Effects of Display Size and Navigation Type on a Classification Task

Can Liu, Olivier Chapuis, Michel Beaudouin-Lafon, Éric Lecolinet, Wendy E. Mackay

HAL Id: hal-00957269
https://hal.archives-ouvertes.fr/hal-00957269v2

Effects of Display Size and Navigation Type on a Classification Task

Can Liu[1,2,3]   **Olivier Chapuis**[1,2]   **Michel Beaudouin-Lafon**[1,2]   **Eric Lecolinet**[3]   **Wendy Mackay**[2,1]

[1]Univ Paris-Sud & CNRS (LRI)
F-91405 Orsay, France

[2]INRIA
F-91405 Orsay, France

[3]Telecom ParisTech & CNRS (LTCI)
F-75013 Paris, France

ABSTRACT

The advent of ultra-high resolution wall-size displays and their use for complex tasks require a more systematic analysis and deeper understanding of their advantages and drawbacks compared with desktop monitors. While previous work has mostly addressed search, visualization and sense-making tasks, we have designed an abstract classification task that involves explicit data manipulation. Based on our observations of real uses of a wall display, this task represents a large category of applications. We report on a controlled experiment that uses this task to compare physical navigation in front of a wall-size display with virtual navigation using pan-and-zoom on the desktop. Our main finding is a robust interaction effect between display type and task difficulty: while the desktop can be faster than the wall for simple tasks, the wall gains a sizable advantage as the task becomes more difficult. A follow-up study shows that other desktop techniques (overview+detail, lens) do not perform better than pan-and-zoom and are therefore slower than the wall for difficult tasks.

Author Keywords

Wall-size display; Classification task; Physical navigation; Pan-and-zoom; Lenses; Overview+detail

ACM Classification Keywords

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

INTRODUCTION

Wall-size displays are becoming more common, [5, 24] raising the question as to whether existing research findings on desktop-size displays still apply to this new environment. Revisiting this question is even more essential as the technology evolves. Projection-based systems with low pixel density (typically 30 dpi) can now be replaced by tiled displays with the same pixel density as desktop monitors, i.e. about 100 dpi. Overall resolution (number of pixels) is thus multiplied by a factor of 10. This increased density, in turn, affords physical navigation. Users simply approach the screen to see detail and step back for an overview, similar to the pan-and-zoom navigation available on a desktop display. This raises the question as to the relative trade-offs between physical navigation with a wall-size display versus virtual navigation on a desktop.

Most previous research has addressed search, visualization and sense-making tasks. However, our observations of actual users during prototyping and real-world tasks shows that they want to *reorganize* data displayed on the wall: users move items around and group them in a way that is meaningful to the task at hand. We are therefore interested in tasks that require explicit data manipulation. Such tasks can be conducted on a desktop computer with multi-scale navigation techniques such as interactive overviews, but the need to manipulate data, e.g. with pick-and-drop, increases the complexity of the interaction. We need to better understand the benefits and drawbacks of these differently sized displays in an interactive context to develop guidelines that better inform their design.

Our challenge is how to design an abstract task that operationalizes the critical aspects of data manipulation in order to conduct controlled experiments that compare task performance for both wall-size and desktop displays. To increase internal validity, the task should reduce the cognitive load associated with the decision-making process and focus on actual data manipulation. To increase external validity, it should feature the same typical interactions found in real-world tasks. Finally, experimenters should be able to vary, in a controlled way, the difficulty of the task.

The rest of this paper provides a more detailed motivation for our work, including specific observations of users that highlighted requirements and influenced the design of the abstract task. We review the related work and then describe our design of an abstract classification task that matches the above requirements. We report the results of two controlled experiments that use this task to compare physical navigation on a wall-size display to three forms of virtual navigation on a desktop monitor. We conclude with a discussion of the main result, which found a robust interaction effect between display type and task difficulty: Although the desktop is often faster for simple tasks, the wall-size display performs significantly better with increased task difficulty.

MOTIVATION

We conducted a series of participatory design workshops with users, mostly scientists, to observe how they interacted with their own data on our ultra-high resolution wall-size display, and we also used the wall for our own real-word tasks. We briefly describe four typical examples of use.

Application Examples

An early use of the wall involved two program committee chairs assigning 145 submitted papers to 13 associate chairs (ACs). We divided the wall into columns, one per AC, plus a scratch area for the papers left to be assigned. The committee
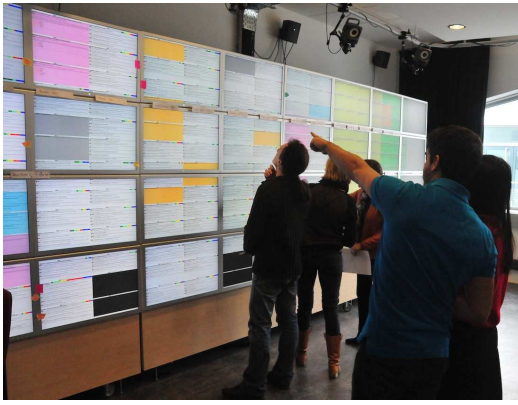
Figure 1. Scheduling a large conference on a wall-size display: Teams and individuals move close for detail or stand back for an overview.

chairs would pick one or more papers from the pile and assign it to an AC by moving it to the corresponding column, while taking conflicts of interest into account. During the process, undecided papers were left in the middle of the wall, to be assigned later. The piles of papers rising from the bottom created a sort of histogram, making it easy to determine the relative load for each AC. As the wall filled up, the chairs rearranged papers to optimize assignments and ensure each AC had a reasonable number of papers to review.

In another application, neuroanatomists wanted to display their collection of several hundred 3D brain scans. We prototyped an application that simultaneously displayed 64 high-resolution brain scans, each oriented in the same direction so that they could analyze the scans from the same angle. Since their goal was to compare, contrast and classify healthy and diseased brains, we provided the ability to rearrange the scans. They could drag and drop the brains into groups or place them side-by-side for easier comparison.

More recently, the wall display was used to schedule a large conference with almost 500 presentations in 200 sessions, organized into 13 parallel tracks over four days (Fig. 1). The scheduling task was highly constrained: related papers had to either be placed in the same session or else not in the same time slot, authors could not be in two places at once, "large draw" events had to be placed in large rooms, etc. Because of the high resolution of the wall, the complete program could be displayed, including titles, authors and a brief description. Various colors and labels served to visualize hard conflicts, softer constraints and inter-event affinities. The most common scheduling task consisted of identifying a misplaced or in-conflict presentation, finding a better slot, and moving it there[1]. Because the schedule was so heavily constrained, one move often triggered another and schedulers, often working in groups, had to juggle sequences of updates.

Finally, we are currently designing an application that lets users organize sequences of video clips on the wall. Each clip appears in one tile of the display. The user can play the clips, individually or in sequence, and reorganize them to try a

different edit. A variant of this application lets users organize a deck of slides as if the wall were a giant light table.

