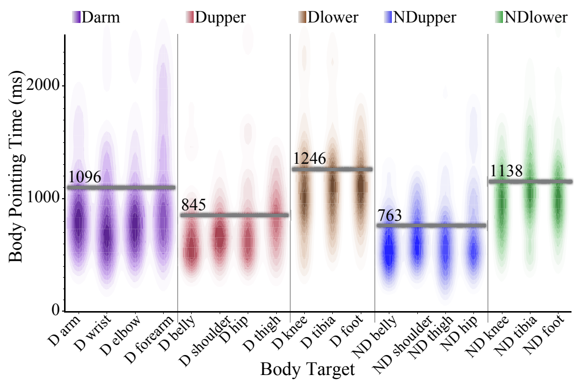


Figure 6. Mean BODY POINTING TIME **is faster for both upper body targets (**D UPPER **and** ND UPPER**) compared to other targets. Horizontal lines indicate group means; performance within groups is consistent.**

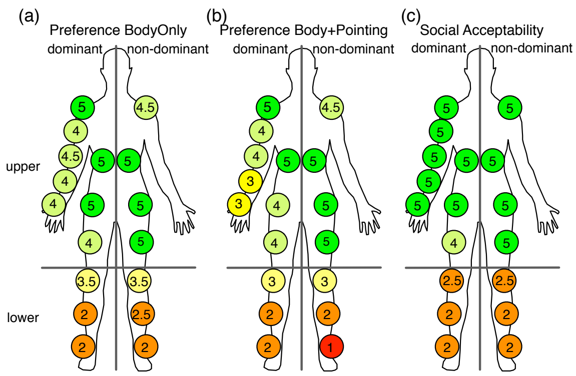


Figure 7. Median preference and acceptability rankings of on-body targets (from green = acceptable to red = unacceptable).

Results

Q1: Efficiency & acceptability of on-body targets

Our first research question asks which on-body targets are most efficient and which are socially acceptable. We conducted a full factorial ANOVA on the BODY ONLY condition, with PARTICIPANT as a random variable based on the standard repeated measures (REML) technique from the JMP 9 statistical package. We found no fatigue or learning effects.

Figure 6 shows the times for touching all 18 on-body targets, grouped into the five body areas. We found significant effects of BODY TARGET on BODY POINTING TIME: touching lower body targets is slower. Since BODY POINTING TIME is consistent for targets within a given target group, we report results according to target group, unless otherwise stated.

Overall, we found a main effect of BODY TARGET GROUP on TRIAL TIME ($F_{4,60} = 21.2$, $p < 0.0001$). A post-hoc Tukey test revealed two significantly different groups: body targets located on the upper torso required less than $1400ms$ to be touched whereas targets on the dominant arm and on the lower body parts required more than $1600ms$. Results are similar for BODY POINTING TIME with a significant effect of BODY TARGET GROUP only for the D UPPER group ($F_{3,45} = 5.07$, $p = 0.004$), specifically, targets on the dominant thigh are touched more slowly than those on the shoulder or torso. For BODY REACTION TIME, despite a significant effect, values are very close for each BODY TARGET GROUP ($530ms \pm 20ms$).

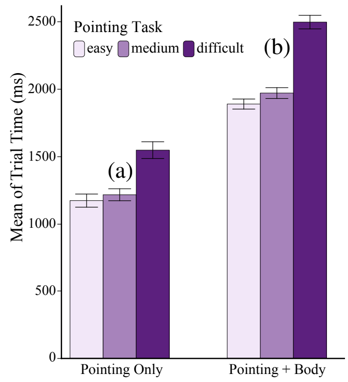


Figure 8. TRIAL TIME **for (a) Pointing Only and (b) Pointing + Body, by pointing difficulty.**

Participants were able to quickly touch on-body targets with an accuracy of 92.4% on the first try. A post-hoc Tukey test showed that targets on the dominant arm were more prone to errors than other body target areas (14.8% vs. 6% for dominant and non-dominant upper body and 2.9% for non-dominant lower body targets). Most errors obtained when targets were close to each other, i.e. when the participant's hand touched the boundary between the goal and a nearby target or when the dominant arm was held close to the torso, making it difficult to distinguish between the torso and arm targets. Touching lower body parts is, not surprisingly, slower, since these targets are further from the starting position and require more complex movements. However, the difference is small, about $200ms$ or 12% of global trial time.

Qualitative measures of Preference and Social Acceptance

Figure 7a shows that participants' preferences (median values of Likert-scale) for and within each BODY TARGET GROUP were consistent with performance measures: targets on the upper body parts were preferred over lower body parts (consistent with [15]) and the torso were slightly more preferred than on the dominant arm.

Interestingly, preferences for non-dominant foot and the dominant arm decrease when on-body touch interaction is combined with mid-air pointing (Fig. 7b). The latter is surprising, given that the most popular location for on-body targets in the literature is on the dominant arm. This suggests that interaction designers should explore alternative on-body targets as well. Social acceptability varies from highly acceptable (upper body) to unacceptable (lower body) (Figure 7c).

Q2: Performance Trade-offs for compound techniques

The second research question examines the effect of combining two atomic interaction techniques, in this case BODY ONLY and POINTING ONLY, into a single compound technique. We treat these atomic techniques as baseline values to help us better evaluate the compound task.

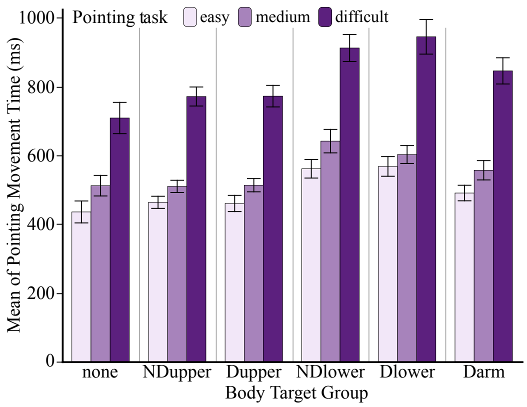


Figure 9. Interaction Pointing×Body on Pointing movement time**.**

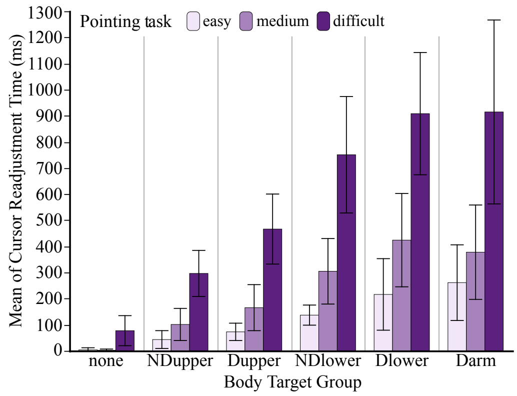


Figure 10. Effect of Pointing difficulty and Body target group **on** Cursor readjustment time**.**

Pointing Only task

Not surprisingly, hard pointing tasks are significantly slower ($1545ms$) (Trial Time ($F_{2,30} = 40.23, p < 0.0001$) than medium ($1216ms$) or easy ($1170ms$) tasks, which are not significantly different from each other (Fig. 8a). Pointing reaction time is also significantly slower for difficult ($498ms$) as opposed to medium ($443ms$) or easy ($456ms$) tasks. Pointing movement time is significantly different for all three levels of difficulty: hard ($708ms$), medium ($511ms$) and easy ($435ms$).

Participants made few errors but occasionally had to relocate the cursor inside the goal target before validating the selection with the mouse. This occurred rarely (1.8% of all trials), but Cursor readjustment time was significantly more likely for difficult pointing tasks (15%) ($F_{2,30} = 8.02$, $p = 0.0016$) and accounts for the differences in Trial Time and Pointing movement time.

Compound Pointing plus Body task

Figure 8b shows that the combined Mid-air Pointing and On-Body touch task is significantly slower than Mid-air Pointing alone for all levels of difficulty. Trial Time is significantly slower for difficult Mid-air Pointing ($2545ms$) than both medium ($1997ms$) and easy ($1905ms$) tasks. In fact, the easiest compound task is significantly slower that the hardest Pointing only task.

Body target group also has an effect on Trial Time ($F_{4,60} = 34.1$, $p < 0.0001$) with the same significant groups as for Body only: Trial Time is significantly faster when touching upper body targets (ND upper $= 1794ms$, D upper $= 1914ms$) than lower body targets (ND lower $= 2267ms$, D lower $= 2368ms$) or the dominant arm (D arm $= 2401ms$). Body reaction time is faster than Pointing reaction time, regardless of pointing difficulty.

Although we can see that the individual techniques are both more efficient than the compound technique, the question is why? Just how does On-Body touch affect Mid-air Pointing? Figure 9 shows interaction effects between the two elements of the compound tasks, by both Body target group and pointing difficulty. While Pointing movement time is close to the pointing baseline for all difficulties when Mid-air Pointing is combined with On-Body touch on the upper body parts, we observe a stronger negative effect for the lower body parts and the dominant arm, especially for difficult pointing tasks.

This impact of On-Body touch on the Mid-air Pointing task does not only relate to the movement phase but also cursor readjustments. For the combined Pointing+Body task, 31% of the trials required the participants to relocate the cursor inside of the target before validating the selection with a body touch, compared to only 6% for Pointing only. Thus, we found significant effects of Mid-air Pointing ($F_{2,30} = 59.64$, $p < 0.0001$), Body target group ($F_{5,75} = 23.03, p < 0.0001$) and Mid-air Pointing×Body target group ($F_{10,150} = 8.45$, $p < 0.0001$) on Cursor readjustment time. As shown in Figure 10, Cursor readjustment time increases significantly for each level of difficulty of Mid-air Pointing but selecting body targets on some Body target group, especially in D lower and D arm, affects the body configuration and requires even more time to relocate the cursor inside of the on-screen target.

