The Perceptual Scalability of Visualization

Beth Yost, *Student Member, IEEE* and Chris North

Abstract— Larger, higher resolution displays can be used to increase the scalability of information visualizations. But just how much can scalability increase using larger displays before hitting human perceptual or cognitive limits? Are the same visualization techniques that are good on a single monitor also the techniques that are best when they are scaled up using large, high-resolution displays? To answer these questions we performed a controlled experiment on user performance time, accuracy, and subjective workload when scaling up data quantity with different space-time-attribute visualizations using a large, tiled display. Twelve college students used small multiples, embedded bar matrices, and embedded time-series graphs either on a 2 megapixel (Mp) display or with data scaled up using a 32 Mp tiled display. Participants performed various overview and detail tasks on geospatially-referenced multidimensional time-series data. Results showed that current designs are perceptually scalable because they result in a decrease in task completion time when normalized per number of data attributes along with no decrease in accuracy. It appears that, for the visualizations selected for this study, the relative comparison between designs is generally consistent between display sizes. However, results also suggest that encoding is more important on a smaller display while spatial grouping is more important on a larger display. Some suggestions for designers are provided based on our experience designing visualizations for large displays.

Index Terms—Information visualization, large displays, empirical evaluation.

1 Introduction

Geospatial intelligence analysts, epidemiologists, sociologists, and biologists all share a common problem. They are all faced with trying to understand potentially massive datasets that involve integrating spatial, multidimensional, and time-series data. The intelligence analyst has to integrate different events occurring at various geographical locations over time to prevent a terrorist attack. The epidemiologist has to integrate medical data, weather patterns, and absenteeism rates over time from various locations to predict or explain the outbreak of a disease. The biologist must consider the relationship between a biological structure and various experimental results measured over time to understand interactions between genes. Information visualizations can provide insight into these datasets.

Because of the size of many of these datasets, scalability is an important issue. Different visualizations are better able to graphically scale (require fewer pixels), which is especially important on typical desktop displays [1]. However, even designs that may not graphically scale well for a desktop display can be scaled up to a greater extent using displays that are larger and/or have a higher resolution (DPI), an example of such a display being used for visualization is shown in Fig. 1. As technology continues to decrease in cost, this is becoming a more viable option. Many places such as NASA and AT&T already have large display walls [2][3]. Theoretically, any dataset could be visualized, regardless of the visualization, on an infinite size display.

Therefore, as larger displays are used for visualization, the scalability limit may be shifted away from the graphical scalability limits imposed by the number of pixels and toward human limits. The most obvious examples of this occur when the display exceeds a resolution such that the human eye cannot perceive the pixels regardless of distance from the display, and when the display size gets to be so large that significant physical movement would be required by a user (as an example, consider the Vietnam Veterans Memorial Wall in Washington, D.C. which is 493.5 feet (150.42 meters) wide with more than 50,000 names inscribed that are each 0.53 inches (1.35 cm) high [4]). As the width of a display increases so does the use of peripheral vision, which is less sensitive to color and more sensitive to motion [5]. In these cases, the limit is created by human abilities rather than caused by the display technology.

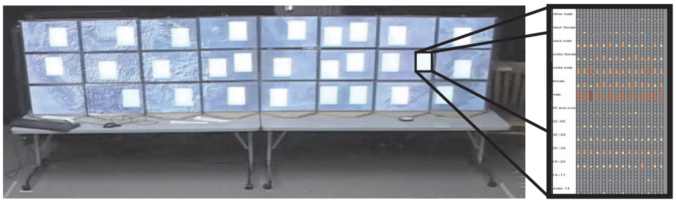

Fig. 1. Bar matrices embedded on a map shown on a 32 Mp display.

This leads to the question of perceptual scalability of visualizations for large displays. When the screen isn't the limiting factor, just how much data can a person effectively perceive? As more data is shown with increasingly larger displays, do we hit a breaking point, the limits of visualization? And how will visualizations for large displays need to fundamentally differ from visualizations on desktop displays? How are basic visualization design principles different on large displays? In this paper, we report on an experiment that begins to answer these questions by comparing three different visualizations across two different display sizes - a 2 Mp display and a 32 Mp display.

2 Related Work

2.1 Large, High-Resolution Displays

Most research on large, high-resolution displays has been about the technology used to create them. Various papers have reported on techniques used to build the displays [6][7] and software such as Chromium and DMX that can be used to distribute graphics and create a single large desktop across multiple monitors [8][9][10]. A survey of these technologies can be found in [11]. While the technology behind the display is important for assuring a usable display in terms of how much delay is introduced during interaction, more relevant to our work are large display user studies.

Large displays naturally lead to collaboration research because of size, cost, and privacy concerns. Various papers have dealt with the use of large displays for collaboration [12][13][14]. In this work we focus on a single user. Single user benefits have been shown for

- *Beth Yost is with Virginia Tech, E-Mail: beyost@vt.edu.*
- *Chris North is with Virginia Tech., E-Mail:north@vt.edu.*

Manuscript received 31 March 2006; accepted 1 August 2006; posted online 6 November 2006.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

physically larger low-resolution displays. Even if the visual angle is maintained, simply having a physically larger display improved performance on spatial tasks [15] and there is evidence that the spatial performance gender gap is narrowed by the wider field of view provided on large displays [16].

Current research on large, high-resolution displays for single users has also shown advantages. Both increased size and increased resolution appear to improve user performance when interacting with virtual environments [17]. People using a large, high-resolution display and physically interacting with it performed a simple search task faster than those using panning and zooming on a smaller display [18]. People were also faster at map reading tasks because of a reduction in the amount of virtual navigation [19]. Curving a large, high-resolution display further improved performance time by bringing the outermost pixels physically closer to the user [20]. However, research dealing specifically with visualization design on large displays is limited.

2.2 Scalability of Visualizations

Eick and Karr defined visual scalability as the ability of visualizations to effectively display large amounts of data. They structured the issue by providing the factors affecting it which include human perception, the visual metaphor, and the display as well as algorithms and computation [1]. Additionally, different techniques have been proposed for scaling up visualizations for a single monitor [21]. However, here our main focus is on the effect of using larger displays for increasing scalability.