Implications

These applications share three elements: A complex decision-making task that relies on the users' expertise as well as their ability to quickly access the full content of the wall; a structured display in which information is logically organized in a grid; and a need to manipulate data by moving items from one cell to another. These applications are also well-suited to collaborative work. Indeed, we observed a range of collaboration patterns, from independent, parallel work to pairs working closely together.

The above examples illustrate how benefits of wall-size displays extend beyond visualization of extremely large images to include tasks requiring human judgment and the active manipulation of large datasets[2]. Here, manipulating content is an integral part of the task, whether to better understand the data set, to form an opinion, or to enact a decision.

The size of the dataset is the primary reason why users want to move off the desktop onto a wall-size display[3]. Users can see all of it at once, yet access details simply by walking toward the wall. Users take better advantage of their spatial memory since it is coupled with their physical movement in space. In contrast, the virtual navigation imposed by a desktop interface can be disorienting, and the overhead of constantly navigating the data set, e.g. with pan-and-zoom, can distract users and increase their cognitive load. Even so, physical locomotion is more time-consuming and tiring than virtual navigation, and manipulating data with well-known devices and widgets may be more efficient than using mid-air techniques on a wall-size display [18]. Our goal is to delve deeper into the questions raised by our observations and systematically evaluate the advantages and disadvantages of manipulating data on ultra-high resolution wall-size displays.

RELATED WORK

Within the extensive body of literature on large displays, we focus on studies of the effects of display size on various tasks. We first consider studies conducted in desktop settings, then move on to those involving physical navigation and finally review relevant research in multiscale interfaces.

Large Displays in Desktop Settings

Previous work has demonstrated the benefits of larger displays for traditional desktop tasks. For instance, Czerwinski et al. [10] observe higher productivity and satisfaction with larger display surfaces when performing complex daily tasks. Bi and Balakrishnan [7] compare a large projected wall display with single and dual desktop monitors. Their results suggest that large displays facilitate tasks with multiple windows and rich information because they offer a more immersive experience, with enhanced peripheral awareness.

In sense-making tasks, high resolution displays not only display more information but also provide a virtual space where

---

[1]A separate application based on affinity matching [16] detected conflicts and suggested alternative moves.

[2]Relative to human cognitive skills, i.e. up to a few hundred items.
[3]The other (correlated) reason is to enable group work, but this is beyond the scope of this work.

meaning is encoded in the spatial relationship between data, documents, the display, and the user [1]. Similarly, increasing display size and resolution both improve user performance in rich-information environments [20].

Tan et al. [26] demonstrate that large projection displays support spatial orientation tasks better than desktop monitors, and argue that these more immersive environments encourage egocentric rotations, leading to improved performance. Similarly, Czerwinski et al. [11] found gender differences on the effect of the field of view in a spatial task.

Although these studies consistently show the benefits of larger displays, most were conducted in traditional desktop settings where users sit before a monitor, with limited or no physical locomotion.

Physical Navigation with Wall Displays
As display size and pixel density increase, standing and moving in front of large displays becomes necessary. Ball et al. [4] show that larger displays promote physical navigation and improve user performance for search, navigation and pattern-finding tasks. However, their tasks do not involve data manipulation, and they do not include a desktop condition for comparison. Ball and North [3] investigate the key advantages of large displays and find that physical navigation is more important than increasing the field of view. Yost et al. [27] show that user performance improves with larger displays despite the need for physical navigation.

Large displays also affects perception. For example, Endert et al. [12] demonstrate the impact of visual encodings on physical navigation on large displays and show that physical navigation improves user performance. Bezerianos and Isenberg [6] find that the perspective distortion that occurs when users do not look in front of them on the wall affects the accuracy of their judgments of relative size and orientation, and that physical movements rarely improve the situation. In summary, these studies show the benefits of physical navigation in some situations for certain tasks, none of which feature data manipulation.

Multiscale Interfaces and Display Size
Multiscale interfaces [9] were designed to visualize large quantities of data on displays that are too small. With few exceptions [18], multiscale interfaces have been deployed and studied on the desktop, presumably to obviate the need for large displays. However, the effect of display size on multiscale navigation has yet to be adequately investigated.

Guiard et al.'s [13] comparison of small-to-medium display sizes for a pan-and-zoom target acquisition task shows a minor performance improvement for the larger display. Jakobsen and Hornbæk [15] evaluate the usability of interactive visualization techniques (overview+detail, focus+context, pan-and-zoom) with three display sizes. Surprisingly, the large display is not always faster, and is sometimes slower than the medium display. The authors suggest that some techniques require increased target searching time on the large display.

These results suggest only a small or no benefit of large displays when using multiscale navigation techniques. However,

these studies were conducted in desktop settings where users sit in front of the display, with tasks that involve only visualization or target acquisition.

Summary and Approach
Despite the growing literature examining the effects of display size and physical navigation, data manipulation tasks have been largely ignored. In fact, we know of no study of such tasks with very large displays where users were standing and moving in front of the display.

Our goal is to build upon previous work to improve our understanding of the trade-offs between wall-size displays and physical navigation on the one hand, and desktop monitors and virtual navigation on the other, for data manipulation tasks. We must identify which input techniques are most appropriate for each setting and construct an abstract data manipulation task that captures the essential elements of the real-world tasks we observed.

ABSTRACT CLASSIFICATION TASK
We chose a simple classification task in which users partition a set of items into classes depending upon their properties. On the desktop display, items from the same class are grouped together into containers, either freely, as when comparing brain scans, or constrained, as when assigning papers. The latter involves explicit containers (columns) to represent each class. The limited capacity of each container adds a constraint, turning this into a resource allocation task. Scheduling tasks are even more complex because they add more constraints, such as avoiding conflicts across parallel sessions.

We sought a middle ground between the simpler and more complex examples. Our task has more containers than classes. Users place like items into containers without letting the container overflow. For example, conference talks (items) must fit into sessions (containers) of limited capacity, but several sessions can have the same theme (class).

To classify the items, we need to know when two items are in the same class. In practice, such decisions are domain-specific, e.g., two conference papers on the same topic or two brains with similar features, and often require expert judgment or incur a heavy cognitive load. In order to properly control this aspect of the task, we needed to find a simple, easily tested relationship, that is well-known to participants. Our solution is to represent each class by a different letter.

We must also determine how much information to display about an item, so users can determine whether two items are in the same class. For example, the conference scheduling task displays the title, keywords and abstracts of the papers whereas the neuroanatomy application displays high-resolution images of the brain. Our experiment uses the simplest solution i.e. we operationalize information density by adjusting font size.