This result reveals two important things: (1) touching the dominant arm while pointing affects the precision of pointing and requires “force-balance” (targets on D arm); (2) touching targets on the lower body parts affects the precision of pointing and requires “movement-balance” (targets on ND lower and D lower). Overall, since the impact of both D lower and D arm is similar, we observe that maintaining force-balance is as difficult as maintaining movement-balance during the pointing task, and that the difficulty in movement-balance is not only caused by standing on one leg, but by simultaneously crossing the body’s sagittal plane (difference between D lower and ND lower).

Similarly, we studied the effect of Mid-air Pointing on On-Body touch by performing an ANOVA with the model Mid-air Pointing[none /easy/medium/difficult]×Body target group. We did not find any effect on Body reaction time. On Body pointing time, we did find a significant effect of Body target group ($F_{4,60} = 38.69, p < 0.0001$), of Mid-air Pointing ($F_{3,45} = 78.15$, $p < 0.0001$) and a significant Mid-air Pointing×Body target group interaction ($F_{12,180} = 2.28$, $p = 0.01$). The

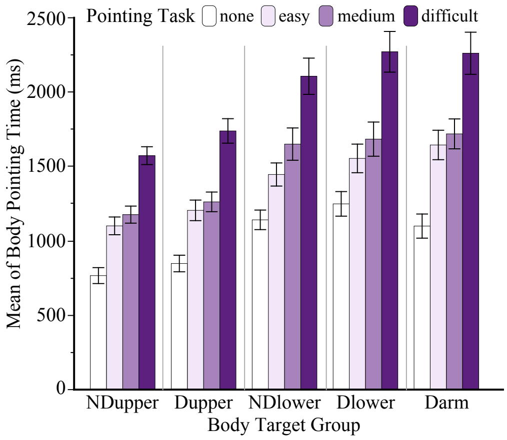


Figure 11. Interaction Pointing×Body on BODY POINTING TIME**.**

main effect of BODY TARGET GROUP is similar to the baseline (with ND UPPER and D UPPER significantly faster than all other groups). The main effect of MID-AIR POINTING is also similar to those observed before, showing that difficult pointing tasks make simultaneous body touching slower than medium or easy pointing task. Obviously, these are all significantly slower than the BODY ONLY baseline.

More interesting, the MID-AIR POINTING×BODY TARGET GROUP interaction effect reveals the actual impact of MID-AIR POINTING on ON-BODY TOUCH. As shown in Figure 11: (i) the increasing difficulty of the pointing task increases BODY POINTING TIME. In fact, despite how our task required body target selection to be the last action, the reaction times indicate that both tasks start almost simultaneously (ON-BODY TOUCH even before MID-AIR POINTING); (ii) this increase of difficulty does not change the difference between the groups of targets, but amplifies these differences. ND UPPER and D UPPER are still the group of targets that require less time to be touched.

In summary, the compound POINTING+BODY task involves interaction effects between the two atomic techniques, which not only incur a time penalty when the tasks are performed simultaneously, but also degrades pointing performance for MID-AIR POINTING (fixed in the world) when combined with a body-relative technique that involves and affects multiple limbs. However, our results also reveal that ON-BODY TOUCH on the lower parts of the body significantly impair the movement phase of pointing, and that the overall negative impact increases with the difficulty of the pointing task, especially when targeting on the pointing arm.

CONCLUSION

We propose a body-centric approach that helps to address the combinatorial explosion of possible interaction techniques in multi-surface environments. The BodyScape design space allows us to classify both existing and novel techniques based on the body's relationship to the environment and the interrelationship among different body parts.

The distributed nature of multi-surface environments highlights the need for combining interaction techniques, in series or in parallel, to accomplish more complex tasks. A body-centric approach can help predict possible interaction effects of body movements by (i) analyzing the spatial body-device relationship and (ii) proposing ways to decompose individual techniques into groups of body parts that are either *involved* in or *affected* by the interaction. We argue that studying composition of interaction techniques and extracting body-centric results regarding motor control, can help us identify powerful guidelines for interaction design, both with and without devices.

We illustrated how to use the BodyScape design space to investigate the combination of two multi-surface interaction techniques: *mid-air pointing* and *on-body touch*. This novel combination enables an eyes-free interaction with on-body targets that offer a rich set of commands that can be applied to the remote virtual target, controlled by mid-air pointing, on a large display.

We conducted a controlled experiment to study both techniques individually and in combination. We investigated performance and acceptability of 18 on-body targets, as well as interaction effects of combining the two individual techniques. Participants were most effective with targets on the torso and least effective with targets on the lower body and on the dominant arm, especially in the combined condition: reaching targets on the lower legs requires additional balance and touching the dominant arm impairs the precision of mid-air pointing because of the force applied on the pointing arm. Users consistently preferred on-body targets located on the upper body.

Our results suggest three guidelines for designing on-body interactions:

D1 *Task difficulty*: Designers should place on-body targets on the most stable locations, such as the upper torso, when the task requires precise or highly coordinated movements.

D2 *Body balance*: Designers should detect anticipatory movements, such as shifts in balance to accommodate corresponding perturbations in a primary task, e.g. freezing an on-screen cursor. The precision of a pointing task can be adversely affected if the user must also touch an on-body target that requires a shift in balance or coordination, in particular, touching the dominant arm while it is performing a separate task.

D3 *Interaction effects*: Designers should consider which body parts negatively affect users' comfort while touching an on-body target. Designers should also consider side effects of each task, such as reduced attention or fatigue that may lead to unexpected body positions or increases in errors.

REFERENCES

1. Augsten, T., Kaefer, K., Meusel, R., Fetzer, C., Kanitz, D., Stoff, T., Becker, T., Holz, C., and Baudisch, P. Multitoe: high-precision interaction with back-projected floors based on high-resolution multi-touch input. In

Proc. UIST (2010), 209–218.

2. Baudel, T., and Beaudouin-Lafon, M. Charade: remote control of objects using free-hand gestures. *CACM 36* (July 1993), 28–35.
3. Beaudouin-Lafon, M., Huot, S., Nancel, M., Mackay, W., Pietriga, E., Primet, R., Wagner, J., Chapuis, O., Pillias, C., Eagan, J., Gjerlufsen, T., and Klokmose, C. Multi-surface Interaction in the WILD Room. *IEEE Computer 45*, 4 (2012), 48–56.
4. Boring, S., Baur, D., Butz, A., Gustafson, S., and Baudisch, P. Touch projector: mobile interaction through video. In *Proc. CHI* (2010), 2287–2296.
5. Card, S., Mackinlay, J., and Robertson, G. A morphological analysis of the design space of input devices. *ACM Trans. Inf. Syst. 9*, 2 (Apr. 1991), 99–122.
6. Chen, X., Marquardt, N., Tang, A., Boring, S., and Greenberg, S. Extending a mobile device's interaction space through body-centric interaction. In *Proc. MobileHCI* (2012), to appear.
7. de Almeida, R., Pillias, C., Pietriga, E., and Cubaud, P. Looking behind bezels: french windows for wall displays. In *Proc. AVI* (2012), 124–131.
8. Dezfuli, N., Khalilbeigi, M., Huber, J., Müller, F., and Mühlhäuser, M. PalmRC: imaginary palm-based remote control for eyes-free television interaction. In *Proc. EuroiTV* (2012), 27–34.
9. Dickstein, R., and Laufer, Y. Light touch and center of mass stability during treadmill locomotion. *Gait & Posture 20*, 1 (2004), 41 – 47.
10. Goldin-Meadow, S., and Beilock, S. L. Action's influence on thought: the case of gesture. *Perspectives on Psychological Science 5*, 6 (2010), 664–674.
11. Gustafson, S., Holz, C., and Baudisch, P. Imaginary Phone: learning imaginary interfaces by transferring spatial memory from a familiar dvice. In *Proc. UIST* (2011), 283–292.
12. Harrison, C., Ramamurthy, S., and Hudson, S. On-body interaction: armed and dangerous. In *Proc. TEI* (2012), 69–76.
13. Harrison, C., Tan, D., and Morris, D. Skinput: appropriating the body as an input surface. In *Proc. CHI* (2010), 453–462.
14. Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., and Fitzgibbon, A. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In *Proc. UIST* (2011), 559–568.
15. Karrer, T., Wittenhagen, M., Lichtschlag, L., Heller, F., and Borchers, J. Pinstripe: eyes-free continuous input on interactive clothing. In *Proc. CHI* (2011), 1313–1322.
16. Klemmer, S., Hartmann, B., and Takayama, L. How bodies matter: five themes for interaction design. In *Proc. DIS* (2006), 140–149.
17. Krueger, M., Gionfriddo, T., and Hinrichsen, K. VIDEOPLACE—an artificial reality. In *Proc. CHI* (1985), 35–40.
18. Latulipe, C., Wilson, D., Huskey, S., Word, M., Carroll, A., Carroll, E., Gonzalez, B., Singh, V., Wirth, M., and Lottridge, D. Exploring the design space in technology-augmented dance. In *CHI Extended Abstracts* (2010), 2995–3000.
19. Li, F., Dearman, D., and Truong, K. Virtual shelves: interactions with orientation aware devices. In *Proc. UIST* (2009), 125–128.
20. Lin, S., Su, Z., Cheng, K., Liang, R., Kuo, T., and Chen, B. PUB - Point Upon Body: exploring eyes-free interactions and methods on an arm. In *Proc. UIST* (2011), 481–488.
21. Mine, M., Brooks Jr, F., and Sequin, C. Moving objects in space: exploiting proprioception in virtual-environment interaction. In *Proc. SIGGRAPH* (1997), 19–26.
22. Nancel, M., Wagner, J., Pietriga, E., Chapuis, O., and Mackay, W. Mid-air pan-and-zoom on wall-sized displays. In *Proc. CHI* (2011), 177–186.
23. Pederson, T., Janlert, L.-E., and Surie, D. Towards a model for egocentric interaction with physical and virtual objects. In *Proc. NordiCHI* (2010), 755–758.
24. Rekimoto, J. Pick-and-drop: a direct manipulation technique for multiple computer environments. In *Proc. UIST* (1997), 31–39.
25. Scott, S., Carpendale, S., and Inkpen, K. Territoriality in collaborative tabletop workspaces. In *Proc. CSCW* (2004), 294–303.
26. Shoemaker, G., Tsukitani, T., Kitamura, Y., and Booth, K. Body-centric interaction techniques for very large wall displays. In *Proc. NordiCHI* (2010), 463–472.
27. Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., and Blake, A. Real-time human pose recognition in parts from single depth images. In *Proc. CVPR* (2011), 1297–1304.
28. Su, R., and Bailey, B. Put them where? towards guidelines for positioning large displays in interactive workspaces. In *Proc. Interact* (2005), 337–349.
29. Tan, D., and Czerwinski, M. Effects of visual separation and physical discontinuities when distributing information across multiple displays. In *Proc. Interact* (2003), 252–255.
30. Wachs, J., Stern, H., Edan, Y., Gillam, M., Handler, J., Feied, C., and Smith, M. A gesture-based tool for sterile browsing of radiology images. *Journal of the American Medical Informatics Association 15*, 3 (2008), 321–323.
31. Wilson, A., and Benko, H. Combining multiple depth cameras and projectors for interactions on, above and between surfaces. In *Proc. UIST* (2010), 273–282.