At least one author has questioned the usefulness of large, high-resolution displays for visualization based on the limits of visual acuity [22]. Ware has also argued that a 4000x4000 display should be adequate for any visual task (not including collaboration) because it most efficiently matches 'brain pixels' to screen pixels [5]. However, little research has been conducted on user performance when using large displays to scale up visualizations.

3 Method

The goal of this study is to examine:

1. *How perceptually scalable are data visualizations for large displays?* In other words, what happens to time/accuracy as both amount of data and number of pixels are increased?

2. *Are some designs more perceptually scalable than others?* In other words, are relative comparisons between designs the same at different screen sizes?

A visualization that is perceptually scalable should not result in an increase in task completion times when time is normalized to the amount of data. It also should not result in decreasing accuracy.

A 2x3x7 mixed design was used. The independent variables were display size (with a proportional increase in data size), visualization design, and task respectively. Display size was treated as a between subjects variable while visualization and task were within subjects. Task completion time, accuracy, and subjective workload were recorded. Each independent variable is described in further detail throughout this section.

3.1 Data

The data used in this study was based on the U.S. Department of Justice, Bureau of Justice Statistics online Crime & Justice database (http://bjsdata.ojp.usdoj.gov/dataonline/). It consisted of percentages of age, race, and gender demographics of homicide victims from 1976-1989. This made it a space-time-attribute dataset in that geospatially-referenced (different states) attributes (demographic groups) were reported over time (1976-1989). To prevent influence of previous knowledge and expectations, participants were not told where the data was from or that it was homicide related until completion of the study. During the study they were only aware of the years, the demographic groups, and that each was associated with a value from 0-100. The data was modified so that a singleton was the answer to every detail/find task and there was always a single definite answer. For temporal overview tasks, the trends generally involved a 5-7 point increase or decrease each year with random variance added. Because the display had a 10 to 3 aspect ratio, the top and bottom portions of the map were cropped to fit, leaving 28 visible states. State lines and state names were displayed.

3.2 Visualizations

Three different visualizations were used. Of these three, two were space-centric designs using embedded visualizations and one was an attribute-centric design. By space-centric we mean a design where the multidimensional data is overlaid onto a single large spatial structure, in this case a United States map. By attribute-centric we mean a design where each attribute of the multidimensional data is on a separate spatial structure. The attribute-centric design strategy is analogous to both small multiples [23] and visualization spreadsheets [24]. The differences are summarized in Table 1. There is a basic trade-off between scanning views when information is separated, and clutter when information is integrated [25].

Table 1. Basic Trade-offs Between Designs

	Space-Centric	**Attribute-Centric**
Design Approach	Single view, multiple embedded attributes	Multiple views each with a single attribute
Type of Visual Complexity	Glyphs/visual encoding	Number of views
Size of View	Full display size	Display size / # of views

3.2.1 Attribute-Centric Design: Small Multiples

[MULTS] A separate United States map was shown for each attribute (where an attribute is a time and demographic group). On each map a single colored bar was displayed at each state location representing that state's population value for that demographic group and year (see Fig. 4). A global legend matching the bar color and height to the value was shown in the bottom left corner.

3.2.2 Space-Centric Designs: Embedded Visualizations

For the space-centric designs a single large map was displayed and each state had an embedded visualization of all of the attributes related to that state. Two different methods of encoding the data were used: bar matrices and time series graphs.

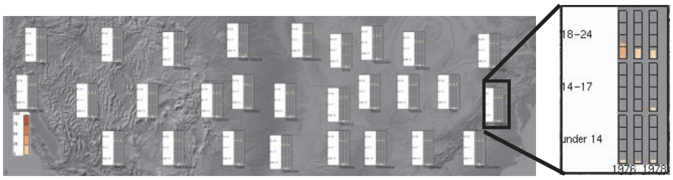

Fig. 2. BARS embedded visualization.

Fig. 3. GRAPHS embedded visualization.

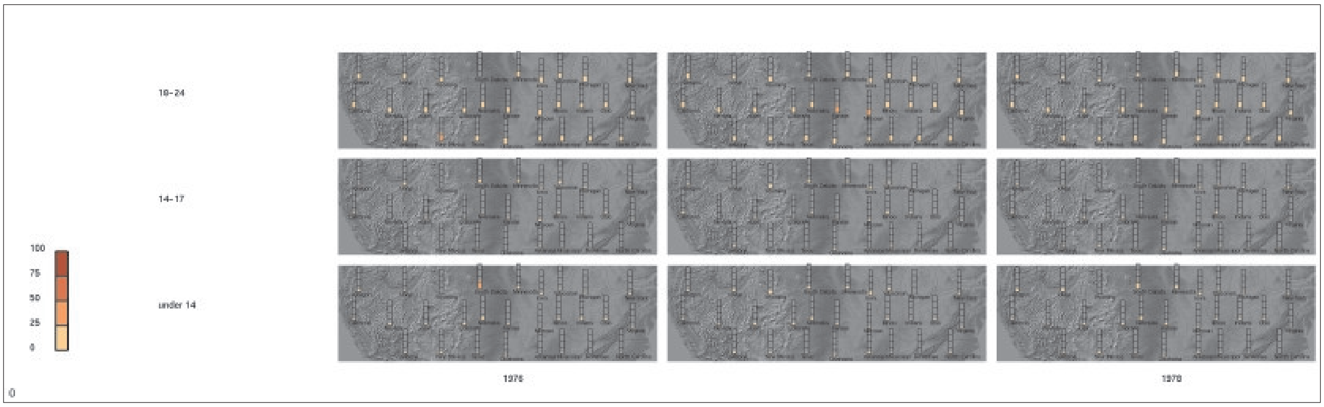

Fig. 4. MULTS attribute-centric design with nine attributes on the 2 Mp display.

Embedded visualizations were used for the space-centric designs because in general, there is a reasonable limit of being able to represent about 8 distinct dimensions with complex glyphs [5]. Additionally, bar matrices were used instead of a technique such as parallel coordinate plots so we could compare structure and attribute-centric techniques using the same visual encoding.