We control the complexity of our classification task via several parameters: number of items, number of classes, number of containers, and representation of the item, including the label font size. These factors define a rich yet easy-to-control design space for experimental tasks based on the abstract task.
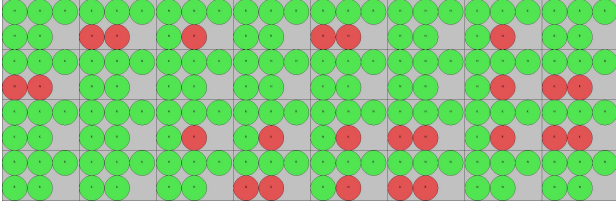
Figure 2. Display layout at the beginning of a trial (in the *Large-Easy* condition). Each cell has the size of one of the 30" monitors of the wall-size display and disks representing items have a diameter of about 20cm.

Experimental Task

For the experiment task, the number of items is set to $32 \times 5 = 160$, the number of containers is 32, and the maximum number of items in a container is 6. Containers are organized in a 8×4 matrix that matches the tiles on the wall. Items are depicted as discs whose diameter is such that 6 discs fill up one container (Fig. 2).

We control difficulty by setting the number of classes. We use two levels for the DIFFICULTY factor: *Easy* (2 classes labeled "C" and "D") and *Hard* (4 classes labeled "H", "K", "N" and "R"). These letters were chosen according to the BS 4274-1:2003 vision test standard [8] to guarantee equal legibility. Since the items display the name of their class, the similarity criterion is very simple: similar items have the same label.

Label size affects legibility at a distance and thus influences the level of physical or virtual navigation required to be able to read a label and make a decision. We use three levels for the LABELSIZE factor: *Small* is a standard computer font (12pt font, letter size about 1.8x2.3mm), *Medium* is twice the small size, *Large* is such that characters have the same size as the *Small* size when the whole scene is scaled down to the size of the desktop monitor used in the experiment (letter size about 15.5x20mm or 100pt).

Finally, we simplify the task by automatically coloring improperly classified items in red. More precisely, when a majority of the items in a container are of the same class, we color these green and the others red. This makes it easy to spot which items are left to be classified and also gives participants a clear goal: "Make everything green".

The task consists of moving disks between containers so that each container holds disks of the same class. Disks are moved with a pick-and-drop interaction: Clicking on a disk picks it up and attaches it to the cursor, a second click drops it into a container, unless the container is full, in which case the disk snaps back to its original container. On the desktop, participants can pan and zoom the scene in order to read the labels, find the target container, and identify which item to move. On the wall-size display, users stand or walk in front of the wall and use a tablet to control the cursor (Fig.3).

Since solving the task with a random configuration takes a long time, we generate initial configurations in which some items are already classified. This not only reduces the time needed to solve each task, but it also has ecological validity, since in the real-world tasks we observed, participants built upon an initial classification made by others or a computer had generated an initial pass for which some errors remain.

Figure 3. Wall condition: Using a trackpad to move a red disk into a container on the wall-size display.

Each task begins with a configuration of 24 incorrectly classified items (Fig. 2) that are evenly distributed across classes but randomly distributed among the containers: 8 containers hold 2 incorrect items and 8 other containers hold 1 incorrect item. To ensure that tasks are of equal difficulty, we generate random layouts and select those that satisfy the following constraint: The average distance[4] between a red disk (incorrectly classified item) and the closest suitable container is between 1.25 and 1.46 for *Easy* tasks and between 2.5 and 2.7 for *Hard* tasks.

We create one layout per LABELSIZE × DIFFICULTY condition. To minimize the effects of different layouts on performance, we create the layouts for the other conditions by permuting the labels and by applying horizontal, vertical or central symmetry. This results in structurally similar but visually different layouts, enabling us to create counter-balanced sets of tasks within and across participants.

EXPERIMENT 1: WALL VS. DESKTOP

Our goal is to investigate the trade-offs between physical and virtual navigation and how they affect task performance. We use the above classification task to compare the performance of a high-resolution wall display (condition *Wall*) with a desktop computer (condition *Desktop*). Based on our review of the literature and our experience using the wall-size display, we formulate three hypotheses:

- *Wall* performs better than *Desktop* for smaller labels;
- *Wall* performs better than *Desktop* for harder tasks;
- *Desktop* performs better than *Wall* for larger labels and simpler tasks.

Participants

We recruited 12 volunteers (5 female), aged 20 to 30, all with normal or corrected-to-normal vision. Seven use a trackpad daily; two had never used one.

Apparatus

In the *Wall* condition, we use a 5.5m × 1.8m wall-size display made of an 8×4 matrix of 30" LCD panels for a total resolution of 20480×6400 pixels (Fig. 3). A cluster of 16 Apple Mac Pro computers running Mac OS X communicate via a dedicated high-speed network and are controlled by a similar

---

[4]Euclidean distance between containers, where a unit is the size of a container (the distance between two adjacent containers is 1).

front-end computer. A VICON motion-capture system tracks the 3D positions of infrared retro-reflective markers attached to a hat worn by the participants with 1mm accuracy.

A 13×13cm Apple Magic Trackpad (weight 165g) controls the cursor of the front-end computer via Bluetooth. The computer displays a scaled-down image of the scene displayed on the wall and maps the cursor position to the wall cursor. Since the targets, i.e. the disks, are large, this technique provides sufficient precision for pointing. Users start and end the pick-and-drop actions by a simple tap on the trackpad.

For the *Desktop* condition, we use the same type of workstation and display as the wall (30" LCD panel, 2560×1600, 100 dpi). We use an Apple Mighty Mouse with default acceleration for input, with a wheel to control zooming. A pilot study suggested that this mouse was better suited to the *Desktop* than the Magic Trackpad used in the *Wall* condition.

The use of a different input device for the *Wall* and *Desktop* conditions is meant to maximize external validity. Since there is no standard input device for wall displays, we choose a trackpad based on previous work [17] and on our own experience and tests. For the desktop, we choose the most well-known input device, the mouse, to ensure that any results in which the wall outperforms the desktop cannot be attributed to an unusual or suboptimal desktop input device.

The experiment software is implemented using jBricks [22], a Java toolkit that supports applications running both on a cluster-driven wall display and on a regular desktop.

Procedure
The experiment is a [2×3×2] within-participants design with three factors:

- DISPLAY: display type, *Wall* or *Desktop*;
- LABELSIZE: label size, *Large*, *Medium* or *Small*;
- DIFFICULTY: number of classes, *Easy* (2 labels) or *Hard* (4 labels).

Prior to the study, participants take a vision and color-blindness test to ensure normal vision. Participants read a standard explanation of the task and perform an initial four-trial training session prior to each display condition. Participants are told to complete the tasks as quickly as possible but to avoid dropping items into the wrong containers, to discourage a trial-and-error strategy.

The experiment is split into two sessions, one per DISPLAY. Half the participants start with *Wall*, the other half with *Desktop*. The order of the DIFFICULTY and LABELSIZE conditions are counterbalanced across participants using Latin Squares. To minimize the potential order effect between DISPLAY conditions, we use the same sequence of trials and symmetric layouts for each participant between the *Wall* and *Desktop* conditions. The experiment lasts about one hour.