Mid-air Pan-and-Zoom on Wall-sized Displays

Mathieu Nancel[1,2] **Julie Wagner**[2,1] **Emmanuel Pietriga**[2,1] **Olivier Chapuis**[1,2] **Wendy Mackay**[2,1]

[1]LRI - Univ Paris-Sud & CNRS
F-91405 Orsay, France

[2]INRIA
F-91405 Orsay, France

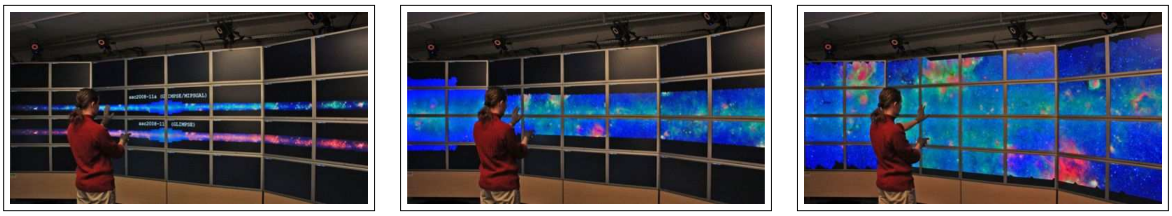


Figure 1. Panning and zooming in Spitzer's 396 032 x 12 000 = 4.7 billion pixels images of the inner part of our galaxy.

ABSTRACT

Very-high-resolution wall-sized displays offer new opportunities for interacting with large data sets. While pointing on this type of display has been studied extensively, higher-level, more complex tasks such as pan-zoom navigation have received little attention. It thus remains unclear which techniques are best suited to perform multiscale navigation in these environments. Building upon empirical data gathered from studies of pan-and-zoom on desktop computers and studies of remote pointing, we identified three key factors for the design of mid-air pan-and-zoom techniques: uni- vs. bimanual interaction, linear vs. circular movements, and level of guidance to accomplish the gestures in mid-air. After an extensive phase of iterative design and pilot testing, we ran a controlled experiment aimed at better understanding the influence of these factors on task performance. Significant effects were obtained for all three factors: bimanual interaction, linear gestures and a high level of guidance resulted in significantly improved performance. Moreover, the interaction effects among some of the dimensions suggest possible combinations for more complex, real-world tasks.

General Terms

Design, Human Factors, Performance

Author Keywords

Multi-scale interfaces, Pan & Zoom, Navigation, Wall-sized Displays, Mid-air interaction techniques.

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User Interfaces. - Graphical user interfaces.

M. Nancel, J. Wagner, E. Pietriga, O. Chapuis and W. Mackay.
Mid-air Pan-and-Zoom on Wall-sized Displays. In CHI '11: Proceedings of the SIGCHI Conference on Human Factors and Computing Systems, 177–186, ACM, May 2011.
Authors Version
doi: http://doi.acm.org/10.1145/1978942.1978969

INTRODUCTION

Very-high-resolution wall-sized displays can accommodate several hundred megapixels and make it possible to visualize very large, heterogeneous datasets in many domains [1, 3, 34]. Astronomers can use them to display telescope images constructed from hundreds of thousands of frames stitched together, such as Spitzer's 4.7 billion pixels images of the inner part of our galaxy (Figure 1). Biologists can explore the docking of complex molecules. Artists can create gigapixel images, such as the 26 gigapixel panorama of Paris based on 2,346 pictures stitched together. Crisis management centers can interact with highly detailed maps of very large areas. For example, OpenStreetMap data range from a view of the world down to street level, resulting in an image that requires 18 peta ($18 \cdot 10^{15}$) pixels at its highest level of detail.

With resolutions up to 100-dpi, these LCD-based displays afford more physical forms of navigation [3, 32, 34] compared to conventional desktop setups or to lower-resolution projection-based large displays: Users simply step back to get an overview of the displayed data and walk forward to see details, including small but legible text. However, as the examples above show, datasets increase in size faster than displays increase in dimensions and pixel density. The display depicted in Figure 1 consists of thirty-two 30-inch tiled monitors and can display a "mere" 131 million pixels. NASA's Hyperwall-2, to our knowledge the largest wall built to date, only doubles that number, and does so by adding some screens that users cannot reach. Virtual navigation is thus still required, as datasets can be several orders of magnitude too large to fit on even wall-sized displays [4].

Many interaction techniques have been specifically designed to help users navigate large multiscale worlds on desktop computers, using zooming and associated interface schemes [11]. However, high-resolution wall-sized displays pose different sets of trade-offs. It is critical to their success that interaction techniques account for both the physical characteristics of the environment and the context of use, including cooperative work aspects. Input should be location-independent and should require neither a hard surface such as a desk nor clumsy equipment: users should have the ability to move freely in front of the display and interact at a distance [3, 34]. This precludes use of conventional input devices such as keyboards and mice, as well as newer interaction techniques: The powerful multi-finger gestural input techniques designed by Malik *et al.* [22] were devised for interaction with lower-resolution large displays from afar. They require sitting at a desk, and are thus not optimal for displays of very high-resolution that afford more physical forms of navigation. The recent Cyclostar approach [21] is very elegant, but requires the display surface to be touch-enabled, a feature that wall-sized displays often lack. Cyclostar is also not well-suited to wall-sized displays, as it requires users to be within arm's reach of the display surface. While this is perfectly acceptable for displays up to 1.5m in diagonal such as SMART Boards™, users of larger displays such as the one in Figure 1 (5.8m in diagonal) would only see a very limited portion of the display while navigating. This lack of an overview would be a non-negligible hindrance as navigation is mostly driven by contextual information.

Our goal is to study different families of location-independent, mid-air input techniques for pan-zoom navigation on wall-sized displays. More specifically, we seek to answer questions related to the performance and subjective preferences of users, including: Beyond their almost universal appeal, do gestures performed in free space work better than those input via devices operated in mid-air? Is bimanual interaction more efficient in this context? Is it more tiring? Do circular, continuous gestures perform better than those that require clutching (restoring the hand or finger to a more comfortable posture)? We ground our work on both theoretical and experimental work on bimanual input [8, 14, 18], the influence of limb segments on input performance [2, 35], on types of gestures [25, 33] and on the integral nature, in terms of perceptual structure, of the pan-zoom task [17]. In particular, we are interested in comparing the following dimensions: bimanual vs. unimanual input; device-based vs. free-hand techniques; degrees of freedom (DOF) and associated kinesthetic and haptic feedback; and types of movements: linear gestures vs. circular, clutch-free gestures.

RELATED WORK

This work is at the intersection of many HCI research areas, including multiscale interfaces, large displays, spatial input and travel in virtual environments. This section highlights strongly related or seminal work that guided our designs and we point to relevant surveys, when available.

Large Displays

Large displays have been the focus of much research and evaluation over the last ten years. Ni *et al.* [27] survey hardware configurations, rendering techniques as well as interaction techniques for many different types of large displays.

Overall, the body of empirical work on large displays suggests that users can greatly benefit from their use. It also shows that the design of interaction techniques has to be carefully adapted to the characteristics of these displays and to their context of use. Early studies investigated how users could benefit from larger displays in different settings. Baudisch *et al.* [4] found advantages to using a large focus+context screen over zooming and overviews to extract information from large documents such as maps and schematics of circuit boards. Improvements to spatial task performance were also identified in several complementary studies [12, 26, 31].

Other works have focused on the size and configuration of high-resolution tiled displays. Ball *et al.* [3] found that for tasks involving pan-zoom, such as navigating to a known location, searching for specific targets or looking for patterns, users perform better with larger viewport sizes that require less virtual navigation, promoting physical navigation instead. Virtual navigation was always performed with the same device: a gyroscopic mouse. Results from other recent studies suggest that large displays are also beneficial for information visualization and analysis tasks thanks to the larger amount of data that can be displayed [1, 34].

