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
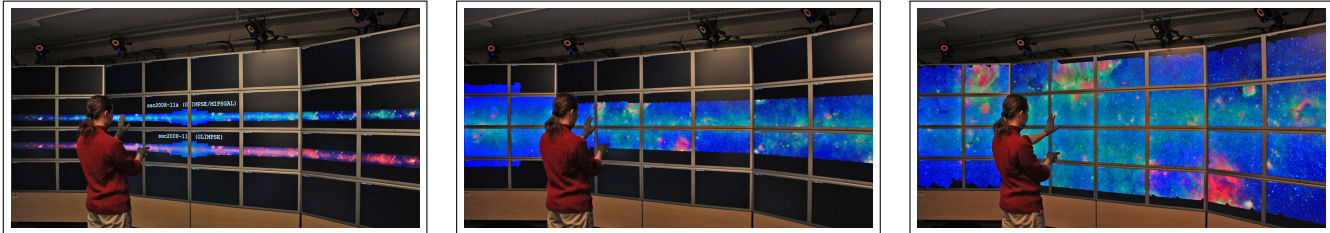F-91405 Orsay, France

[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.

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, in-

cluding 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[TM], 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

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

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.

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
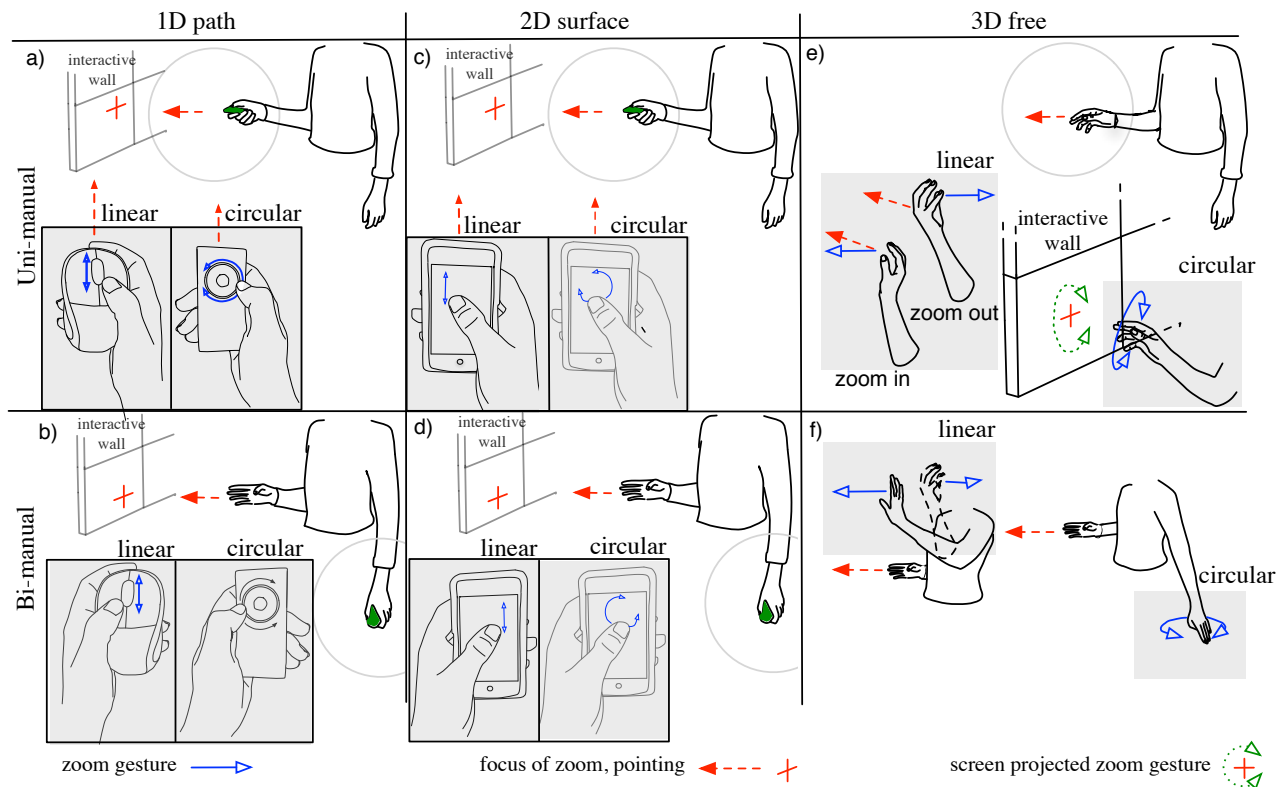
Figure 2. Matrix of the 12 techniques organized according to key characteristics: uni- vs. bimanual, degree of guidance, linear vs. circular gestures. 1D path involves guiding gestures along a particular path in space; in 2D surface gestures are made on a touch-sensitive surface; while in 3D free gestures 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 ∈ {*OneHanded*, *TwoHanded*}, GESTURE ∈ {*Circular*, *Linear*}, and GUIDANCE ∈ {*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
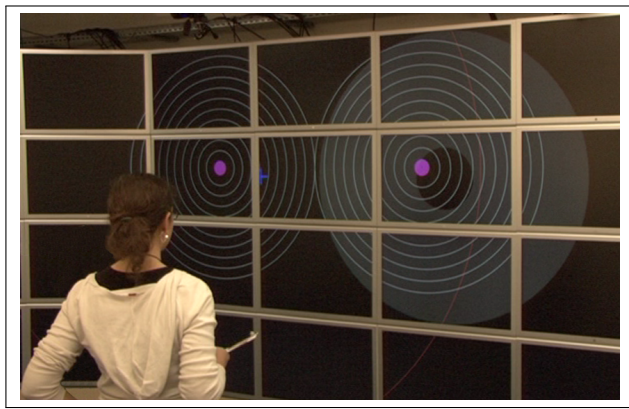
Figure 3. Participant performing the task