Data collection
We collected 288 measured trials (2 DISPLAY × 3 LABELSIZE × 2 DIFFICULTY × 2 replications x 12 participants). We measure Task Completion Time, *TCT*, and count the number of pick-and-drop actions to complete each trial. We also log kinematic data about cursor movements and user navigation,
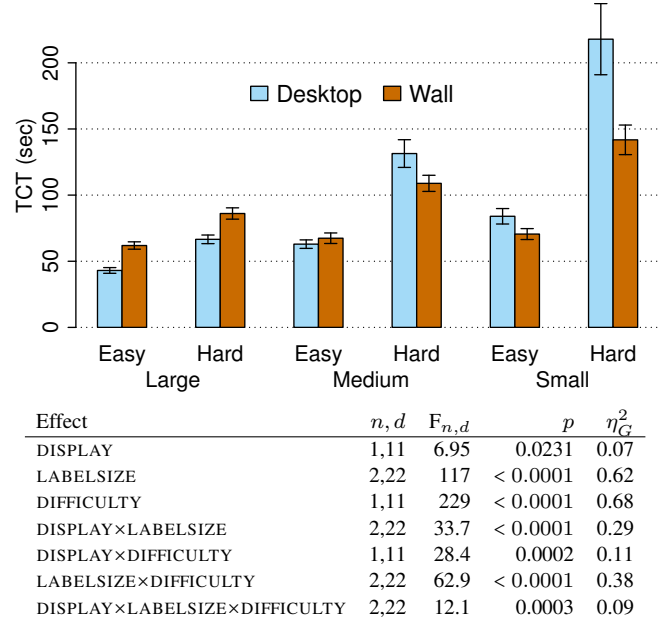

Effect	n, d	$F_{n,d}$	p	η_G^2
DISPLAY	1,11	6.95	0.0231	0.07
LABELSIZE	2,22	117	< 0.0001	0.62
DIFFICULTY	1,11	229	< 0.0001	0.68
DISPLAY×LABELSIZE	2,22	33.7	< 0.0001	0.29
DISPLAY×DIFFICULTY	1,11	28.4	0.0002	0.11
LABELSIZE×DIFFICULTY	2,22	62.9	< 0.0001	0.38
DISPLAY×LABELSIZE×DIFFICULTY	2,22	12.1	0.0003	0.09

Figure 4. Top: Task completion time (*TCT*) for each condition. Bottom: full factorial ANOVA with participant as random factor.

i.e. physical navigation using the tracking system for the *Wall* condition, panning and zooming for the *Desktop* condition. Participants fill out a questionnaire about their subjective work load at the end of each DISPLAY condition and another about their preferences at the end of the experiment.

Results

Task Completion Time
An analysis of outliers showed that 95% percent of the trials were within 15% of the mean completion time per condition. Three trials were more than 20% slower. We kept all trials in the analyses below. The Shapiro-Wilk normality test showed evidence of non-normality only for the *Desktop-Large-Easy* condition, with two participants being very slow. This does not affect the results described below.

Fig. 4 highlights the differences between *Wall* and *Desktop* for task completion time[5]. It also displays the results of the full factorial ANOVA for the model *TCT* ~ DISPLAY×LABELSIZE×DIFFICULTY×Rand(Participant). All main and interaction effects are significant but with varying effect sizes[6]. While the effect sizes of DISPLAY and the triple interaction are moderate, others can be considered large.

Focusing our analysis on the effect of DISPLAY and given the significant interaction effects, we compare *TCT* for *Wall* and *Desktop* for all LABELSIZE×DIFFICULTY conditions with a t-test (Bonferroni correction, $n = 6$). We find that:

- For *Large* labels, *Desktop* is faster than *Wall* for both *Easy* ($p < 0.0001$, 30.5% faster) and *Hard* ($p = 0.0001$, 17.1% faster);

---

[5] All bargraphs display the average of all trials per condition, with the error bars showing the corresponding confidence intervals.
[6] The η_G^2 statistic measures effect size. 0.02 is considered a small effect size, 0.13 medium and 0.26 large. However, Bakeman [2] recommends that each field develop their own guidelines.

- For *Medium* labels there is no significant difference for *Easy* ($p = 1$), but *Wall* is faster than *Desktop* for *Hard* ($p = 0.0222$, 22.7% faster);
- For *Small* labels, *Wall* is faster than *Desktop* for both *Easy* ($p = 0.0315$, 16.0% faster) and *Hard* ($p = 0.0059$, 34.9% faster).

As predicted, the desktop is faster for *Large* labels and the wall is faster for *Small* labels, in both DIFFICULTY conditions. However, we also see that the magnitude of the difference depends on DIFFICULTY. The wall shows a larger advantage for *Hard* tasks, while the desktop performs better for *Easy* tasks. In the *Medium* condition, the wall and desktop perform similarly in the *Easy* condition, but the wall is faster in the *Hard* condition.

We now turn our attention to DIFFICULTY and LABELSIZE and observe the expected effects (Fig. 4): *TCT* grows with higher DIFFICULTY and smaller LABELSIZE. The effect of smaller LABELSIZE is more important for the difficult task (hence the LABELSIZE×DIFFICULTY interaction). However, for the *Easy* task on the wall, the task completion times are very close across the three label sizes. To confirm this observation we compare *TCT* for the three LABELSIZE for all DISPLAY×DIFFICULTY conditions with a t-test (Bonferroni correction, $n = 12$). All differences are significant ($p < 0.005$) except for *Small* vs *Medium* and *Medium* vs *Large* in the *Wall-Easy* condition. This suggests that, for the wall, label size does not affect performance as much for easy tasks.

In contrast, the absolute difference between *Wall* and *Desktop* in the *Small-Hard* condition is large, with the wall being about 35% faster. This supports our hypothesis that complex tasks become intractable on the desktop but still manageable on the wall-size display.

Number of Pick-and-Drop Actions
All participants were able to solve the task in an almost optimal number of steps: while all configurations can be solved in 24 steps, participants performed 25.3±0.20 pick-and-drop actions on average, with no significant difference between *Wall* and *Desktop* overall. The only significant result is that participants performed more actions in the *Small-Hard* condition (27.0±1.43) than on any other LABELSIZE×DIFFICULTY conditions (24.91±0.15).

Movements of the Virtual Viewpoint and of the Participants
The wall and desktop displays have the same pixel density and render the exact same scene. Moreover, the maximum scale available when zooming in on the desktop displays the scene at the exact same size as on the wall. We can therefore compare the physical movements of participants in front of the wall with their virtual movements on the desktop.