Spatial Input and Mid-air Interaction Techniques

Spatial input has been studied for years in the context of travel in immersive virtual environments and other 3D user interfaces based on virtual camera control with techniques using gloves, bimanual input and leaning, or high degrees of freedom devices [7, 24, 35]. Hinckley *et al.* [16] present a survey of design issues in spatial input, including fatigue, recalibration, clutching, motion and orientation, unimanual vs. bimanual interaction. One important issue they raise is the interdependency of all these aspects, that makes formal studies challenging, as we will see later.

Several input devices make it possible to point in mid-air on large displays: commercial devices such as gyroscopic mice, or *soap* [5], based on hardware found in a conventional optical mouse wrapped in elastic fabric. ARC-Pad [23] enables seamless absolute+relative pointing on large displays through a mobile touchscreen. The VisionWand [10] is a passive wand whose colored tips are tracked in 3D by two webcams. The multiple degrees of freedom enable a richer interaction vocabulary, that includes pan-zoom navigation.

Recent advances in motion tracking and dynamic gesture recognition technologies now make it possible to investigate freehand input techniques. Vogel and Balakrishnan [32] propose three pointing and clicking techniques that work with bare hands, with emphasis on important design characteristics such as accuracy, performance, but also comfort of use. Zigelbaum *et al.* [36] describe a gestural interface based on Oblong's *g-speak* spatial operating environment to navigate in a collection of videos arranged in a 3D interface through a set of twenty hand-centric gestures.

Multi-scale Navigation on the Desktop

Pan-zoom navigation techniques have been studied for many years in the more general context of multiscale interfaces for the desktop. Cockburn *et al.* [11] provide a thorough survey of the many zooming, overview + detail and focus + context techniques, as well as empirical work that evaluated them.

Of particular interest to us is the work by Guiard *et al.* on multiscale pointing. Multiscale pointing consists of panning and zooming the view so as to bring the target in view, followed by a cursor pointing action to that target [15]. They performed several empirical studies, showing that multiscale pointing obeys Fitts' law, and that performance bandwidth is proportional to view size (up to a ceiling that we far exceed on wall-sized displays). They introduced an experimental task adapted from Fitts' reciprocal pointing task, that we further adapt to take into account potential overshoots in the scale dimension. An earlier paper [6] evaluated pan-zoom performance with uni- and bimanual input, suggesting that performance is enhanced with two hands, as it affords better pan-zoom coordination. Pan-zoom navigation has however not received much attention beyond desktop interfaces, except for the recent work by Malacria *et al.* on Cyclostar [21], specifically designed for touch-enabled surfaces and discussed in more detail in the next section.

PANNING AND ZOOMING IN MID-AIR

A large body of literature is devoted to the design and evaluation of input devices that feature a high number of degrees of freedom (DOF). Available degrees of freedom have a direct impact on the potential for parallelization of actions required to achieve the task. For example, 6DOF input devices can increase the degree of parallelization of docking tasks [35], though studies report limits in terms of human capacity to handle all DOFs simultaneously. Pan and zoom is a 3DOF task: the user controls the view's position (x, y) and its scale (s). The possible solutions for mapping pan and zoom to three input channels are endless.

The film industry offers interesting and visually attractive scenarios with movies such as *Minority Report* which show users interacting via freehand gestures to navigate in a seemingly fluid and efficient way. The technology to achieve this type of interaction is now available in research laboratories and beyond [36]. However, it remains unclear how freehand gestures actually fare when compared to device-based input techniques that take advantage of the human ability to use physical tools [10] and suffer less from problems commonly associated with spatial input [16], such as precision and fatigue. Years of research in virtual reality have demonstrated that devising efficient navigation techniques for immersive virtual environments is still a challenge.

Our goal is to study families of input techniques that let users pan and zoom from any location in front of very high-resolution, wall-sized displays. We made no *a priori* assumptions about relevant metaphors or technologies and considered freehand as well as device-based techniques.

An extensive design and testing phase allowed us to limit the number of candidates for the subsequent formal evaluation. For instance, the apparently intuitive solution that consists in using two hands or two fingers to zoom with pinch and stretch gestures was considered but quickly discarded: while these gestures work well on touch-sensitive surfaces such as tabletops, they are much less natural when performed in mid-air. Most importantly, they proved quite inaccurate, and tiring. Another category of techniques that was discarded are those based on first-order-of-control and operated via an elastic or isometric input device. As reported in the literature in the case of pointing, e.g., [9], our pilot tests revealed that techniques based on first-order-of-control allow for fast and comfortable coarse navigation, but perform poorly during the final precise positioning phase, causing numerous overshoots.

Factors		Advantages	Disadvantages
Hands	One	• One hand available for other actions	• Pan and zoom are performed sequentially
	Two	• Pan and zoom can be performed in parallel	• No hand available for other actions
Gesture	Linear	• Direct, natural mapping to zoom actions	• Potentially requires clutching
	Circular	• No clutching (continuous gesture)	• Less natural mapping to zoom actions
Degree of Guidance	1D path	• Input guided by strong haptic feedback • Mainly involves fingers	• Only 1 degree of freedom
	2D surface	• Many degrees of freedom • Mainly involves fingers	• Input guided by limited haptic feedback
	3D free hand	• Many degrees of freedom • No device	• No haptic feedback • Mainly involves whole hand and arms

Table 1. Key Dimensions of the Design Space

We eventually identified a set of twelve candidate techniques. Their design was informed by related empirical studies reported in the literature and refined through prototyping and pilot testing. These techniques can be organized according to three key dimensions forming a design space (Table 1), and introduced in the following sections. In addition to performance (task time and accuracy), we took into account other usability issues, such as fatigue and ease of use.

Unimanual vs. Bimanual Input

In their paper on the perceptual structure of multidimensional input, Jacob and Sibert claim that panning and zooming are integrally related: the user does not think of them as separate operations, but rather as a single, integral task like "*focus on that area over there*" [17]. Buxton and Myers [8] and later Bourgeois and Guiard [6] observed high levels of parallelism for pan-zoom operations, further supporting this argument. The level of parallelism correlates with task performance and is typically well afforded by the use of bimanual input techniques [14, 18]. While we expect bimanual techniques to outperform unimanual ones, we are still interested in comparing their performance, as the latter might still be of interest in more complex, real-world tasks that require input channels for other actions.

Linear vs. Circular Gestures

Navigating in the scale dimension (zooming in and out) is a task typically performed through vertical scroll gestures on, e.g., a mouse wheel or a touchpad. The mapping from input to command is natural, but often entails clutching as the course of mouse wheels and touchpads is very limited. An alternative consists in mapping continuous circular gestures to zooming. Clockwise gestures zoom in; counter-clockwise

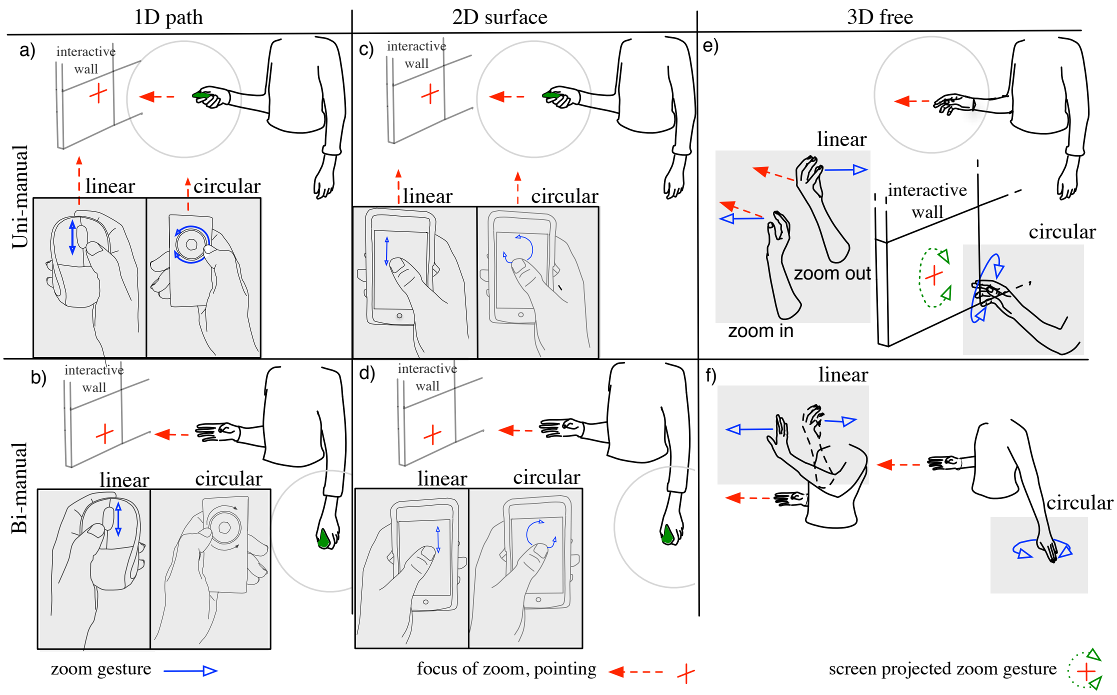


Figure 2. Matrix of the 12 techniques organized according to key characteristics: uni- vs. bimanual, degree of guidance, linear vs. circular gestures. 1D path involves guiding gestures along a particular path in space; in 2D surface gestures are made on a touch-sensitive surface; while in 3D free gestures are totally free.

gestures zoom out. Despite the less natural mapping from input to commands, such continuous, clutch-free gestures have been successfully applied to vertical scrolling in documents [25, 33], and to pan and zoom on large, touch-sensitive surfaces in CycloStar [21]. Circular gestures potentially benefit from an automatic Vernier effect [13]: as zooming is mapped to angular movements, the larger the circular gesture's radius, the greater the distance that has to be covered to make a full circle, and consequently the more precise the input.