[BARS] Bar encoding: This visualization represented values as a matrix of colored bars at each geographic location. It was created so the encoding was consistent with the MULTS visualization, but the bars were grouped by their geographic location rather than by their space/time pair (see Fig. 2). A global legend matching bar colors to values was shown in the bottom left.

[GRAPHS] Line encoding: This was an attempt to improve on BARS by combining the y-axes and also increasing familiarity since many people have seen time series graphs before (see Fig. 3). Because of the difficulty matching colors, a local legend was included with each embedded time series graph. The legend was created so that it matched the order of the last time point for that location. However, while the legends were in a different order based on the data at each location, the colors used for each demographic group remained the same across locations.

3.3 Display Size

We used a 24 monitor tiled display (Fig. 1). It was arranged in 8 columns that were 3 monitors high. Each monitor was a 17-inch diagonal LCD, 1280x1024 (~96 DPI) for a total resolution of 10,240 x 3,072 or approximately 31.5 million pixels. The total size of the display was roughly 9 feet wide and 3.5 feet high.

Two different display conditions were used:

1) An approximately 2 Mp 2-monitor portion of the display with 3 time points by 3 demographic groups (9 attributes) and 252 total data points

2) An approximately 32 Mp 24-monitor display with 14 time points by 14 demographic groups (196 attributes) and 5488 total data points.

The data to screen size ratio is not a perfect match because some of the display was left blank to maintain an equal number of time points and demographic groups in the MULTS condition. However, the size of each individual map in the MULTS condition was constant between screen sizes as well as the size of each of the bars and the text in the BARS condition. An example of the scaled up version of the attribute-centric visualization (MULTS) is shown in Fig. 5, and examples of the embedded visualizations (BARS and GRAPHS) are shown in Fig. 6.

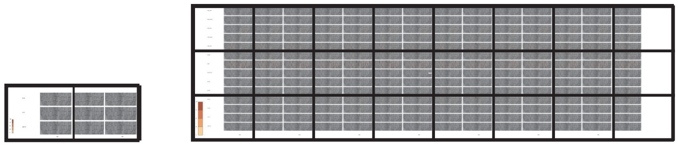

Fig. 5. Small multiples (MULTS) in two display conditions.

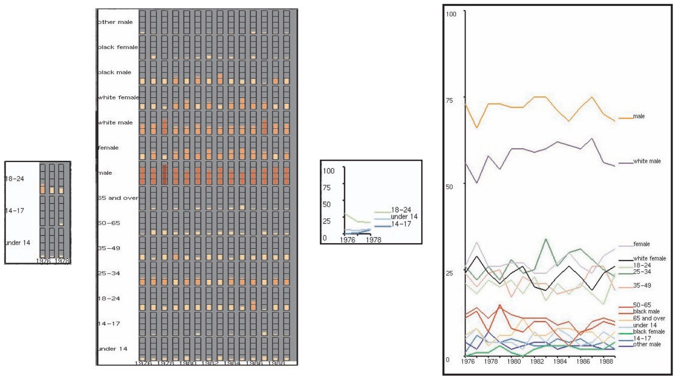

Fig. 6. Embedded visualizations (BARS, GRAPHS) in two display conditions.

3.4 Tasks

There were 7 types of tasks, 3 detail tasks and 4 overview tasks. For each of these there was a task related to time, attributes, and space. Additionally, there was a spatiotemporal overview task. The tasks are shown in Table 2. A modified version of the task was used for each of the different visualizations and each type of task was performed twice with a given visualization. This meant participants completed 3 visualization x 7 tasks x 2 trials = 42 tasks.

Table 2. Task Types and Examples

(D = detail, O = overview, T = time, A = attribute, S = space)

	Task Structure	Example Task
DT	Find a year, given an attribute and location.	Which year was the population of 14-17 the highest in Kansas?
DA	Find an attribute, given a year and location.	Which population was highest in South Dakota in 1976?
DS	Find a location, given a year and an attribute.	Which state had the highest population of under 14 in 1977?
OT	Identify a trend across time for all attributes and locations.	In general, populations have gone [up then down, down then up, up, down]?
OA	Identify a trend in attribute values for all years and locations.	In general, most populations values are in the range [<25, 25-50, 50-75, >75]?
OS	Identify a trend in location for all years and attributes.	In general, populations are the highest in the [North, South, East, West]?
OST	Identify a relationship between locations and times for all attributes.	In general, populations increase the fastest in the [North, South, East, West]?

3.5 Procedure

Participants were randomly assigned to either the 2 or 32 Mp display condition. After signing a consent form, participants filled out a demographic form. Participants in the 32 Mp condition were given a stool that could easily be moved, and in the 2 Mp condition they were given a regular office chair. For the 2 Mp condition the second and third displays from the right in the middle row were used. There was also a video camera placed overhead recording physical interaction with the display.

Participants answered all questions with one of the visualizations before moving on to the next design. There were three sets of tasks and datasets that were isomorphic. The tasks were presented in the same order to each participant, but each used a different ordering of visualizations. Detail tasks were asked before overview tasks and a modified version of the NASA TLX was used after each type of task. This asked users to rate the mental demand, physical demand, overall effort, perceived performance, and frustration for the previous tasks. Upon completion of the experiment the users were asked to subjectively compare visualizations.

3.6 Participants

There were 9 male and 3 female participants in this study. Most (10) were undergraduate and graduate computer science majors. Participants were recruited from a graduate level information visualization class; hence they were relatively experienced visualization users. No reimbursement was given for their voluntary participation. Participants were randomly assigned to either the 2 or 32 Mp condition, with one female in the 32 Mp condition and two females in the 2 Mp condition.

On a pre-experiment questionnaire participants were asked to rate their familiarity with computers, large displays, information visualization, and geographic information systems on a scale from 1-5 with 1 being strongly disagree and 5 being strongly agree. The mean familiarity ratings for the two groups of subjects are shown in Table 3. None of the reported differences in familiarity between the 2 and 32 Mp groups were statistically significant.

