The Cost of Display Switching: A Comparison of Mobile, Large Display and Hybrid UI Configurations

Umar Rashid
University of St Andrews
ur9@st-andrews.ac.uk

Miguel A. Nacenta
University of St Andrews
mans@st-andrews.ac.uk

Aaron Quigley
University of St Andrews
aquigley@st-andrews.ac.uk

ABSTRACT

Attaching a large external display can help a mobile device user view more content at once. This paper reports on a study investigating how different configurations of input and output across displays affect performance, subjective workload and preferences in map, text and photo search tasks. Experimental results show that a hybrid configuration where visual output is distributed across displays is worst or equivalent to worst in all tasks. A mobile device-controlled large display configuration performs best in the map search task and equal to best in text and photo search tasks (tied with a mobile-only configuration). After conducting a detailed analysis of the performance differences across different UI configurations, we give recommendations for the design of distributed user interfaces.

Categories and Subject Descriptors

H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

General Terms

Design, Experimentation, Human Factors.

Keywords

Distributed user interfaces, multi-display environments, mobile input, text search, photo search, map search.

1. INTRODUCTION

Mobile devices are used by billions of people for voice calls or messaging and by hundreds of millions to access e-mail, browse photographs, use applications or access geographical data. However, mobile devices have intrinsic limitations due to their necessarily small size. A very apparent drawback of mobile devices is that their screens do not allow for the display of large amounts of information at once without requiring interaction, which limits the possibilities for information access and manipulation on the go.

One attractive approach taken to address this problem is to complement the small display of a mobile device with a larger display, which might be borrowed opportunistically from the environment (see Figure 1 and [32]). Such configuration has several perceived advantages, including: a) the large display might supplement the screen real estate lacking in the mobile display, b) input can still be provided through the mobile device, which is familiar to, proximate to and easily manipulated by the user, and c)implementation of this kind of configuration is already possible with a currently available infrastructure (e.g., network-connected digital signage [18]). Many research projects (e.g., [4, 6, 9, 10, 11], and numerous new commercial products (e.g., Wii U, Apple Remote App, multi-device Scrabble) point to a future where we might expect operating system-level support for multi-display ecosystems with seamless interaction across opportunistically available displays.

Figure 1. Hybrid configuration for information seeking.

However, before this can happen, we need to better understand the implications on performance and use when the interface to an application exists across multiple displays. Indeed, the distribution of the interface may introduce new problems that are not present in single-screen user interfaces [31, 25] and we need to know whether the added problems outweigh the advantages of the extra display. In this paper we describe a series of experiments that investigate the effects of three configurations (mobile-only, mobile-controlled large display, and hybrid) on performance, subjective workload and user preference in three tasks common to mobile scenarios (map, text, and photo search).

We contribute important new empirical evidence applicable to the design of interfaces for mobile multi-display scenarios. The experiments show that distributing visual information across the two displays (hybrid) causes significant overheads due to the necessary gaze shifts, resulting in the worst performance across configurations. The best configuration across all tasks was the mobile-controlled large display, where output is constrained to the large display and input is provided in the mobile device.

2. RELATED WORK

2.1 Distributed and Multi-Device UI

Multi-Display Environments (also known as Distributed Display Environments) are computing systems that present visual output to more than one display [19, 24]. Previous work has explored the challenges of designing, enabling, and implementing generic MDEs (e.g., [30, 21, 22]).

In this paper we focus on MDEs that use a large situated display to complement a mobile device (mobile+large display), which has been explored by several research groups in the past [10, 23]. Some prototypes demonstrate varied applications where the mobile device is used for both input and output [28, 15, 18, 6, 27, 9], or exclusively for input [2, 18, 23]. LensMouse [33] is also an example of this kind of configuration motivated instead as enhanced input device for personal computers. In addition to research prototypes, several commercial systems exist that take advantage of similar MDE configurations such as the upcoming Wii U, or other applications to control PCs or TVs from a smartphone (e.g., [29]). Phone projector systems [14, 17] are also similar to our scenario in that they produce two visual spaces.

http://dx.doi.org/10.1145/2254556.2254577

The copy of record of the paper can be found in:
http://dl.acm.org/citation.cfm?doid=2254556.2254577

2.2 Studies on Distributed Interaction

Although many mobile+large display systems exist with alternative distributions of input and output, comparisons of the effects of different configurations are scarce.