Guidance through Passive Haptic Feedback

Two main categories of techniques have been studied for mid-air interaction on wall-sized displays: freehand techniques based on motion tracking [32, 36]; and techniques that require the user to hold an input device [5, 10, 19, 23]. Input devices provide some guidance to the user in terms of what gesture to execute, as all of them provide some sort of passive haptic feedback: A finger operating a knob or a mouse wheel follows a specific path; gestures on touch-enabled devices are made on planar surfaces. Freehand techniques, on the contrary, provide essentially no feedback to the user who can only rely on proprioception [24] to execute the gesture. We call this dimension the *degree of guidance*. Gestures can be guided to follow a particular path in space (*1D path*); they can be guided on a touch-sensitive surface (*2D surface*) ; or they can be totally free (*3D free*). These three values correspond to decreasing amounts of passive haptic feedback for the performance of input gestures.

DESIGN CHOICES

Panning. For all techniques, controlling the cursor's position is achieved naturally by ray-casting from the dominant hand to the wall display (dashed arrows in Figure 2). As mentioned earlier, first order of control was discarded for both pan and zoom operations. Panning is achieved by dragging, as in applications such as Adobe Illustrator™ or Google Maps™ with their typical hand-shaped cursor.

Zooming. As in desktop applications such as Google Maps or NASA's WorldWind, *linear* techniques zoom in by moving forward towards the display and zoom out by moving backwards; *circular* techniques zoom in by turning clockwise and zoom out by turning counter-clockwise (solid arrows in Figure 2). Pointing plays an important role when zooming, as it specifies the focus of expansion (zoom in)/contraction (zoom out). Letting users specify this focus point is very important on displays of that physical size, as they will typically not be standing right in the center. A focus of expansion implicitly located at the center of the screen would make zooming operations tedious and hard to control as every zoom operation would require multiple panning actions to compensate drifts induced by the offset focus.

Bi-manual interaction. All bimanual techniques (Figure 2, bottom row) are grounded in Guiard's study of asymmetric division of labor in bimanual actions that led to the Kinematic chain model [14]. Following the observation that motion of the dominant hand typically finds its spatial reference in the results of motion of the non-dominant-hand, we assign pointing and panning to the dominant hand, while the non-dominant hand controls zoom, as is typically the case for bimanual pan-zoom techniques on the desktop [6, 8].

Input Gestures via a Device

The main limb segments involved in the input of gestures via a device are the fingers and, to a lesser extent, the forearm (for the dominant hand). This group of techniques is illustrated in Figure 2, columns *1D path* and *2D surface*.

Column *1D path* illustrates techniques that provide a high degree of guidance for executing the zooming gestures. The first row corresponds to *one handed* techniques: the device is operated by the dominant hand, which also controls pointing via ray-casting. The second row corresponds to *two handed* techniques: the dominant hand controls pointing via ray-casting, while the non-dominant hand controls zoom using the device. *linear* gestures can be input using, e.g., a wireless handheld mouse featuring a scroll wheel; *circular* gestures using, e.g., any type of handheld knob. Depressing a button on the device activates drag mode for panning.

Column *2D surface* illustrates techniques that use a touch-sensitive surface for input, providing a lesser degree of guidance. The surface is divided horizontally in two areas. Users zoom in the upper area either by moving the thumb up and down (*linear* case), or by drawing approximate circles (*circular* case). Touching the lower area activates drag mode for panning. Users just rely on proprioceptive information to switch between both areas and do not have to look at the device. These techniques can be implemented with a touch-sensitive handheld device such as a PDA or smartphone.

1D path techniques employing *circular* gestures will provide more guidance, but will not benefit from the earlier-mentioned Vernier effect, as input is constrained to one specific trajectory. However, the range of amplitudes that can be covered with the thumb is limited [30]. This should minimize the trade-off between *1D path* and *2D surface* in that respect. For *2D surface* techniques, rubbing gestures [28] were considered to avoid clutching when performing linear gestures, but were found to be impractical when performed with the thumb on a handheld touch-sensitive surface. As a technique designed specifically for thumb input, we were also interested in MicroRolls [30]. However, these were originally designed for discrete input. Cardinal MicroRolls would have had to be mapped to first order of control, which we discarded as discussed earlier, and circular MicroRolls are not precise enough for zoom control.

Input Gestures in Free Space

The main limb segments involved in performing gestures in free space are the wrist, forearm and upper arm. This group of techniques is illustrated in Figure 2, column *3D free*.

The first row illustrates *one handed* techniques using either *linear* or *circular* gestures. The technique using *circular* gestures is actually very close to the CycloStar zooming gesture, but performed in mid-air, without touching any surface. Users perform circular gestures with the dominant hand and forearm oriented toward the display. As in CycloStar, the focus of expansion is the centroid of the round shape corresponding to the cursor's circular path, here projected on the display surface (dotted arrow in Figure 2-e). The technique using *linear* gestures consists in pushing the dominant hand forward to zoom in, as if reaching for something, with the palm towards the target. Turning the hand and pulling backward (away from the display) zooms out. Users point orthogonally to the palm of the same hand (blue arrows in Figure 2-e, left side), with the arm slightly tilted for greater comfort. The second row illustrates *two handed* techniques (Figure 2-f). The *linear* zooming gestures are similar to the ones above, but are performed with the non-dominant hand, the dominant hand still being used for pointing and specifying the focus of expansion. In the *circular* case, users adopt a potentially less tiring posture, pointing at the floor with their non-dominant hand and making circular movements. All other postures and movements being ignored by the system for the non-dominant hand, the user can easily clutch. Several options can be considered for engaging drag mode: specific hand postures such as pinching, or using a small wireless actuator (e.g., a button).

EXPERIMENT

We conducted an experiment using a [2×2×3] within-subjects design with three primary factors: HANDEDNESS $\in$ $\{$*OneHanded*, *TwoHanded*$\}$, GESTURE $\in$ $\{$*Circular*, *Linear*$\}$, and GUIDANCE $\in$ $\{$*1DPath*, *2DSurface*, *3DFree*$\}$ to evaluate the 12 unique interaction techniques described above. We controlled for potential distance effects by introducing the DISTANCE between two consecutive targets as a secondary within-subjects factor. We systematically varied these factors in the context of a multiscale navigation task within a wall-sized display environment.

Measures include performance time and number of overshoots, treated as errors. Overshoots occur when participants zooms beyond the target zoom level, and indicate situations in which the participant has less precision of control over the level of zoom. For instance, from an overview of Canada, zooming down to street level in Google Maps when what the user actually wanted was to get an overview of Vancouver.

Hypotheses

Based on the research literature and our own experience with the above techniques, we made the following 7 hypotheses.

Handedness: prior work [6, 8, 15, 18] suggests that two-handed gestures will be faster than one-handed gestures (*H1*) because panning and zooming are complementary actions, integrated into a single task [17]. Two-handed gestures should also be more accurate and easier to use (*H2*).

Gesture: Linear gestures should map better to the zooming component of the task, but should eventually be slower because of clutching, the limited action space compared to zoom range requiring participants to repeatedly reposition their hand/finger (*H3*). Prior work [25, 33] suggests that users will prefer clutch-free circular gestures (*H4*).

Device vs. Free Space: Zhai et al. [35] suggest that techniques using the smaller muscle groups of fingers should be

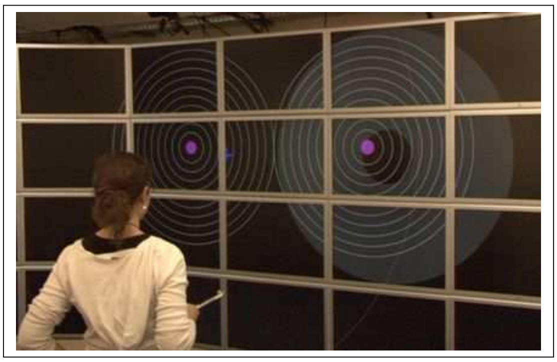


Figure 3. Participant performing the task

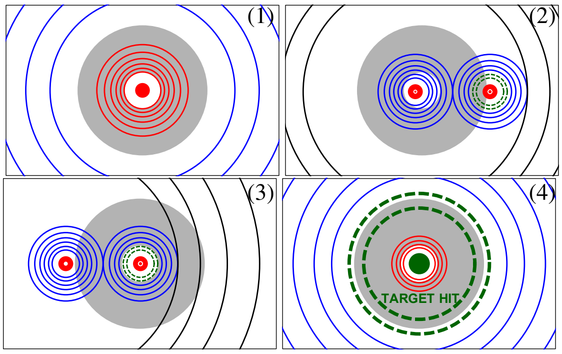


Figure 4. Task (schematic representation using altered colors): (1) Groups of concentric circles represent a given position and zoom level. (2) Zooming out until the neighboring set of circles appears. (3-4) Pan and zoom until the target (green inner disc and circles, dashed for illustration purposes only) is positioned correctly with respect to the stationary gray ring.

more efficient than those using upper limb segments. Balakrishnan *et al.* [2] moderate this observation with findings suggesting that the fingers are not performing better than forearm or wrist for a reciprocal pointing task. Nevertheless, they acknowledge that differences exist in the motor system's ability to control the different limb segments. Based on the gestures to be performed and taking into account the physical size and mass of the segments involved, we predicted that techniques using fingers (*1DPath* and *2DSurface* conditions), should be faster than those requiring larger muscle groups (hands and arms, *3DFree* conditions) (*H5*).