Table 3. Mean Familiarity Ratings per Display Condition

(1 = strongly disagree, 5 = strongly agree)

	Computer	Large Display	Info Vis	Geographic Information Systems
2 Mp	4.83	4.25	4.42	3.75
32 Mp	4.7	4.2	4.5	4.0

4 Results

4.1 Performance Time

The task completion times were first normalized by the number of data attributes so that a meaningful comparison between display sizes could be made. This meant that the times in the 2 Mp condition were divided by 9 and the times in the 32 Mp condition were divided by 196 (the respective numbers of attributes). Without normalizing the data it would be expected and unsurprising to see that all 32 Mp times were significantly longer than all 2 Mp times. Normalizing the data allowed for a fair comparison. While the rest of the statistics are done using the normalized times, the actual mean task completion times are shown in Table 4.

Table 4. Task Completion Time Means (seconds)

	2 Mp			32 Mp		
	Mults	**Bars**	**Graphs**	**Mults**	**Bars**	**Graphs**
DT	6.85	5.31	4.64	33.10	11.31	22.64
DA	5.59	3.96	4.13	18.84	15.84	11.50
DS	3.34	18.75	21.35	16.04	63.27	53.08
OT	14.17	8.40	4.86	27.96	11.39	6.07
OA	6.78	5.89	6.22	21.15	17.59	18.76
OS	3.99	2.05	7.44	18.94	7.82	17.36
OST	12.79	6.39	5.43	29.04	12.96	11.54

A 3-way mixed model ANOVA with visualization and task being within subjects factors and display size being a between subjects factor was used to analyze normalized task completion times per attribute. There was a significant 3-way interaction between visualization, display size, and task. Also relevant was a significant main effect by display size $F(1,10)=54.67$, $p<0.001$, such that the normalized time per attribute on the 32 Mp display (0.11s) was faster than on the 2 Mp display (0.84s). Note that although the normalized time per attribute was faster, the actual task completion time was longer (see Fig. 7). While there was more than a 20x increase in data size (from 9 to 196 attributes), there was less than a 3x increase in task completion times (from 7.54s to 21.25s).

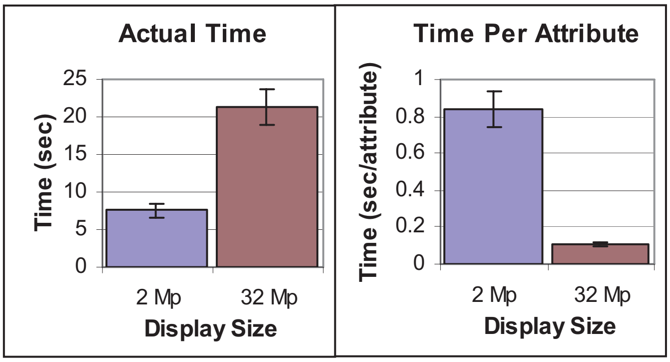

Fig. 7. Task completion times.

Because of the significant 3-way interaction, we first looked for display size by visualization interactions for each task. We next considered, for each display size, the visualization by task interactions. Post-hoc analysis was done using Tukey's HSD. Visual representations of all results are shown in Fig. 8 and Fig. 9.

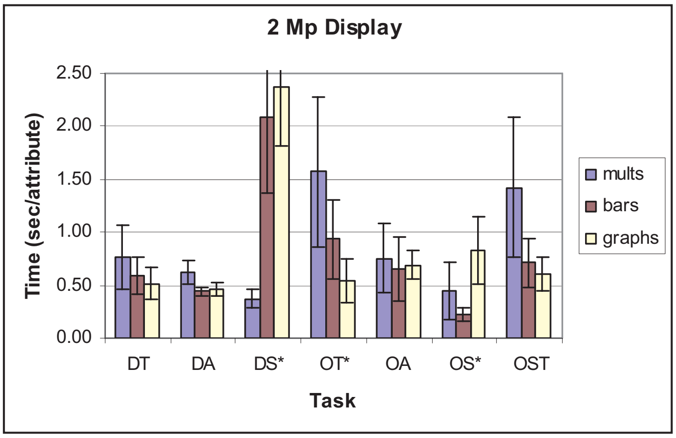

Fig. 8. 2 Mp differences in time per attribute. Bars are 95% confidence intervals. Tasks with significant differences are marked with a '*', and have non-overlapping bars.

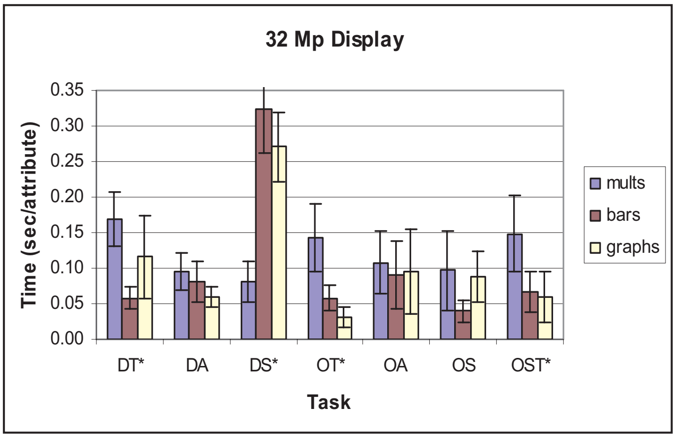

Fig. 9. 32 Mp differences in time per attribute. Bars are 95% confidence intervals. Tasks with significant differences are marked with a '*', and have non-overlapping bars.

There were significant display size by visualization interactions for the DS task F(2,20)=11.49, p<0.001, the OT task F(2,20)=4.11, p=0.032, and the OS task F(2,20)=5.19, p=0.015. These interactions are shown in Fig. 10, Fig. 11, and Fig. 12. The space-centric/embedded visualizations saw the most improvement on the DS task with the large display. Graphs saw an 8.76x improvement and bars had a 6.45x improvement while mults only had a 4.54x improvement in time per attribute when moving from the 2 Mp to 32 Mp display. This is likely because finding a location when given a year and attribute was already perfectly suited for mults so there was less room for improvement. The decrease in time for the embedded visualizations may be a result of users mentally filtering a greater portion of the information by remembering where the year x-coordinate was with respect to an embedded visualization. Some users reported remembering the relative location of an attribute and using that to filter as they scanned the embedded visualization for each state. However, while graphs and bars did see a greater percentage of improvement, mults was still significantly faster than both bars and graphs for the DS task, regardless of display size (see Table 5 for p values).