ARCPad [23] and LensMouse [33] contain evaluations of pointing and menu access capability; although these are useful for the design of input interaction techniques, we seek to investigate higher-level tasks that go beyond selection. Finke et al. [11] investigate several strategies for the design of applications across personal and public displays and test several examples; however, these provide qualitative evidence about specific applications that might be difficult to generalize, and they did not find differences between configurations. Closest to our research is Gostner et al.'s study [13], which investigated a text search and entry task in large display-only, mobile-only, and hybrid configurations and found that the large-display and hybrid condition were best. Our study extends their work in three ways: we test three tasks, we isolate text search (which they studied in combination with text entry), and we study a large display configuration that is comparable in input to the hybrid and mobile (their large display condition had input in the large display, which could introduce a confound – results could be due to input or output differences).

Although not identical, results from studies on projector phone devices are also relevant to our work. Studies about picture browsing [14] and map search tasks [17] found performance advantages for the projected-only configuration for map search tasks. Besides the differences in the physical environment (theirs was a hanging mini-projector attached to a mobile phone), our work differs from theirs in two main ways: we use touch as input (they used buttons and a joystick), and we provide a statistical analysis of differences in performance and user preference. Cauchard et al. [5] studied the effects of the positioning of the projected image on hybrid (mobile+projected display) configurations. They found that having two screens in the same visual field encouraged context shifts, but that did not affect performance.

2.3 Issues in Mobile-LD Interactions

Other researchers have investigated issues that are relevant, but they focused on other scenarios. For example, mobile+large-display configurations can be considered to be bifocal display systems used for *overview and detail* [8]. Grudin [16] highlights how the partition between displays can be beneficial; Tan et al. [31] showed that comparing information across displays at different depths has a small cost in terms of performance; and Bi et al. [3] found that bezels can negatively affect visual search in vertical displays if objects are split. Finally, map navigation in large displays was found to be more efficient with physical navigation than through panning and zooming [1].

3. EMPIRICAL STUDY

The main goal of the studies is to find out which configuration is best and worst for each task in terms of performance, workload, and participant preference. The secondary goal is to investigate the possible causes of these results. We chose tasks among those identified as the most common information tasks on mobile scenarios: map search, text search, and photo search [7].

Elements common to all three experiments are described in the following sub-sections. Specifics of each experiment are explained in their corresponding experiment sections.

3.1 Apparatus

The apparatus consisted of two displays: a HTC Desire HD™ mobile device with a 480x800px (240ppi), 5.7x9.6cm (4.3") screen and a 1980x1080px (54ppi), 92.5x52cm (42") LG High Definition Liquid Crystal™ display. The large display was attached to a Windows 7™ PC and both devices ran custom experimental Java™ software connected through a 13Mbps IEEE 802.11 wireless connection (no noticeable delay).

Participants sat on a chair, with the large vertical display perpendicular to them and centered in front, at a distance of about 120cm. Participants were not movement-constrained although we kept the chair in a fixed position. Participants were allowed to hold the untethered mobile device in the non-dominant hand, in a portrait orientation (content rotation was disabled).

3.2 Conditions

The main factor for all three experiments was *UI configuration*. UI Configuration had three levels:

Mobile. Only the mobile device was used for input and output. Panning and paging through touch input allowed virtual navigation of data space that did not fit on screen.

Large Display. The mobile device was used only as input device; output was shown only on the large display. Input was captured through a modified version of RemoteDroid [29] which works like a buttonless touchpad.

Hybrid. The mobile device was used for input and output as in the *mobile* configuration, but output was also shown on the large display. In the case of a large data set, the large display showed the whole data, while the handheld only showed a partial view. The part of the dataset visible on the mobile device was dynamically represented with the *scope window frame,* a rectangular bounding box on the large display (Figure 2).