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 unam-
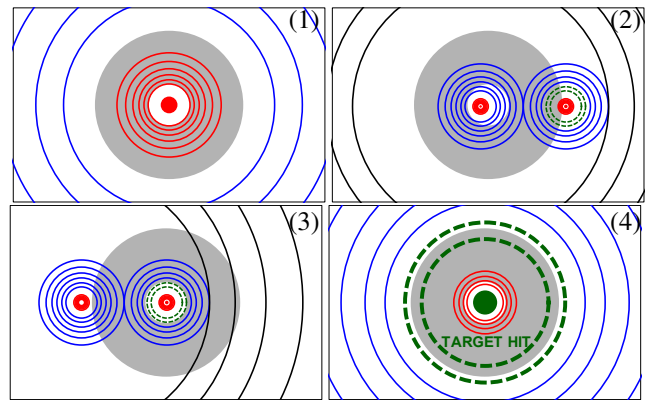

Figure 4. Task (schematic representation using altered colors): (1) Groups of concentric circles represent a given position and zoom level. (2) Zooming out until the neighboring set of circles appears. (3-4) Pan and zoom until the target (green inner disc and circles, dashed for illustration purposes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

biguously 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

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.

main factors. As the factors were counter-balanced, this created no adverse effects in the analysis.

Table 2 details results of the full factorial ANOVA for the model $MT \sim$ 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 ($\pm StdErr \times 1.96$).
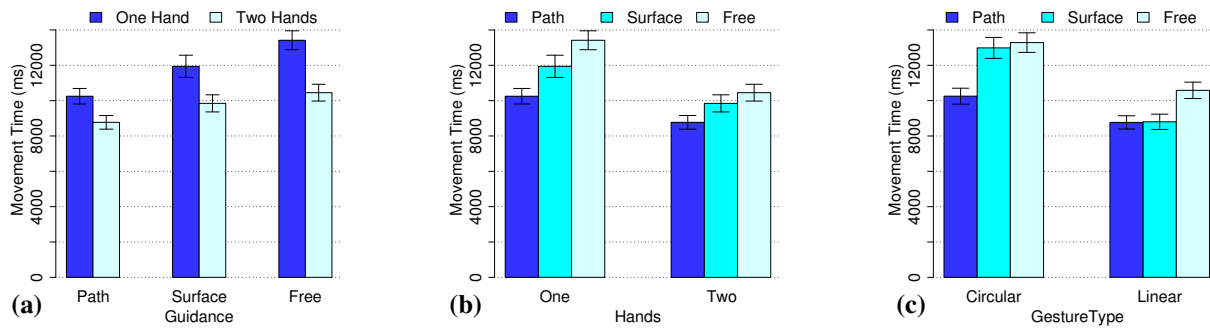
Figure 5. (a): *MT* per HANDS × GUIDANCE. **(b)** *MT* per GUIDANCE × HANDS. **(c)** *MT* per GUIDANCE × GESTURE.

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).
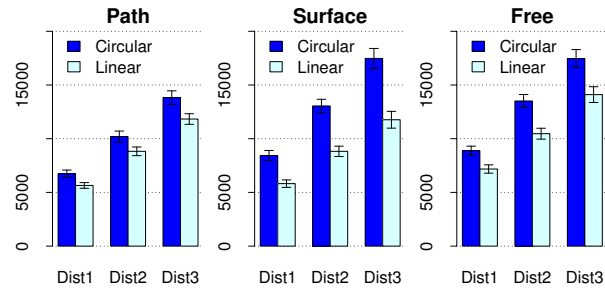

Figure 6. *MT* per DIST × GESTURE, **for each** GUIDANCE

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

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.