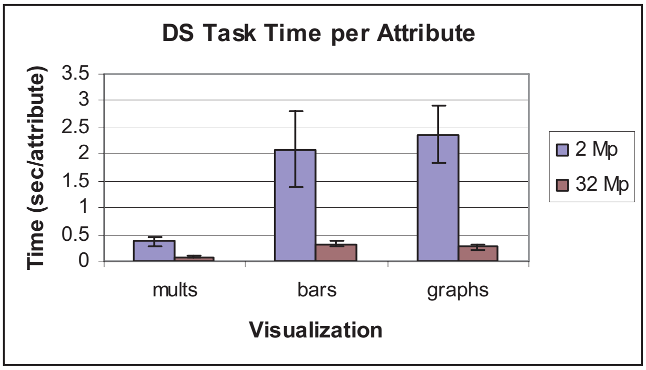

Fig. 10. Interaction between display and visualization for DS task.

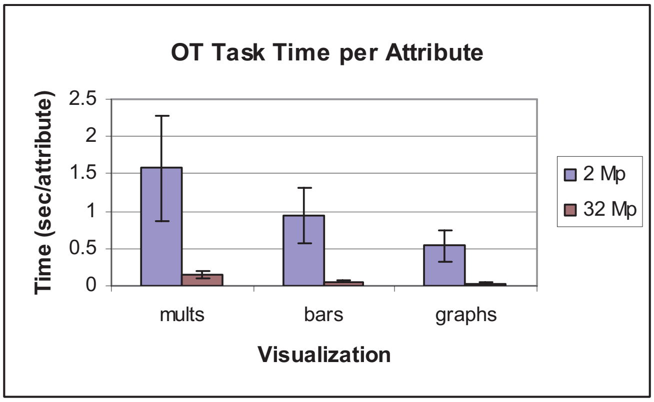

Fig. 11. Interaction between display and visualization for OT task.

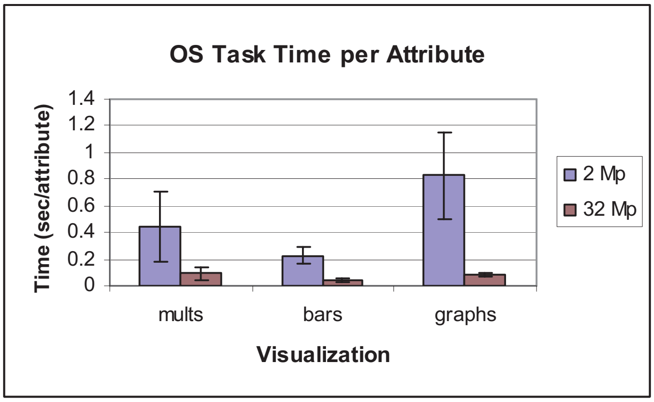

Fig. 12. Interaction between display and visualization for OS task.

For identifying trends across time (OT), mults saw a great improvement on the large display (11x), but improved less than bars (16x) and graphs (17x). However, while mults was only significantly slower than graphs on the small display, it was significantly slower than both graphs and bars on the large display. Again, the relative comparisons still held.

For the OS task, which required identifying a spatial trend, graphs improved the most (9.34x compared to 5.72x for bars and 4.59x for mults). In the 2 Mp condition, bars were significantly faster than graphs for the OS task. This was the only significant difference between the bar encoding and the line encoding. However, this difference was no longer significant in the 32 Mp condition. This is also the only significant difference that appeared on the small display but not on the large display. This suggests that the encoding may play a greater role on the smaller display while the spatial grouping is of greater importance on the larger display.

An additional observation from Table 5 is that for all three tasks involving time (DT, OT, and OST), the advantages of bars over mults did not show up on the small display but did appear on the large display. This also suggests that the spatial grouping is of greater importance on the larger display and encoding on the smaller display since grouping was the only difference between these designs.

Table 5. Significant Normalized Time Differences

	2 Mp	32 Mp
DT	-----	MULTS(0.17s) >BARS(0.06s,p=0.0206)
DA	-----	-----
DS	MULTS (0.37s) <BARS (2.08s, p=0.0042) <GRAPHS (2.37s, p=0.0014)	MULTS(0.08s) <BARS (0.32s, p<0.0001) <GRAPHS (0.27s, p=0.0001)
OT	MULTS (1.57s) >GRAPHS (0.54s), p=0.0237)	MULTS (0.14s) >BARS (0.06s, p=0.0035) >GRAPHS (0.03s, p=0.0005)
OA	-----	-----
OS	BARS (0.23s) <GRAPHS (0.83s, p=0.0153)	-----
OST	-----	MULTS(0.15s) >BARS (0.07s,p=0.0063) >GRAPHS (0.06s,p=0.0036)

4.2 Accuracy

In general, users were able to correctly answer almost all of the questions; the actual numbers are shown in Table 6. For accuracy there was a visualization by task interaction (p=0.001) along with a main effect by visualization and main effect by task, but no other interactions. The difference between display sizes was not significant (p=0.312). The only significant differences occurred in task DS such that graphs (67%) were significantly less accurate than both bars (92%, p=0.0192) and mults (100%, p=0.0019). This was likely the result of the integration of y-axes and perhaps difficulty in distinguishing colors.

Table 6. Total Correct Answers (12 max)

	2 Mp			32 Mp		
	Mults	**Bars**	**Graphs**	**Mults**	**Bars**	**Graphs**
DT	12	12	10	9	12	8
DA	12	12	11	11	12	12
DS	12	10	7	12	12	9
OT	11	11	12	10	12	11
OA	12	12	12	11	11	11
OS	12	12	12	12	12	10
OST	12	12	12	10	12	12

4.3 Task Workload

Users reported mental demand, physical demand, effort, perceived performance, and frustration on rating scales from 1 to 10 after all detail tasks with each visualization and again after all overview tasks with each visualization. A MANOVA showed all interactions and main effects as significant. Therefore, univariate analysis was done followed by post-hoc analysis using Tukey's HSD for individual comparisons. The mean scores for each condition are shown in Table 7.