For the wall, we compute the length of the participants' path from the tracking data. For the desktop, we compute two measures of the length of the movements of the viewpoint by considering the following two spaces. (i) *Scene space* is the space of the rendered scene at the maximum level of detail; panning movements are scaled by the current zoom factor in order to match the equivalent wall physical navigation. (ii) *Screen space* is the space of the desktop screen where panning movements are not scaled according to the zoom factor,
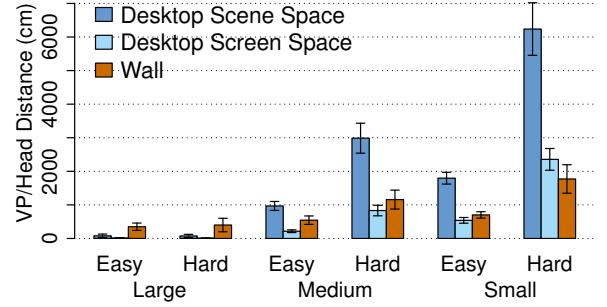

Figure 5. Average distance traveled by participants for each DISPLAY **condition. (See the text for the definition of scene vs. screen space).**

in order to better match the mouse physical actions. Both measures are converted to centimeters for comparison.

Fig. 5 shows the average distances covered by the participants according to these measures. Fig. 6 and 7 show the actual trajectories for one participant. With large labels, no navigation is needed to perform the task, and indeed almost no viewpoint movement occurs in the *Desktop* condition. However, most participants did move in front of the wall (about 482cm per trial on average), with no evidence that this depends upon DIFFICULTY. For the other conditions, the amount of movement increases significantly both with smaller LABELSIZE and higher DIFFICULTY. These differences correlate with the differences in task completion time. In particular, viewpoint movements increase sharply for *Small-Hard* (in scene space).

For *Small* and *Medium* sizes, in scene space, the length of the virtual navigation is longer than that of physical navigation (Fig. 5). This is not surprising, and can be attributed to the users' ability to move their head and eyes [4]. For example, the position of the head has very low variability in the y dimension, between 5.4±2.0cm for the *Medium-Easy* condition and 15±11cm for the *Small-Hard* condition, while the viewpoint has a larger amplitude (Fig. 7, bottom row): from 101±33cm for the *Medium-Easy* condition to 164±33cm for the *Small-Hard* condition.

However, the distance in *screen* space (Fig. 5) for the desktop is shorter or close to that of the wall. This indicates that virtual navigation competes with physical navigation in terms of distance covered in motor space, and therefore the difference in performance between *Wall* and *Desktop* for difficult tasks must have another explanation.

Physical vs Virtual Reach
To complement our analysis of participants' movements, we now look at their ability to interact with distant targets. Indeed, the larger size of the wall-size display enables users to reach targets at a distance without moving while on the desktop they must pan and/or zoom the scene. We did not use an eye-tracker to collect accurate data of where users were looking, however we can calculate the distance between the user and the target when they pick or drop an item. Below we report results at pick time. Results at drop time are similar.

For the *Wall* condition, Fig. 8 plots the cursor positions at pick time, relative to the orthogonal projection of the position of the head. For the *Desktop* condition, Fig. 9 plots the cur-
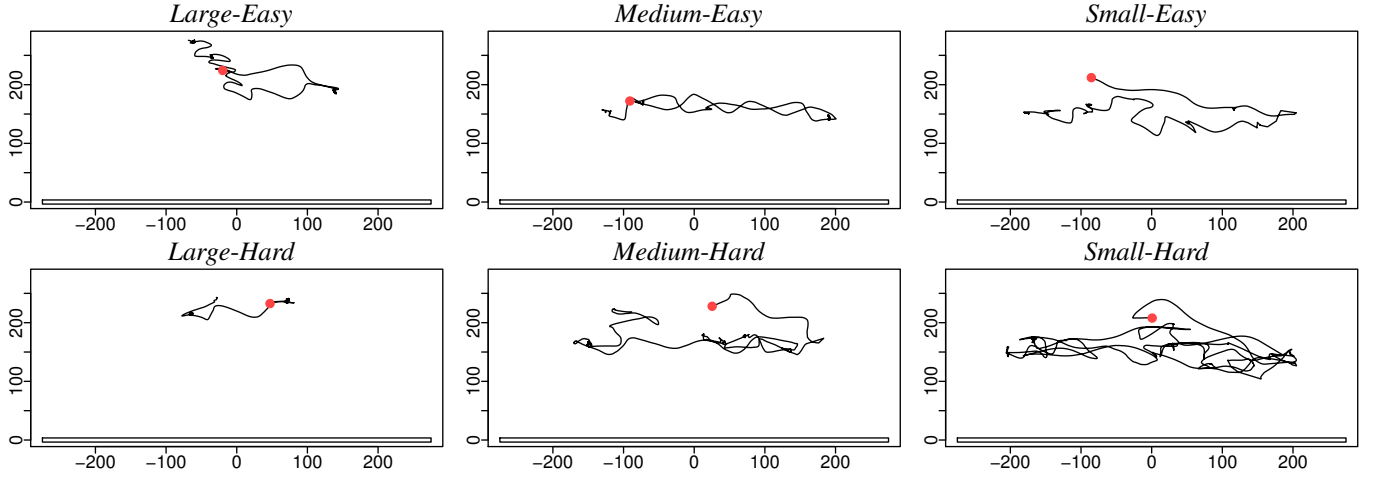
Figure 6. Movements of the head of participant P06 in front of the wall for the first measured block for each LABELSIZE×DIFFICULTY **condition. This is a bird's eye view with the wall at the bottom of each graph. The axes represent the distance of the head to the wall in centimeters. These graphs are consistent with those reported in Ball et al. [4].**
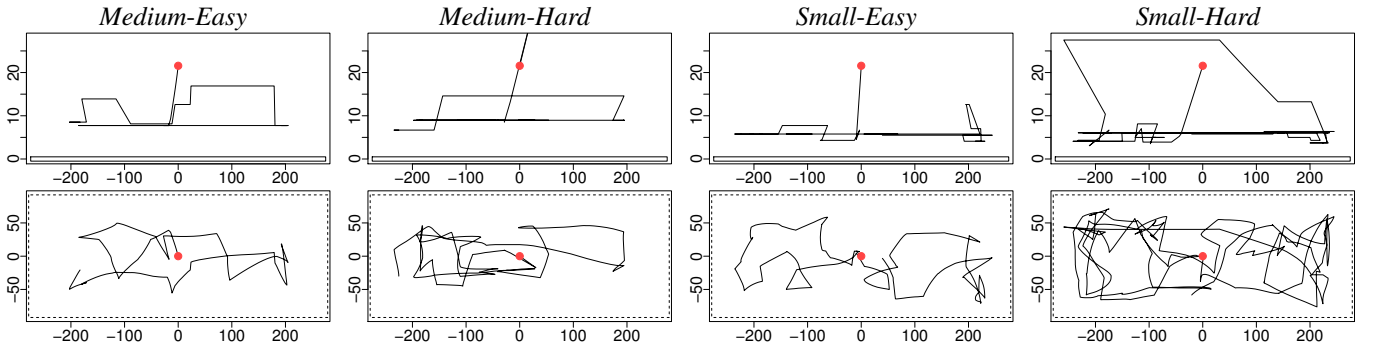