Resolution and input ratio were kept constant across the three configurations. Each map, photo and text was represented by the same amount of pixels regardless of the display. Input was calibrated so that the same drag gesture would result in the same amount of pixel movement. The C-D gain was approximately 4.6 for mobile-controlled movements on the large display and 1 for content on the mobile display (direct touch).

The secondary factor in all three experiments was *data size*, that had two levels: *small*, and *large*. Small data was calculated to fit completely on the mobile display, while large data did not, requiring some kind of navigation dependent on the task.

3.3 Measures

All experiments measured:

Completion time (CT). Calculated from the time the trial starts until the last selection is made.

Errors. Proportion of trials with incorrect answer(s).

Length of interaction (LI). Finger drags distance logged by the mobile screen (in pixels) during the trial.

Gaze shifts (#GS). The number of times a participant shifts gaze, or gaze and head pose, between displays, per trial. Measured by examining the video of each trial.

Subjective Workload. For each configuration and task, participants filled a six-question NASA TLX survey.

Overall Preference. Each participant ranked each configuration in order of preference.

3.4 Participants

Twenty-six participants (age 19-33, 7 females) were recruited from the local university in exchange for a £5 gratuity. All participants had normal or corrected-to-normal vision. Except two, all had used touch phones. All participants did all three experiments.

3.5 Procedure

Participants rehearsed with each condition before the real trials. For each task, they performed three blocks of eight trials, each on a different configuration. The order of the conditions was balanced across participants, although each participant would see the same order of configurations across all three experiments. Within each block, participants performed four trials under the small data condition followed by four trials under the large data condition. The first and fifth trial in each block were considered as training trials, and are excluded from the analyses. Participants performed a total of 8x3x3 = 72 trials of which 18 (25%) were considered training. After each block, participants filled the TLX questionnaire, and after each experiment they ranked the configurations in order of preference. Each session took approximately 1h.

3.6 Data Filtering

We filtered out the data of participants that did not comply with criteria established before the analysis. We discarded a participant's data for an experiment when more than $^2/_3$ of the trials contained at least one error, when the participant showed signs of not understanding the task during the real trials, and when a disruption took place during the test (e.g., loud external noise).

4. EXPERIMENT 1: MAP SEARCH

This task represents map search situations where a user needs to find a location with a choice criterion (e.g., the cheapest hotel) and is modeled upon previous experimental tasks [1, 17]. Participants had to find and tap on the marker with the lowest price label out of 15 markers distributed on the map shown on a fixed zoom level. The trial would not finish until the unique lowest-price marker was tapped. For *small data*, all markers were visible within the mobile screen and panning was disabled.

In *large data* trials, the markers were spread out across the full map, which was the same size as the large screen (1920x1080), and was accessible from the mobile device through standard 2D panning. In the hybrid condition, a rectangle on the large screen dynamically represented the viewport currently shown on the mobile display (Figure 2).

Figure 2. Map search task on mobile and large display (hybrid configuration). Devices not to scale.

4.1 Results

No participant data was excluded for the map experiment.

4.1.1 *Performance*

The main measure of the experiment was completion time. Due to the non-normality of the time measure distributions all time statistical analyses were performed on log-transformed measures. Average times and graphs are presented in non-transformed measures (seconds). The same applies to the other experiments.

An omnibus ANOVA with *configuration* and *data size* as main factors and participant as random factor revealed a strong effect of configuration ($F_{2,50} = 11.29$, $p < 0.001$, $\eta_p^2 = 0.31$), and data size ($F_{1,25} = 84.93$, $p < 0.001$, $\eta_p^2 = 0.77$) on completion time, as well as interaction between the two main factors ($F_{2,50} = 11.77$, $p < 0.001$, $\eta_p^2 = 0.32$). To follow up the interaction, we performed separate analysis on the large and small data conditions.