Table 7. Task Workload Means

DETAIL TASKS						
	2 Mp			32 Mp		
	Mults	Bars	Graphs	Mults	Bars	Graphs
Mental	3.08	3.83	7.00	6.08	5.42	7.50
Physical	3.00	3.00	4.08	8.00	6.67	5.92
Effort	3.42	3.42	6.08	7.08	6.33	7.42
Performance	8.17	7.50	6.67	8.25	8.25	7.75
Frustration	0.83	1.00	3.75	5.67	3.92	5.08
OVERVIEW TASKS						
	2 Mp			32 Mp		
	Mults	Bars	Graphs	Mults	Bars	Graphs
Mental	4.75	3.17	3.42	7.25	4.67	4.42
Physical	3.08	2.25	3.33	5.75	2.92	2.67
Effort	4.67	2.75	3.17	7.08	3.92	3.58
Performance	7.50	8.00	7.75	6.67	8.92	8.25
Frustration	1.50	0.75	1.08	5.58	2.25	2.83

Overall, the large display users were significantly more frustrated (4.22 vs. 1.49, p=0.019) and reported more physical demand (5.32 vs. 3.13, p=0.055) than the small display users (see Fig. 13). The only visualization specific differences that were seen were between mults on the different display sizes. Users in the large display condition reported more mental demand (6.08 vs. 3.08, p=0.043) and physical demand (6.88 vs. 3.04, p=0.007) for mults than users in the small display condition. Within the small display conditions, users reported more physical demand with graphs (3.71) than with bars (2.63, p=0.0024). Within the large display condition, users reported more physical demand from mults (6.88) than graphs (4.29, p=0.0272).

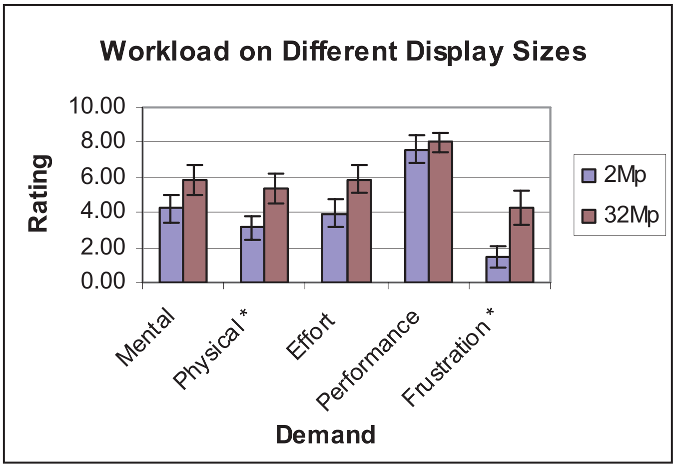

Fig. 13. Workload on display sizes. Significant differences have a '*'.

There were also some differences between visualizations based on the task. For detail tasks, users reported more mental demand with graphs (7.25) than both bars (4.63, p<0.001) and mults (4.85, <0.001) as well as more effort with graphs (6.75) compared to both bars (4.88, p=0.0116) and mults (5.25, p=0.0468). For overview tasks, mults required more effort (5.88 vs 3.33 for bars and 3.38 for graphs, p<0.001), had a lower perceived performance compared to bars (7.08 vs. 8.46 for bars, p=0.0256) and also had higher frustration levels (3.45 vs. 1.5 for bars, p=0.03). These findings match the proximity compatibility principle [25] in that the embedded time series graphs required more mental demand and effort for detail tasks – where mental separation would need to occur, and for overview tasks mults required more effort, resulted in more frustration, and lower perceived performance – tasks where integration was necessary.

4.4 User Preference

User preference was different based on the display size (see Fig. 14 and Fig. 15). In the 2 Mp condition, four of five users preferred mults followed by bars and then graphs (the 6th user gave a circular answer). However, in the 32 Mp condition four of five users preferred bars first followed by graphs and then mults. Users always preferred bars to graphs, but mults was the most preferred on the smaller display and the least preferred on the large display. Therefore user preference for visualizations was different based on the display size/amount of data.

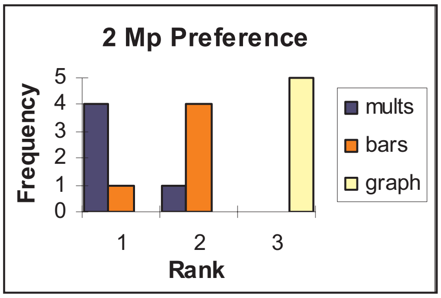

Fig. 14. Frequency of user visualization preference for 2 Mp display (rank of 1=best, 3=worst).

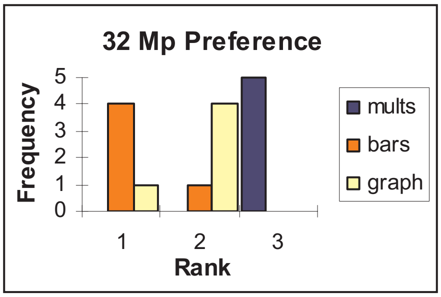

Fig. 15. Frequency of user visualization preference for 32 Mp display.

5 Discussion

As large displays continue to decrease in cost, we must continue to explore how various visualizations scale for these large displays. In this study, although overall task time increased, the time per number of attributes actually decreased. There was more than a 20x increase in data size for the large display, but less than a 3x increase in task completion times. This means that overall, across tasks and visualizations, these designs were perceptually scalable in terms of time. We also did not find a significant decrease in accuracy on the large display, with a change only from 95% to 92%, again suggesting these designs are perceptually scalable. Despite having 32 Mp, we did not hit the limits of visualization. Users were able to successfully physically navigate to complete detail tasks and were also able to gain a large overview and perceptually integrate information across a display almost 9 feet in width.