We also predicted that *1DPath* gestures would be faster, with fewer overshoots than techniques with lesser haptic feedback, i.e., *2DSurface* and *3DFree* (*H6*). Finally, we predicted that *3DFree* gestures would be more tiring (*H7*).

Participants

We recruited 12 participants (1 female), ranging in age from 20 to 30 years old (average 24.75, median = 25). All are right-handed daily computer users. None are color-blind.

Apparatus

Hardware. The display wall (Fig. 1 and 3) consists of 32 high-resolution 30" LCDs laid out in an 8×4 matrix, 5.5 meters wide and 1.8 meters high. It can display 20480×6400 pixels. A cluster of 16 computers, each with two high-end nVidia 8800GT graphics cards, communicate via a dedicated high-speed network through a front-end computer. Our goal is to identify the performance characteristics of each technique from the user's perspective. It is thus essential that each technique operates equally well from a purely technological perspective. We use a VICON motion-capture system to track passive IR retroreflective markers and provide 3D object coordinates with sub-millimeter accuracy at 200Hz (although gesture recognition technologies are constantly improving, such a system is still necessary to get reliable and precise 3D position/orientation information). The *Linear 1DPath* condition uses the wheel of a wireless Logitech M305 mouse (Fig. 2-a,b). The *Circular 1DPath* condition uses a wireless Samsung SM30P pointing device, normally used for presentations (Fig. 2-a,b). All *2DSurface* conditions use an iPod Touch. So as to avoid failures from gesture segmentation algorithms that would impact task performance in an uncontrolled manner, we use an explicit mode switch to unambiguously engage drag mode (panning). As mentioned earlier, we use the device's main button for *1DPath* conditions, and the lower area of the touch-sensitive surface for *2DSurface* conditions. While in real-world applications we would use specific hand postures such as pinching in *3DFree* conditions, for the sake of robustness we use a wireless mouse button whose activation is seamlessly integrated with the gesture.

Software. The experiment was written in Java 1.5 running on Mac OS X and was implemented with the open source ZVTM toolkit [29] (`http://zvtm.sf.net`) modified to run on clusters of computers driving display walls. Touchstone [20] was used to manage the experiment.

Pan-Zoom Task

The task is a variation of Guiard *et al.*'s multiscale pointing task [15], adapted to take overshoots into account. Participants navigate through an abstract information space made of two groups of concentric circles: the *start group* and the *target group*. Each group consists of seven series of 10 concentric circles symbolizing different zoom levels, each designated by a different color (Fig. 4.2). The target group features two additional green circles (dashed in Fig. 4.4) and a disc, referred to as C_1, C_2 and C_3 from smallest to largest.

Participants start at a high zoom level in the start group (Fig. 4.1). They zoom out until the neighboring target group appears (Fig. 4.2). It may appear either on the left or right side of the start group. Then they pan and zoom into the target group until they reach the correct zoom level and the target is correctly centered. A stationary gray ring symbolizes the correct zoom level and position (Fig. 4-(1-4)). Its radii are $r_1 = 4400$ and $r_2 = 12480$ pixels. All three criteria must be met for the trial to end: $\mathcal{A}$) C_1 is fully contained within the stationary ring's hole (radius $= r_1$), $\mathcal{B}$) $radius(C_2) < r_2$, $\mathcal{C}$) $radius(C_3) > r_2$. Overshoots occur when the zoom level is higher than the maximum level required to meet criteria $\mathcal{B}$ and $\mathcal{C}$, in which case participants have to zoom out again (C_1 becomes white instead of green in that situation). When all conditions are met, the message TARGET HIT appears and the thickness of C_2 and C_3 is increased (Fig. 4.4). The trial ends when the position and zoom level have stabilized for at least 1.2 seconds (all trials must be successfully completed).

Procedure

The experiment presents each subject with six replications of each of the 12 techniques at three DISTANCES. The experiment is organized into four sessions that each present three techniques: One combination of the GESTURE and HANDEDNESS factors and all three degrees of GUIDANCE. Each session lasts between 30 and 90 minutes, depending on techniques and participant. Participants are required to wait at least one hour between two consecutive sessions, and to complete the whole experiment within four days or fewer, with a maximum of two sessions per day to avoid too much fatigue and boredom. Participants stand 1.7m from the wall and are asked to find a comfortable position so they can perform gestures quickly, but in a relaxed way.

Practice Condition: Participants are given a brief introduction at the beginning of the first session. Each technique begins with a practice condition, with trials at three different DISTANCES: (49 920, 798 720 and 12 779 520 pixels). Measures for the experimental condition start as soon as 1) participants feel comfortable and 2) task performance time variation for the last four trials is less than 30% of the task time average in that window.

Experimental Condition: Each technique is presented in a block of 18 trials consisting of 6 replications at each DISTANCE. Trials, blocks and sessions are fully counter-balanced within and across subjects, using a Latin square design.

Measures: We measure movement time *MT* and number of overshoots for each of 2592 trials: 2 GESTURE × 2 HANDEDNESS × 3 GUIDANCE × 3 DISTANCE × 12 participants × 6 replications. Participants also answer questions, based on a 5-point Likert scale, about their perceived *performance*, *accuracy*, *ease of learning*, *ease of use*, and *fatigue*. They rank the techniques with respect to the GUIDANCE factor after each session. When they have been exposed to both conditions of HANDEDNESS or GESTURE, they rank those as well. After the last session, they rank the techniques individually and by factor. Participants are encouraged to make additional observations and comments about any of the above.

Results and Discussion: Movement Time

Prior to our analysis, we checked the performance for unwanted effects from secondary factors. We checked for individual performance differences across subjects and found that, for all 12 participants, movement time and number of overshoots were perfectly correlated with the overall performance measures. As expected, movement time data are skewed positively; replications of unique experimental conditions are thus handled by taking the median (note that taking the mean yields similar results). In all remaining analysis, we handled participant as a random variable, using the standard repeated measures REML technique. We found no significant fatigue effect although we did find a significant learning effect across sessions. Participants performed about $1.4\,s$ more slowly in the first session and then became slightly faster over the next three sessions. However, we found no significant interaction between session orders and main factors. As the factors were counter-balanced, this created no adverse effects in the analysis.

Factor	DF, DFDen	F Ratio	p
HANDS	1,11	24.65	0.0004 *
GESTURE	1,11	42.87	< 0.0001 *
GUIDANCE	2,22	58.80	< 0.0001 *
DIST	2,22	228.8	< 0.0001 *
HANDS×GESTURE	1,11	2.060	0.1790
HANDS×GUIDANCE	2,22	4.914	0.0172 *
GESTURE×GUIDANCE	2,22	10.38	0.0007 *
GESTURE×DIST	2,22	17.27	< 0.0001 *
HANDS×DIST	2,22	11.57	0.0004 *
GUIDANCE×DIST	4,44	3.828	0.0094 *
HANDS×GESTURE×GUIDANCE	2,22	1.127	0.3420
HANDS×GESTURE×DIST	2,22	0.790	0.4661
HANDS×GUIDANCE×DIST	4,44	0.650	0.6301
GESTURE×GUIDANCE×DIST	4,44	3.750	0.0104 *
HANDS×GESTURE×GUIDANCE×DIST	4,44	1.049	0.3929

Table 2. Results of the full factorial ANOVA for MT.

Table 2 details results of the full factorial ANOVA for the model *MT* ∼ HANDS × GUIDANCE × GESTURE × DIST × Rand(Participant). We observe that HANDS has a significant effect on *MT* (Figure 5-a[1]). A post-hoc Tukey test shows that *TwoHanded* gestures are significantly faster than *OneHanded* gestures (avg. $9690\,ms$ vs. $11869\,ms$). We found a significant interaction effect of HANDS × GUIDANCE (Figure 5-a). The interaction does not change the significance of the post-hoc test, but indicates that the magnitude of the difference is greater for *3DFree* than for *2DSurface* and greater for *2DSurface* than for *1DPath* techniques.

Unsurprisingly, performance data strongly support (*H1*): all other conditions being equal, two-handed techniques are consistently faster than one-handed techniques. An interesting observation is that using two hands is more advantageous when the degree of guidance for achieving gestures is low.

GUIDANCE has a significant effect on *MT* (Figure 5-b). A post-hoc Tukey test shows that *1DPath* (avg. $9511\,ms$) is significantly faster than *2DSurface* ($10894\,ms$), which in turn is significantly faster than *3DFree* ($11934\,ms$). This time the HANDS × GUIDANCE interaction changes the significance of the test (Figure 5-b). The difference is that a post-hoc Tukey test shows no significant difference between *2DSurface* and *3DFree* for *TwoHanded*.

Both hypotheses (*H5*) and (*H6*) are supported: involving smaller muscle groups improves performance; providing higher guidance further contributes to this. However, this effect is less pronounced in *TwoHanded* conditions. This confirms the previous observation that a higher degree of guidance is especially useful when a single hand is involved.

GESTURE also has a significant effect on *MT*. A post-hoc Tukey test shows that *Linear* movements (avg. $9384\,ms$) performed significantly faster than *Circular* gestures ($12175\,ms$). However, we have a strong significant interaction of GESTURE ×

[1]Error bars in all the figures represent the 95% confidence limit of the mean of the medians per participants ($\pm StdErr \times 1.96$).