Figure 7. Movements of the viewpoint of participant P06 in scene space (*Desktop*** condition). The top graphs plot the trajectory of the virtual navigation from a bird's eye view, as in the figure above (zoom factor is converted to a distance on the *y* axis). The bottom graphs plot the trajectory from a front view of the display (orthogonal projection of the viewpoint on the scene). All distances are in centimeters.**
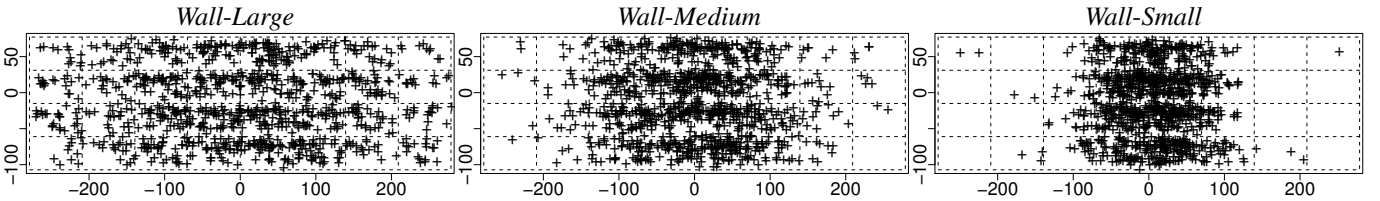

Figure 8. Positions of the cursor at pick time relative to the orthogonal projection of the head of the participants on the wall for each LABELSIZE **condition (all participants, in centimeters). The dotted lines show the size of the wall for reference. The wall is at this position only when the user is in front of the center of the wall. Note that, in the *Large* condition, about 20 points are outside the wall boundaries and are thus not shown in the figure. These correspond to pick actions where the distance between the projection of the head and the cursor was greater than half the wall width, e.g. picking a disk on the left side of the wall while the user is on the right side.**
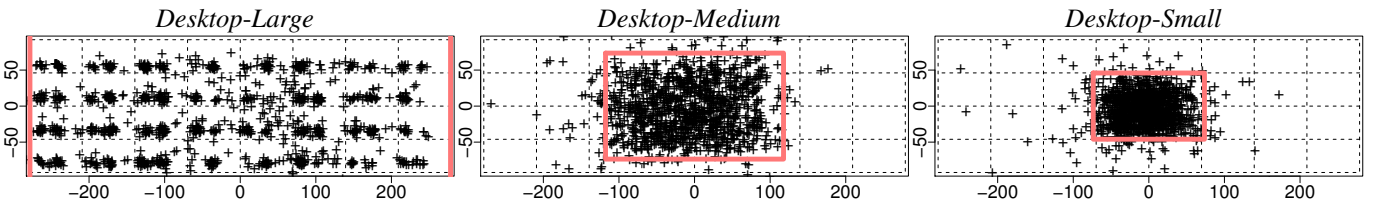

Figure 9. Positions of the cursor at pick time relatively to the position of the center of the view (translated to the origin of the graphs) for each LABELSIZE **(all participants, in centimeters). The red rectangles show the average size of the scene that was displayed on the desktop at pick time. The dotted lines show the size of the scene and containers for reference. The scene is at this position only when the center of the view is at $(0, 0)$, which is almost always the case for *Large* since no panning is needed.**

sor positions at pick time, relative to the center of the view. With *Medium* and *Small* labels, the points are more closely clustered for the desktop than for the wall, indicating that the participants' reach is larger on the wall. Indeed, while on the desktop users must bring the target into view with pan-and-zoom, they can act at a larger distance on the wall, reducing the need for navigation. On the other hand, with *Large* labels no navigation is needed for the *Desktop* nor the *Wall*. However the *Wall* requires more head movements, which might explain why the desktop is faster.

Fig. 9 also shows the average size of the area of the scene that was displayed on the desktop display. For *Small* labels, it is close to 4 containers (2×2). This explains the performance differences between *Easy* and *Hard* tasks for the *Desktop* condition. In the *Easy* condition, most misplaced items can be moved to an adjacent container while in the *Hard* condition, they often need to be moved to a container further away, requiring the participant to pan-and-zoom during the pick-and-drop action. With *Medium* labels, the average displayed size is about 9 containers (3×3), which reduces the chance that the destination of the move was out of sight, thus reducing virtual navigation. This is confirmed by the table below: the average number of pan and zoom actions during pick-and-drop more than doubles between the *Easy* and *Hard* conditions, albeit with large variability, probably due to different participant strategies.

	Easy-Medium	Hard-Medium	Easy-Small	Hard-Small
number of pan	14.4±5.2	42.5±18	30.6±11	97.8±29
number of pan in pick-drop	9.71±3.4	18.8±7.6	21.2±5.8	42.0±19
number of zoom	5.46±2.4	12.2±7.3	17.0±6.2	47.2±23
number of zoom in pick-drop	4.92±2.2	8.54±4.3	14.0±4.7	30.9±20

Angular Size of the Labels

In our design, LABELSIZE operationalizes information density: smaller text size forces participants to get closer to the display through either physical or virtual navigation in order to make an informed decision. We compute the angular sizes of labels when an item is picked, taking into account the perspective distortion due to the view angle of view. On the wall, we use the tracking data; on the desktop, we assume a 60cm distance between the participant's eyes and the display.

The table below shows the average angular width of the labels at pick time in arcminute for the *Medium* and *Small* labels:

DISPLAY	Easy-Medium	Hard-Medium	Easy-Medium	Hard-Medium
Wall	6.78 ± 0.37	6.74 ± 0.32	3.90 ± 0.16	4.20 ± 0.17
Desktop	5.78 ± 0.26	6.27 ± 0.25	4.22 ± 0.17	4.76 ± 0.24

These values are close to 5 arcminute, corresponding to 20/20 vision acuity. The differences between *Wall* and *Desktop* are small, suggesting that participants optimized their physical and virtual navigation according to their visual acuity.

Qualitative Results

At the end of each session (*Desktop* or *Wall*) we asked participants to rate their degree of fatigue, mental load and frustration while performing the task (Fig. 10). We use pairwise Wilcoxon rank sum tests with Bonferroni corrections to test
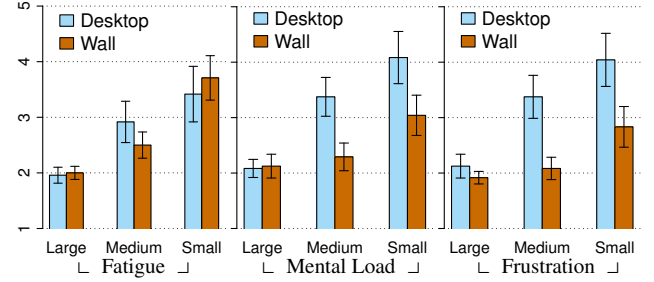

Figure 10. Physical fatigue, mental load and frustration on a five-point Likert item (1 is best, 5 is worst) for each LABELSIZE condition.

significant differences among DISPLAY×LABELSIZE conditions and focus on the differences between *Desktop* and *Wall*.