It also appears that in general, based on time and accuracy, the relative comparison of these three designs was consistent across display sizes. However, there were some differences. The fact that the only time difference between bars and graphs disappeared on the large display, and that the difference between mults and bars showed up on almost every task on the large display suggests that the line vs. bar encoding was most important on the small display. This matches previous research suggesting that visual encoding can be more important than grouping on small displays [26]. However, as the visualization was scaled up using the display, the spatial grouping became much more important. This spatial grouping likely increased visual aggregation and reduced the amount of physical navigation. This was echoed both by the shift in user preference from mults on the 2 Mp display to bars on the 32 Mp display and also by the significant increase with the mults design in physical and mental effort. Results also support the proximity compatibility principle [25] and go against the idea that small multiples might show the most benefit when more than 16 views are displayed [27]. This shows great promise for using embedded visualizations for geospatially-referenced data on large displays.

While many open issues regarding information visualization on large, high-resolution displays remain, there were some basic observations from this study and from the pilot tests that may prove useful for designers. In general, **consider how various encodings will be affected by viewing distance and angle**. As an example, the size of bars will be affected by the distance from a display. In our pilot study, users preferred the colored bars to plain white bars because it was very hard to compare the size of bars from one side of the display to the other, but color could still easily be compared. As another example, the orientation of a glyph may be affected by the viewing angle, therefore these are not good choices. However, if the displays are of good quality then color may not be affected, making it a good choice. Our use of dual encoding of color and size seemed an effective compromise.

A second suggestion is that if you want to use a large display to scale a visualization, **choose a visualization with scalable graphical encodings**. For example, 3D occlusion problems and the number of different perceivable colors cannot be improved with a larger display, so these graphical encodings are not scalable. With respect to our study, using 14 demographic groups was already pushing the limits of embedded time series graphs because of the slight difference in color used to encode each line. On the other hand, the bar design does use scalable graphical encodings because it can be scaled infinitely as long as there are enough pixels.

A third observation is that even if legends and labels are larger, as a user physically navigates the display they often lose sight of that that information. Therefore, **consider having both local and global legends on a large display.** The local legends on the graphs meant that users had to walk close to the display to read the legend before moving further back to get an overview. If they had a global legend then that would not have been necessary. Bars and mults only had global legends, which meant sometimes users had to step back to see the global legend in the middle of doing a detail task. Additionally, the demographic group labels and the year labels were placed along the left and the bottom for the mults design. Some users lost track of which column or row they were viewing. Therefore, it would have been useful to **place labels at multiple strategic locations**. Carefully placing labels may help users maintain physical context.

6 Conclusion

In this paper we reported the results of a study comparing three different visualizations on both a small and large display. Results showed that the designs used were perceptually scalable – not resulting in an increase in normalized performance time or a significant decrease in accuracy. Accuracy only decreased from 95% to 92%, and a 20x increase in data resulted in only a 3x increase in task completion times. Using a combination of perceptual abilities and physical navigation people were able to effectively use a 32 Mp display (2x Ware's proposed 16 Mp display [5]).

Results also showed that relative comparison between designs with respect to time and accuracy was typically the same regardless of the display size. However, based on user preference and workload, graphical encoding seems to be more important with less data on a small display whereas spatial grouping seems to be more important with more data on a large display. On the large display, both embedded visualization designs were generally significantly faster than small multiples. User preferences also switched on large displays, with most users preferring both of the embedded designs. In addition to the results of the study, we presented some of our observations from designing visualizations for large displays.

7 Future Work

In the future we would like to explore how the results would be affected if we had increased the number of spatial locations as opposed to the number of data attributes. Additionally, we would like to explore interactive visualizations with larger datasets and on larger high-resolution displays. Basic issues for interaction on large displays such as what types of input devices best allow you to interact and the discovery of new techniques [28][29] have been explored. Less attention has been given to how visualization specific interactions such as brushing and linking, overview, and navigation techniques will need to be modified for these displays. Because of the use of peripheral vision, should propagated changes in linked views be delayed until you look at that view or perhaps occur more gradually? How might we use cascading overviews or adjust the size of overview and detail views with a very large display? In addition to interaction related issue there are many basic perceptual issues that remain to be answered. Is the order of effectiveness of graphical encoding [30][31] the same on a large display as on a smaller display? How does visual aggregation compare to computational aggregation and when does it become advantageous simply because it reduces visual complexity? Those are just a few of the many open issues with respect to designing visualizations for these displays.

Acknowledgements

This research is partially supported by the National Science Foundation grant #CNS-04-23611. This study was also supported and monitored by the Advanced Research and Development Activity (ARDA) and the National Geospatial-Intelligence Agency (NGA) under Contract Number HM1582-05-1-2001. The views, opinions, and findings contained in this report are those of the authors and should not be construed as an official US Department of Defense position, policy, or decision, unless so designated by other official documentation.

References

[1] S. G. Eick and A. F. Karr, "Visual Scalability," *Journal of Computational & Graphical Statistics*, vol. 11, no. 1, pp. 22-43, 2002.

[2] T. A. Sandstrom, C. Henze, and C. Levit, "The Hyperwall," *Proc. Coordinated and Multiple Views in Exploratory Visualization (CMV)*, pp. 124-133, 2003.

[3] B. Wei, C. Silva, E. Koutsofios, S. Krishnan, and S. North, "Visualization Research with Large Displays," *IEEE Computer Graphics and Applications*, vol. 20, no. 4, pp. 50-54, 2000.

[4] "Stones & Mortar," *National Park Service*, http://www.nps.gov/mall/mortar/mortar.htm#vive, 2004.

[5] C. Ware, *Information Visualization: Perception for Design*: Morgan Kaufmann, 2000.

[6] K. Li, H. Chen, Y. Chen, D. W. Clark, P. Cook, S. Damianakis, G. Essl, A. Finkelstein, T. Funkhouser, T. Housel, A. Klein, Z. Liu, E. Praun, R. Samanta, B. Shedd, J. P. Singh, G. Tzanetakis, and J. Zheng, "Building and Using a Scalable Display Wall System," *IEEE Computer Graphics and Applications*, vol. 20, no. 4, pp. 29-37, 2000.