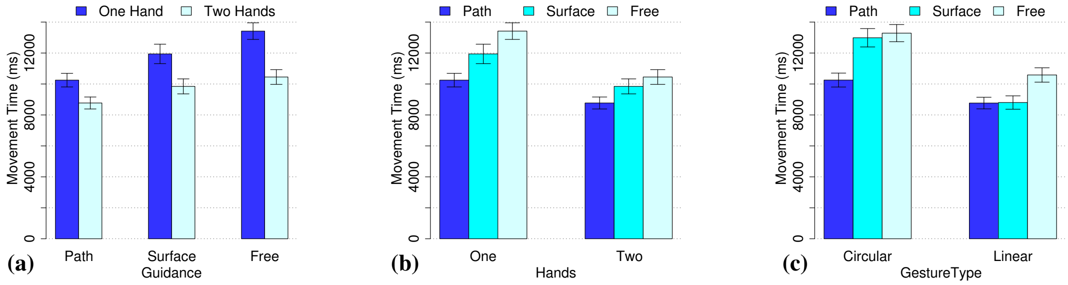


Figure 5. (a): ***MT*** **per** HANDS × GUIDANCE**. (b)** ***MT*** **per** GUIDANCE × HANDS**. (c)** ***MT*** **per** GUIDANCE × GESTURE**.**

GUIDANCE (Figure 5-c). A post-hoc Tukey test shows that (i) for *Circular* gestures: *1DPath* guidance is faster than both *2DSurface* and *3DFree* with no significant difference between *2DSurface* and *3DFree*; (ii) for *Linear* gestures, there is no significant difference between *1DPath* and *2DSurface*, but a significant difference between *2DSurface* and *3DFree*; (iii) for *1DPath* guidance there is no significant difference between *Circular* and *Linear* gestures, but there is a significant difference between *Circular* and *Linear* for *2DSurface* and *3DFree* guidance.

Surprisingly, *Linear* gestures are generally faster than *Circular* ones. (*H3*), that claimed that *Linear* gestures should be slower because of clutching, is not supported. Performance differences between gesture types are however affected by the degree of guidance: *Circular* gestures with *1DPath* guidance (e.g., a knob) are comparable to *Linear* gestures with low guidance. We tentatively explain the lower performance of *Circular* gestures with *2DSurface* guidance by the difficulty of performing circular gestures with the thumb [30], also observed here.

Another interesting observation is that our analogue of CycloStar in mid-air (*Circular* gestures with *3DFree* guidance) performs poorly. It seems that the lack of a surface to guide the gesture significantly degrades this technique's usability. Another factor contributing to its poor performance in our study is likely related to overshoots, as discussed below.

As expected, distance to target (DIST) has a significant effect on *MT*. A post-hoc Tukey test shows that *MT* increases significantly with *distance*. There are several significant interactions between DIST and the main factors (Fig. 6), but none of these change the relative performance ordering for the main factors. These interactions are due to a change in the magnitude of the difference across conditions, confirming that the choice of an efficient technique is of increasing importance as the task becomes harder.

Results and Discussion: Overshoots

As detailed earlier in the description of task design, overshoots correspond to zooming beyond the target zoom level and are treated as errors. We consider the model *Overshoots* ∼ HANDS × GUIDANCE × GESTURE × DIST × Rand(Participant).

We observe significant simple effects on *Overshoots* for GESTURE ($F_{1,11} = 21.04$, $p = 0.0008$) and GUIDANCE ($F_{2,22} = 53.80$, $p < 0.0001$), and one significant interaction effect for GESTURE × GUIDANCE ($F_{2,22} = 8.63$, $p = 0.0017$). *Circular* gestures exhibit more overshoots than *Linear* gestures (1.65 vs. 2.71).

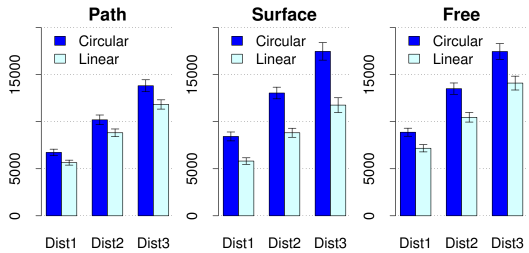


Figure 6. ***MT*** **per** DIST × GESTURE**, for each** GUIDANCE

2DSurface gestures exhibit more overshoots than *1DPath* and *3DFree* gestures (3.75 for *2DSurface* vs. 1.52 for *1DPath*, and 1.26 for *3DFree*). There is a significant difference between *Linear* and *Circular* gestures for *2DSurface* and *3DFree*, but not *1DPath*. Moreover, overshoots exhibit the same interaction effect for *2DSurface* gestures: *Circular 2DSurface* result in significantly more overshoots than *Linear 2DSurface* (4.68 vs. 2.82).

The observed higher number of overshoots for *Circular* techniques helps explain the generally lower *MT* performance measured for this type of gestures. The best-fitting ellipse algorithm involved in the recognition of *Circular* gestures has an inherently higher cost of recovery, introducing a delay when reversing course. The poor performance of our analogue of CycloStar is at least partially due to this, knowing that there was a major difference between the zooming experiment reported in [21] and the present one: we included overshoots in our task design, whereas the CycloStar experiment apparently did not (there is no report of such a measure in task design or results analysis), thus ignoring this issue.

Results and Discussion: Qualitative Results

Qualitative data confirms our results. Participants generally preferred *TwoHanded* to *OneHanded* techniques (8/12) and *Linear* to *Circular* gestures (10/12). Subjective preferences about degree of guidance were mixed, with 4 participants preferring the high degree of guidance provided by *1DPath* techniques, only 1 for both of *2DSurface* and *3DFree* techniques, and all others expressing no particular preferences. Looking at the details of answers to our 5-point Likert scale questions about perceived speed, accuracy, ease of use and fatigue, significant results ($p < 0.002$) were obtained only for degree of GUIDANCE, with *1DPath* being consistently rated higher than *2DSurface* and *3DFree*; and for HANDS, *TwoHanded* techniques being considered less tiring than *OneHanded* techniques ($p < 0.03$).

GROUP	HANDS	GESTURE	GUIDANCE	Figure	*MT* (ms)
Gr1	*TwoHanded*	*Linear*	*2DSurface*	2-d	8 100
	TwoHanded	*Linear*	*1DPath*	2-b	8 377
Gr2	*OneHanded*	*Linear*	*1DPath*	2-a	9 160
	TwoHanded	*Circular*	*1DPath*	2-b	9 168
	TwoHanded	*Linear*	*3DFree*	2-f	9 185
	OneHanded	*Linear*	*2DSurface*	2-c	9 504
Gr3	*OneHanded*	*Circular*	*1DPath*	2-a	11 340
	TwoHanded	*Circular*	*2DSurface*	2-d	11 591
	TwoHanded	*Circular*	*3DFree*	2-f	11 718
	OneHanded	*Linear*	*3DFree*	2-e	11 981
Gr4	*OneHanded*	*Circular*	*2DSurface*	2-c	14 380
	OneHanded	*Circular*	*3DFree*	2-e	14 851

Table 3. Groups of techniques according to *MT*

Comments from participants suggest that in the *OneHanded* condition, zoom gestures interfere with pointing as they introduce additional hand jitter and consequently lower accuracy. Some participants also commented that pointing and zooming were confounded in the *OneHanded* conditions, making the techniques difficult to use (*H2*). However, two participants strongly preferred one-handed gestures, arguing that they were less complex and less tiring. They assumed their performance was better (even though it was not), probably because they experienced more overshoots in the *two handed* condition, which may have led to their conclusions. One of them mentioned that for the *one handed* condition there was *"no need for coordination"*; techniques were *"more relaxed"* and made it *"easier to pan and zoom"*.

Linear gestures were preferred to *Circular* ones, participants commenting that circular gestures were difficult to perform without guidance, that circular gestures for zooming interfered with linear gestures for panning, and that circular gestures were hard to map to zoom factor. All but one participants preferred linear gestures overall although one commented that he liked *"the continuity of circular gestures"*. Others commented that *"making good circles without a guide is hard"* and did not like having to turn their hands. These findings contradict our hypothesis that users would prefer clutch-free circular gestures (*H4*). This hypothesis was based on observations made for techniques operated on a desktop, not in mid-air, and involved different limb segments. In many of our conditions, the gestures had to be performed with the thumb, and were thus more complex to achieve than when using, e.g., the index finger in conjunction with hand or forearm movements. Several participants commented on this interaction effect: *"[It is] too hard to do circle gestures without a guide"*, *"Linear movements are easier on the iPod"* and *"[Is it] impossible to do circular movements on a surface, maybe with some oil?"*.

Finally, as hypothesized (*H7*), participants found *1DPath* guidance least tiring while *3DFree* caused the most fatigue.

Results and Discussion: Individual Techniques

The analysis of variance for the model *MT* $\sim$ HANDS $\times$ GUIDANCE $\times$ GESTURE $\times$ DIST $\times$ Rand(Participant) does not show a significant triple interaction between the three main factors (Table 2). Formally, we cannot say more than the above about the ranking of the twelve techniques. However, based on the results about *MT* above, we can observe four distinct groups of techniques, shown in Table 3. As a side note, if we consider the model *MT* $\sim$ GROUP $\times$ Rand(Participant), the ANOVA shows a significant effect of GROUP ($F_{3,33} = 65.35$, $p < 0.0001$) and a post-hoc Tukey test shows a significant difference between each groups.

Gr1 contains the two fastest techniques with similar *MT* : *TwoHanded*, *Linear* gestures with either *2DSurface* or *1DPath* degrees of guidance. Optimal performance in terms of movement time implies the use of two hands and a device to guide gestural input.

Gr2 contains the four techniques that come next and also have close *MT* : the *OneHanded* version of the two fastest techniques, the *TwoHanded Circular 1DPath* and the *TwoHanded Linear 3DFree* techniques. Techniques in this group are of interest as they exhibit a relatively good level of performance while broadening possible choices for interaction designers. For instance, the unimanual techniques in this group make one hand available to perform other actions. The *3DFree* technique is also of interest as it does not require the user to hold any equipment and is generally appealing to users.