The results on fatigue are not significant ($p \geq 0.5$), in contrast to previous studies that found wall-size displays to be more tiring than desktop settings. Most participants found the small labels with four letters (hard task) tiring, in both conditions. One participant said: "*Desktop's repetitive work was somehow tiring. However the wall was very tiring after a while. If I could perform the wall task by resting my hands on a desk it would be ideal.*"

We find no significant differences for *Large* labels ($p \geq 0.9$) on mental load and frustration. However, for medium and large labels, the *Wall* causes significantly lower subjective mental load ($p = 0.0007$ for *Medium*, $p = 0.02$ for *Small*) and frustration ($p < 0.0001$ for *Medium*, $p = 0.01$ for *Small*). A few participants mentioned memory load: "*with the small labels, it was more difficult to get a mental map of the layout.*"

Fig. 11 summarizes participants preferences between the desktop and the wall. Except for *Large* labels, almost all participants preferred the wall. These results are stronger than the quantitative measure of performance, where the *Medium* label sizes performed about the same in both environments. This may be due to the novelty effect of using a wall-size display as well as other factors yet to be identified, including spatial memory.

11 out of 12 participants tried to remember the positions of the items and/or containers, and 7 of them commented that it was easier to remember the positions of the items when they were in front of the wall: "*...because I remembered the spatial location (in the room) of some particular rectangles*"; "*I have better vision with the wall. It was more fun standing up and walking. It was also easier to remember where to go because of the movement memory.*" This is consistent with the participants' description of their strategies: most participants
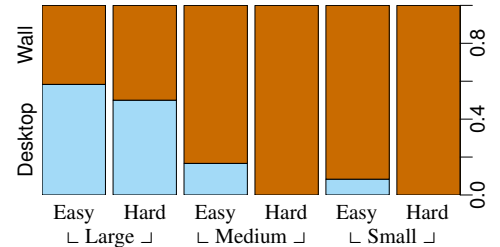

Figure 11. Proportion of the participants that preferred *Desktop* (■) or *Wall* (■) for all LABELSIZE×DIFFICULTY conditions.

tried to pick and drop locally in the easy conditions. In the hard conditions, they tried to remember the positions of the misplaced items and of the containers to reduce navigation.

Some participants explain the interaction effect between LA-BELSIZE and DISPLAY: "*The desktop with large labels is very fast, but exploring small and medium labels is painful*"; "*With the small and four letters in the wall, I didn't have to pan and zoom all the time, which was tiring. I just had to move a little bit, which was fine.*" Other comments reflect the different sense of engagement between the desktop and the wall, e.g., "*For the desktop, I use the mouse [...] I feel I am under control. For the wall, I can move around, I feel I am a part of the interaction, and I feel I am controlling everything.*"

In summary, these results show a robust interaction effect between display type and overall task difficulty (label size and number of classes), with the wall up to 35% faster in the hardest condition. This difference can be attributed to the ability to use more efficient strategies on the wall, as evidenced by the larger reach of users. Other factors are likely at work, though, such as a better use of spatial memory.

EXPERIMENT 2: THREE DESKTOP TECHNIQUES
Experiment 1 showed a strong performance advantage of physical navigation on a wall-size display when compared with pan-and-zoom navigation on a desktop interface for difficult classification tasks. Could these results be different with other types of virtual navigation? To test this hypothesis, we compared three desktop techniques in a second experiment: the baseline pan-and-zoom technique, an overview+detail technique and a focus+context technique [9].

Overview+detail adds a miniature view of the scene (the overview) displayed in a corner of the main view (the detail view). Many implementations let the user move the detail view by interacting with the overview, e.g. by dragging the rectangle that represents the detail view in the overview. The literature suggests that adding an overview to a pan-and-zoom interface increases user satisfaction [14, 19] and that an interactive overview can be very efficient for search tasks [21].

We tested an interactive overview but found that it slowed users down. The switching cost between views was too high when they performed pick-and-drop actions. This shows how a data manipulation task can affect the usability of a technique that has been tested only for search or visualization tasks. Instead, we chose to test a *PZ+OV* technique, which adds a passive overview in the lower-right corner of the screen, with a rectangle showing the current position of the detail view.

Lenses [25] are another way to combine focus and context in a single view. We implemented a fisheye lens that is permanently attached to the cursor and has the same radius as the disks. The entire scene is scaled down to fit the display and the lens has a magnifying factor of 6, making the small labels readable. To avoid occlusion during pick-and-drop, the disk being picked is attached to the bottom of the lens.

Method
We recruited 12 volunteers (6 female), aged 22 to 38, all with normal or corrected vision. Half had participated in Experi-
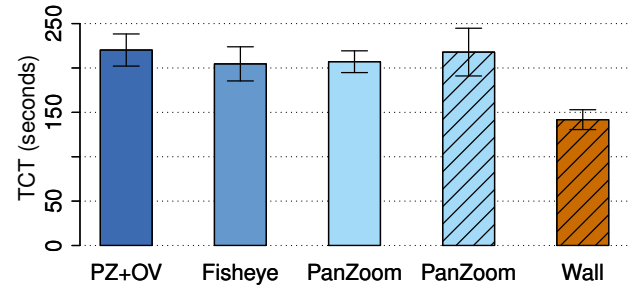

Figure 12. Task completion time (*TCT*) for the three techniques in Experiment 2. The hashed bars on the right show the results from the corresponding condition of Experiment 1.

ment 1. We use the same desktop apparatus as in Experiment 1, and the task is the same as in Experiment 1. Since we want to test whether desktop techniques can beat the wall-size display, we use a single task, corresponding to the *Small−Hard* condition of the first experiment. We also use the same initial configurations as in Experiment 1.

The experiment is a within-subjects design with one factor (TECHNIQUE): *PanZoom*, *PZ+OV*, *Fisheye*. Trials are grouped by technique. The 6 possible orders are used once for each participant from Experiment 1 and once for each new participant. Participants start with a training trial with *PanZoom* to learn or recall the task. Then, they perform one training trial and two measured trails for each TECHNIQUE. At the end of the experiment, we ask participants for their preferences. The experiment lasts about 35 minutes. We collected the same data as in Experiment 1 for 72 trials: 3 TECHNIQUE x 2 REPLICATION x 12 participants.