An ANOVA test on the small data trials showed that configuration had a significant effect on completion time ($F_{2,50} = 6.28$, $p < 0.01$, $\eta_p^2 = 0.20$). For the small data condition, mobile was fastest (M=8.99s), followed by large display (M=9.29s, 3.3% slower) and hybrid (M=11.49s, 28% slower). Post-hoc tests corrected for multiple comparisons (Tukey's HSD) showed significant differences between the hybrid (the slowest) and the other configurations, but not between mobile and the large display.

The ANOVA on the large data trials was also significant for configuration ($F_{2,50} = 18.45$, $p < 0.001$, $\eta_p^2 = 0.42$), but the post-hoc comparisons showed a different pattern, where large display is significantly faster (M=11.17s) than both hybrid (M=14.01, 25% slower) and mobile (M=16.52, 47% slower) configurations. Figure 3 shows a summary of completion times for the map task where * sign on bars indicates a significant difference.

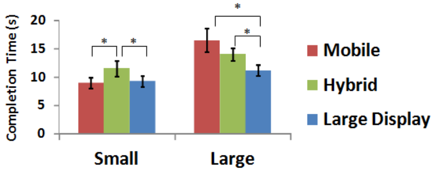

Figure 3. Map search: mean completion times (error bars indicate 95% confidence intervals).

Error data was non-parametric and was analyzed using a Friedman test, which showed no significant differences between configurations ($\chi^2(2) = 4.23$, $p > 0.05$). The average number of trials with errors was lowest in large display (M=17.3%) followed by hybrid (M=18.3%) and mobile (M=24.5%) configurations.

4.1.1.1 *Subjective Evaluation*

The 10-point scale NASA TLX questionnaire questions were analyzed separately using non-parametric Friedman paired measures tests. We found significant differences among UI configurations regarding physical demand temporal demand, performance, effort and frustration level, but not for mental demand (see statistics and averages in Table 1). The ratings show that participants ranked large display as best across all workload questions and mobile as worst, with hybrid in the middle.

Table 1. Map search: average ratings and statistics for the TLX questions. For performance, higher means better. Green, yellow, red indicate best, middle and worst configuration.

Factor	χ² (2)	p	Mobile	Hybrid	Large D.
Physical Demand	11.06	<0.01*	2.98	2.58	2.27
Mental Demand	5.01	>0.05	2.75	2.19	2.27
Temporal Demand	9.32	<0.01*	2.9	2.33	2.27
Performance	7.27	<0.05*	7.85	8.42	8.75
Effort	6.93	<0.05*	3.36	2.92	2.48
Frustration	8.44	<0.05*	3.17	2.38	2.07

Table 2. Map search: configuration preference rankings.

	Best		Worst
Mobile	2	7	17
Hybrid	12	11	3
Large Display	12	8	6

In the overall ranking, mobile was overwhelmingly the least preferred configuration (17 out of 26 participants, 65%). Large display and hybrid were ranked similarly (each preferred by 12 participants), although hybrid was more often the intermediate choice. The complete preference choices are shown in Table 2. A Friedman test of the rankings shows significant differences between configurations ($\chi^2(2) = 12.0$, $p < 0.01$).

Comments from the participants serve to further explain the preference rankings. Large display was preferred to hybrid because it did not require switching attention between displays. Participants that preferred hybrid often mentioned the availability of both detail and overview in different displays. They also highlighted the ability to easily keep track of the overall position as one of the advantages of hybrid vs. mobile. The extensive panning required in the mobile configuration was cited as a disadvantage for mobile, although not for hybrid.

4.2 Auxiliary Analyses

In addition to the measures analyzed above, we collected *gaze shift* and *length of interaction* measures to explore explanations for the performance differences. Note that these analyses are not the main focus of the study and should be interpreted and generalized with caution.