Gr3 contains techniques that again have very close *MT* but about $2.3\,s$ slower than the techniques of *Gr2*. This group consists of *OneHanded Circular 1DPath*, *TwoHanded Circular 2DSurface* and *3DFree*, and *OneHanded Linear 3DFree*. Techniques in this group are of lesser interest, except maybe for the *OneHanded Linear 3DFree* technique, which is the fastest unimanual technique using gestures performed in free space.

Gr4 contains the 2 techniques performing worst, *OneHanded Circular 2DSurface* and *3DFree*. These are significantly slower than all others, about $3\,s$ slower than the techniques of *Gr3* and about $6\,s$ slower than the techniques of *Gr1*. Our data suggest that these techniques should be rejected.

SUMMARY AND FUTURE WORK

We studied different families of location-independent, mid-air input techniques for pan-zoom navigation on wall-sized displays. After an extensive exploratory design phase, we identified the following key factors for the design of such techniques: handedness (uni- vs. bimanual input), gesture type (linear or circular), and level of guidance (movements restricted to a 1D path, a 2D surface or free movements in 3D space). We systematically evaluated each combination of these factors through a controlled experiment in which participants performed pan-and-zoom navigation in an abstract, very large multiscale environment, with distances up to 12 million pixels.

Experimental results identify several successful mid-air input techniques that can be used to navigate efficiently in very large datasets on wall-sized displays. In addition to identifying groups of alternative techniques based on performance, but each with specific characteristics, the experiment also suggests clear results with respect to the factors that constitute our design space. For example, despite their inherent and almost universal appeal, gestures performed in free space prove to be generally less efficient and more prone to fatigue than device-based input techniques. Adding guidance to input gestures increases, rather than decreases, accuracy. In accordance with the research literature, bimanual input techniques perform very well. Unimanual techniques perform honorably, and may still be considered in contexts of use where, for example, tools must be held in one hand to perform a domain/task specific action. A more surprising result is the generally higher efficiency of linear gestures when compared to circular, clutch-free gestures.

As future work, we plan to investigate how these pan-zoom techniques combine with other interaction techniques. Indeed, in real-world applications, users must also handle text entry, menu selection, copy and paste, drag and drop, and other activities. This implies trade-offs among techniques: a technique with optimal performance in this experiment may prove less easy to integrate with other techniques because of its requirements in terms of handedness or type of device. We have started to explore these questions in the context of real-world activities involving scientists visualizing and manipulating extremely large sets of multi-scale data.

ACKNOWLEDGEMENTS

We wish to thank Clément Pillias and Romain Primet, who helped implement the experiment on the wall-sized display, as well as Caroline Appert, Stéphane Huot and the anonymous reviewers for their feedback on drafts of this paper. We also wish to thank all the volunteers who participated in our experiments. This research was supported by a Région Île-de-France/Digiteo grant.

REFERENCES

1. C. Andrews, A. Endert, and C. North. Space to think: large high-resolution displays for sensemaking. In *Proc. CHI '10*, 55–64. ACM, 2010.
2. R. Balakrishnan and I. S. MacKenzie. Performance differences in the fingers, wrist, and forearm in computer input control. In *Proc. CHI '97*, 303–310. ACM, 1997.
3. R. Ball, C. North, and D. Bowman. Move to improve: promoting physical navigation to increase user performance with large displays. In *Proc. CHI '07*, 191–200. ACM, 2007.
4. P. Baudisch, N. Good, V. Bellotti, and P. Schraedley. Keeping things in context: a comparative evaluation of focus plus context screens, overviews, and zooming. In *Proc. CHI '02*, 259–266. ACM, 2002.
5. P. Baudisch, M. Sinclair, and A. Wilson. Soap: a pointing device that works in mid-air. In *Proc. UIST '06*, 43–46. ACM, 2006.
6. F. Bourgeois and Y. Guiard. Multiscale pointing: facilitating pan-zoom coordination. In *CHI '02 EA*, 758–759. ACM, 2002.
7. D. A. Bowman, D. Koller, and L. Hodges. Travel in immersive virtual environments: An evaluation of viewpoint motion control techniques. In *Proc. VRAIS '97*, 45–52. IEEE, 1997.
8. W. Buxton and B. Myers. A study in two-handed input. *SIGCHI Bull.*, 17(4):321–326, 1986.
9. B. Campbell, K. O'Brien, M. Byrne, and B. J. Bachman. Fitts' law predictions with an alternative pointing device (wiimote®). *Proc. Hum. Factors Ergon. Soc. Meeting*, 52(19):1321–1325, 2008.
10. X. Cao and R. Balakrishnan. Visionwand: interaction techniques for large displays using a passive wand tracked in 3d. In *Proc. UIST '03*, 173–182. ACM, 2003.
11. A. Cockburn, A. Karlson, and B. B. Bederson. A review of overview+detail, zooming, and focus+context interfaces. *ACM CSUR*, 41(1):1–31, 2008.
12. M. Czerwinski, D. Tan, and G. Robertson. Women take a wider view. In *Proc. CHI '02*, 195–202. ACM, 2002.
13. K. Evans, P. Tanner, and M. Wein. Tablet-based valuators that provide one, two, or three degrees of freedom. In *Proc. SIGGRAPH '81*, 91–97. ACM, 1981.
14. Y. Guiard. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *J. Motor. Behav.*, 19:486–517, 1987.
15. Y. Guiard and M. Beaudouin-Lafon. Target acquisition in multiscale electronic worlds. *IJHCS*, 61(6):875–905, 2004.
16. K. Hinckley, R. Pausch, J. Goble, and N. Kassell. A survey of design issues in spatial input. In *Proc. UIST '94*, 213–222. ACM, 1994.
17. R. Jacob and L. Sibert. The perceptual structure of multidimensional input device selection. In *Proc. CHI '92*, 211–218. ACM, 1992.
18. A. Leganchuk, S. Zhai, and W. Buxton. Manual and cognitive benefits of two-handed input: an experimental study. *ACM ToCHI*, 5(4):326–359, 1998.
19. R. Lindeman, J. Sibert, and J. Hahn. Hand-held windows: Towards effective 2d interaction in immersive virtual environments. In *Proc. VR '99*, 205–212. IEEE, 1999.
20. W. E. Mackay, C. Appert, M. Beaudouin-Lafon, O. Chapuis, Y. Du, J.-D. Fekete, and Y. Guiard. Touchstone: exploratory design of experiments. In *Proc. CHI '07*, 1425–1434. ACM, 2007.
21. S. Malacria, E. Lecolinet, and Y. Guiard. Clutch-free panning and integrated pan-zoom control on touch-sensitive surfaces: the cyclostar approach. In *Proc. CHI '10*, 2615–2624. ACM, 2010.
22. S. Malik, A. Ranjan, and R. Balakrishnan. Interacting with large displays from a distance with vision-tracked multi-finger gestural input. In *Proc. UIST '05*, 43–52. ACM, 2005.
23. D. McCallum and P. Irani. Arc-pad: absolute+relative cursor positioning for large displays with a mobile touchscreen. In *Proc. UIST '09*, 153–156. ACM, 2009.
24. M. Mine, F. Brooks, Jr., and C. Sequin. Moving objects in space: exploiting proprioception in virtual-environment interaction. In *Proc. SIGGRAPH '97*, 19–26. ACM Press/Addison-Wesley, 1997.
25. T. Moscovich and J. F. Hughes. Navigating documents with the virtual scroll ring. In *Proc. UIST '04*, 57–60. ACM, 2004.
26. T. Ni, D. Bowman, and J. Chen. Increased display size and resolution improve task performance in information-rich virtual environments. In *Proc. GI '06*, 139–146. Canadian Inf. Proc. Soc., 2006.
27. T. Ni, G. Schmidt, O. Staadt, M. Livingston, R. Ball, and R. May. A survey of large high-resolution display technologies, techniques, and applications. In *Proc. VR '06*, 223–236. IEEE, 2006.
28. A. Olwal, S. Feiner, and S. Heyman. Rubbing and tapping for precise and rapid selection on touch-screen displays. In *Proc. CHI '08*, 295–304. ACM, 2008.
29. E. Pietriga. A toolkit for addressing hci issues in visual language environments. In *Proc. VL/HCC'05*, 145–152. IEEE, 2005.
30. A. Roudaut, E. Lecolinet, and Y. Guiard. Microrolls: expanding touch-screen input vocabulary by distinguishing rolls vs. slides of the thumb. In *Proc. CHI '09*, 927–936. ACM, 2009.
31. D. Tan, D. Gergle, P. Scupelli, and R. Pausch. Physically large displays improve performance on spatial tasks. *ACM ToCHI*, 13(1):71–99, 2006.
32. D. Vogel and R. Balakrishnan. Distant freehand pointing and clicking on very large, high resolution displays. In *Proc. UIST '05*, 33–42. ACM, 2005.
33. E. Wherry. Scroll ring performance evaluation. In *CHI '03 EA*, 758–759. ACM, 2003.
34. B. Yost, Y. Haciahmetoglu, and C. North. Beyond visual acuity: the perceptual scalability of information visualizations for large displays. In *Proc. CHI '07*, 101–110. ACM, 2007.
35. S. Zhai, P. Milgram, and W. Buxton. The influence of muscle groups on performance of multiple degree-of-freedom input. In *Proc. CHI '96*, 308–315. ACM, 1996.
36. J. Zigelbaum, A. Browning, D. Leithinger, O. Bau, and H. Ishii. g-stalt: a chirocentric, spatiotemporal, and telekinetic gestural interface. In *Proc. TEI '10*, 261–264. ACM, 2010.

Typeset January 7, 2013