[7] G. K. Starkweather, "DSHARP-A Wide-Screen Multi-Projector Display," *Journal of Optics: Pure and Applied Optics*, vol. 5, no. 5, pp. S136-S139, 2003.

[8] M. Hereld, I. R. Judson, and R. L. Stevens, "Tutorial: Introduction to Building Projection-based Tiled Display Systems," *IEEE Computer Graphics and Applications*, vol. 20, no. 4, pp. 22-28, 2000.

[9] G. Humphreys, M. Houston, R. Ng, R. Frank, S. Ahern, P. D. Kirchner, and J. T. Klosowski, "Chromium: A Stream-Processing Framework for Interactive Rendering on Clusters," *ACM Transactions on Graphics (TOG)*, vol. 21, no. 3, pp. 693-702, 2002.

[10] O. G. Staadt, J. Walker, C. Nuber, and B. Hamann, "A Survey and Performance Analysis of Software Platforms for Interactive Cluster-Based Multi-Screen Rendering," in *Workshop on Virtual Environments*. Zurich, Switzerland: ACM, 2003.

[11] T. Ni, G. S. Schmidt, O. G. Staadt, M. A. Livingston, R. Ball, and R. May, "A Survey of Large High-Resolution Display Technologies, Techniques, and Applications," *Proc. IEEE Virtual Reality*, pp. 223-234, 2006.

[12] S. Elrod, R. Bruce, R. Gold, D. Goldberg, F. Halasz, W. Janssen, D. Lee, K. McCall, E. Pedersen, K. Pier, J. Tank, and B. Welch, "Liveboard: A Large Interactive Display Supporting Group Meetings, Presentations, and Remote Collaboration," *Proc. Human Factors in Computing Systems (CHI)*, pp. 599-607, 1992.

[13] B. Johanson, A. Fox, and T. Winograd, "The Interactive Workspaces Project: Experiences with Ubiquitous Computing Rooms," *IEEE Pervasive Computing*, vol. 1, no. 2, pp. 67-74, 2002.

[14] N. A. Streitz, J. Geibler, T. Holmer, S.-i. Konomi, C. Muller-Tomfelde, W. Reischl, P. Rexroth, P. Seitz, and R. Steinmetz, "i-LAND: An Interactive Landscape for Creativity and Innovation," *Proc. Human Factors in Computing Systems (CHI)*, pp. 120-127, 1999.

[15] D. Tan, D. Gergle, P. Scupelli, and R. Pausch, "With Similar Visual Angles, Larger Displays Improve Spatial Performance," *Proc. Human Factors in Computing Systems (CHI)*, pp. 217-224, 2003.

[16] M. Czerwinski, D. Tan, and G. Robertson, "Women Take a Wider View," *Proc. Human Factors in Computing Systems (CHI)*, pp. 195-202, 2002.

[17] T. Ni, D. Bowman, and J. Chen, "Increased Display Size and Resolution Improve Task Performance in Information-Rich Virtual Environments," *Proc. Graphics Interface*, pp. 139-146, 2006.

[18] R. Ball and C. North, "Effects of Tiled High-Resolution Displays on Basic Visualization and Navigation Tasks," *Proc. Extended abstracts of Human Factors in Computing Systems (CHI)*, pp. 1196-1199, 2005.

[19] R. Ball, M. Varghese, A. Sabri, E. D. Cox, C. Fierer, M. Peterson, B. Carstensen, and C. North, "Evaluating the Benefits of Tiled Displays for Navigating Maps," *Proc. International Conference on Human-Computer Interaction (IASTED-HCI)*, pp. 66-71, 2005.

[20] L. Shupp, R. Ball, B. Yost, J. Booker, and C. North, "Evaluation of Viewport Size and Curvature of Large, High-Resolution Displays," *Proc. Graphics Interface (GI)*, pp. 123-130, 2006.

[21] J. D. Fekete and C. Plaisant, "Interactive Information Visualization of a Million Items," *Proc. IEEE Symposium on Information Visualization*, pp. 117-124, 2002.

[22] E. Wegman, "Huge Datasets and the Frontiers of Computational Feasibility," *Computational and Graphical Statistics*, vol. 4, no. 4, pp. 281-295, 1995.

[23] E. R. Tufte, *Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press, 1983.

[24] E. H. Chi, J. Riedl, P. Barry, and J. Konstan, "Principles for Information Visualization Spreadsheets," *IEEE Computer Graphics and Applications (Special Issue on Visualization)*, vol. 18, no. 4, pp. 30-38, 1998.

[25] C. D. Wickens and C. M. Carswell, "The Proximity Compatibility Principle: Its Psychological Foundation and Relevance to Display Design," *Human Factors*, vol. 37, no. 3, pp. 473-494, 1995.

[26] B. Yost and C. North, "Single Complex Glyphs Versus Multiple Simple Glyphs," *Proc. Extended Abstracts of Human Factors in Computing Systems (CHI)*, pp. 1889-1892, 2005.

[27] A. M. MacEachren, *How Maps Work: Representation, Visualization, and Design*. New York: The Guilford Press, 1995.

[28] A. Bezerianos and R. Balakrishnan, "View and Space Management on Large Displays," *IEEE Computer Graphics and Applications*, vol. 25, no. 4, pp. 34-43, 2005.

[29] G. Robertson, M. Czerwinski, P. Baudisch, B. Meyers, D. Robbins, G. Smith, and D. Tan, "The Large-Display User Experience," *IEEE Computer Graphcs and Applications*, vol. 25, no. 4, pp. 44-51, 2005.

[30] W. S. Cleveland and R. McGill, "Graphical Perception: Theory, Experimentation and Application to the Development of Graphical Methods," *Journal of the American Statistical Association*, vol. 79, no. 387, pp. 531-554, 1984.

[31] J. D. Mackinlay, "Automating the Design of Graphical Presentations," *ACM Transactions on Graphics (TOG)*, vol. 5, no. 2, pp. 110-141, 1986.