Results
The analysis of variance reveals *no* significant effect of TECHNIQUE on task completion time *TCT*: $F_{2,22} = 0.68$, $p = 0.5149$, $\eta_G^2 = 0.03$. Figure 12 shows that the three techniques are very close. The results from Experiment 1 show that none of these techniques comes close to the wall for this task.

Nine participants preferred the *Fisheye* technique, three preferred *PZ+OV*. Those who preferred the lens noted that they did not have to zoom and pan all the time. Although the lens was heavily preferred, some participants complained that it was hard to focus on the labels with the lens, despite the size of the lens being large enough to show a disk. This might be due to the high magnification factor we used, which made the lens more difficult to control, but this level of magnification was needed to make the labels readable.

Eight participants stated that the overview in *PZ+OV* was not very helpful while some mentioned that it helped them locate the red circles and empty slots. But they also said that they could do so by zooming out, so the overview was not needed.

In summary, this experiment confirmed that the wall-size display out performs the desktop for difficult data classification tasks. Although new techniques could be devised to improve the desktop condition, e.g., using multiple or adaptive lenses [23], we believe that they are unlikely to help the desktop beat the wall for complex data manipulation tasks.

CONCLUSION AND FUTURE WORK

This paper introduces a classification task that abstracts out a wide category of tasks that involve data manipulation and operationalizes two key factors: information density and task difficulty. This abstract task was informed by our observations of users of an ultra-high-resolution wall-size display, raising the question of the advantages of this type of display over a traditional desktop display.

We ran a controlled experiment comparing physical navigation in front of a wall-size display vs. virtual navigation on a desktop display for a data classification task. Our results show a robust interaction effect, such that the desktop is more efficient for easy tasks, but the wall is significantly more efficient (up to 35%) for difficult tasks. We tested three other desktop techniques with the difficult task in a follow-up experiment, but none could compete with the wall-size display.

This is but a first step in understanding the interaction environment provided by wall-size displays. Our next goal is to extend this research to collaborative work, where multiple users perform the classification task in various settings, e.g. a shared wall-size display vs. several desktop workstations working on a shared scene. A deeper understanding of spatial memory and of the respective advantages of physical and virtual navigation should also inform the design of new techniques for the wall and desktop environments that improve user performance and reduce fatigue and cognitive load.

ACKNOWLEDGEMENTS

REFERENCES

1. Andrews C., Endert A. & North C. Space to think: large high-resolution displays for sensemaking. *CHI '10*, ACM (2010).

2. Bakeman R. Recommended effect size statistics for repeated measures designs. *Behav. Res. Meth. 37*, 3 (2005).

3. Ball R. & North C. The effects of peripheral vision and physical navigation on large scale visualization. *GI '08*, CIPS (2008).

4. Ball R., North C. & Bowman D. A. Move to improve: promoting physical navigation to increase user performance with large displays. *CHI '07*, ACM (2007).

5. Beaudouin-Lafon M., Chapuis O., Eagan J., Gjerlufsen T., Huot S., Klokmose C., Mackay W., Nancel M., Pietriga E., Pillias C., Primet R. & Wagner J. Multi-surface interaction in the WILD room. *IEEE Computer 45*, 4 (2012).

6. Bezerianos A. & Isenberg P. Perception of visual variables on tiled wall-sized displays for information visualization applications. *IEEE TVCG 18*, 12 (2012).

7. Bi X. & Balakrishnan R. Comparing usage of a large high-resolution display to single or dual desktop displays for daily work. *CHI '09*, ACM (2009).

8. British Standards Institution. Test charts for clinical determination of distance visual acuity, 2003. London.

9. Cockburn A., Karlson A. & Bederson B. B. A review of overview+detail, zooming, and focus+context interfaces. *ACM CSUR 41*, 1 (2009).

10. Czerwinski M., Smith G., Regan T., Meyers B., Robertson G. & Starkweather G. Toward characterizing the productivity benefits of very large displays. *INTERACT '03*, IOS (2003).

11. Czerwinski M., Tan D. S. & Robertson G. G. Women take a wider view. *CHI '02*, ACM (2002).

12. Endert A., Andrews C., Lee Y. H. & North C. Visual encodings that support physical navigation on large displays. *GI '11*, CHCCS (2011).

13. Guiard Y., Beaudouin-Lafon M., Bastin J., Pasveer D. & Zhai S. View size and pointing difficulty in multi-scale navigation. *AVI '04*, ACM (2004).

14. Hornbæk K., Bederson B. B. & Plaisant C. Navigation patterns and usability of zoomable user interfaces with and without an overview. *ACM ToCHI 9*, 4 (2002).

15. Jakobsen R. M. & Hornbæk K. Sizing up visualizations: effects of display size in focus+context, overview+detail, and zooming interfaces. *CHI '11*, ACM (2011).

16. Kim J., Zhang H., André P., Chilton L. B., Mackay W., Beaudouin-Lafon M., Miller R. C. & Dow S. P. Cobi: A community-informed conference scheduling tool. *UIST '13*, ACM (2013).

17. Nancel M., Chapuis O., Pietriga E., Yang X.-D., Irani P. & Beaudouin-Lafon M. High-precision pointing on large wall displays using small handheld devices. *CHI '13*, ACM (2013).

18. Nancel M., Wagner J., Pietriga E., Chapuis O. & Mackay W. Mid-air pan-and-zoom on wall-sized displays. *CHI '11*, ACM (2011).

19. Nekrasovski D., Bodnar A., McGrenere J., Guimbretière F. & Munzner T. An evaluation of pan & zoom and rubber sheet navigation with and without an overview. *CHI '06*, ACM (2006).

20. Ni T., Bowman D. & Chen J. Increased display size and resolution improve task performance in information-rich virtual environments. *GI '06*, CIPS (2006).

21. Pietriga E., Appert C. & Beaudouin-Lafon M. Pointing and beyond: an operationalization and preliminary evaluation of multi-scale searching. *CHI '07*, ACM (2007).

22. Pietriga E., Huot S., Nancel M. & Primet R. Rapid development of user interfaces on cluster-driven wall displays with jbricks. *EICS '11*, ACM (2011).

23. Pindat C., Pietriga E., Chapuis O. & Puech C. Jellylens: Content-aware adaptive lenses. *UIST '12*, ACM (2012).

24. Rooney C., Endert A., Fekete J.-D., Hornbæk K. & North C. Powerwall: int. workshop on interactive, ultra-high-resolution displays. *CHI EA '13* (2013).

25. Sarkar M. & Brown M. H. Graphical fisheye views of graphs. *CHI '92*, ACM (1992).

26. Tan D. S., Gergle D., Scupelli P. & Pausch R. Physically large displays improve performance on spatial tasks. *ACM ToCHI 13*, 1 (2006).

27. Yost B., Haciahmetoglu Y. & North C. Beyond visual acuity: the perceptual scalability of information visualizations for large displays. *CHI '07*, ACM (2007).