An omnibus ANOVA of *length of interaction* with configuration and data size as main factors and participant as random factor yielded main effects of data size ($F_{1,25} = 118.79$, $p < 0.001$, $\eta_p^2 = 0.83$) and configuration ($F_{2,50} = 97.56$, $p < 0.001$, $\eta_p^2 = 0.80$), as well as an interaction between the two ($F_{2,50} = 105.25$, $p < 0.001$, $\eta_p^2 = 0.81$). A post-hoc analysis analogous to the one performed above showed statistical differences between the interaction length in pixels between mobile ($M_{small} = 2381$px, $M_{large} = 39278$px) and other configurations (large display: $M_{small} = 849$px, $M_{large} = 1270$px; hybrid: $M_{small} = 645$px, $M_{large} = 2361$px), but not between hybrid and large display. Figure 4 shows the length of interaction in configurations.

To investigate the relationship between gaze shifts and completion time, we ran a regression test for the data in the hybrid configuration (the only one with gaze shifts). Trials from small and large data were analyzed separately to avoid causing the regression to appear significant due to the differences in task requirements.

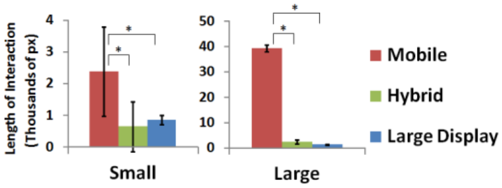

Figure 4. Map search: length of interaction (LI)

For small data, there is a linear relationship between the number of gaze shifts (#GS) and completion time (*CT(seconds)=1.8*#GS+ 8.21*) as shown in Figure 5. This regression is statistically significant ($F_{1,25} = 6.55$, $p < 0.05$) and explains 21% of the completion time variance ($R^2=0.214$). For large data, the regression is very similar (*CT=1.8*#GS+9.53*), and also significant ($F_{1,25} = 15.01$, $p < 0.01$), although it explains a larger portion of the CT variance ($R^2=0.385$, 38%).

4.3 Summary

In the map task, *configuration* made a large difference to performance; for data that fits within a mobile screen, mobile and large display are equivalent, and better than hybrid, whereas for large data, large display is faster than hybrid and mobile. The overall disadvantage of the hybrid condition seems to stem from the gaze shifts needed in this configuration. Our measures indicate that each shift might cost up to 1.8 seconds. The auxiliary analyses suggest that, for the small data, the main factor affecting performance was gaze shifts, which made hybrid worst.

The large data forced participants to pan much more in the mobile condition, making it the slowest; although this increased level of interaction was not necessary with the hybrid configuration (which showed low interaction), the gaze shifts in this configuration still mattered, making hybrid and mobile equivalently slow for different reasons.

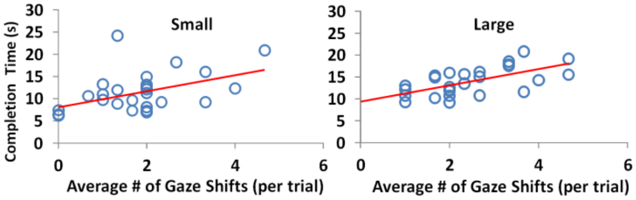

Figure 5. Map search: average gaze shifts (per trial) in the hybrid configuration.

These results put the large display configuration as the best option across all data sizes. Although this was recognized in the workload assessments (large display ranked as the least demanding configuration), more participants ranked hybrid above large display than the opposite when asked for their overall preference.

5. EXPERIMENT 2: TEXT SEARCH

This task was designed to represent text tasks such as those involved in search-engine use. Participants had to find and tap on page-result descriptors that contained a specific text fragment. The text to be found was presented at the top of the screens and was reproduced verbatim within the target page-result descriptors. Descriptors were arranged in columns (see Figure 6).

The text page-result descriptions consisted of Wikipedia featured-page text fragments of 30 words in average and were selected to have Flesch readability scores [12] between 50 and 62, representing text at 10-11th grade reading level. The 12-point Sans-Serif text was left-aligned. Each page-result text was 23x31cm on the large display, and 5.6x9cm on the mobile screen, using a similar number of pixels across displays. The search fragments to be found were, on average, 3 words long.

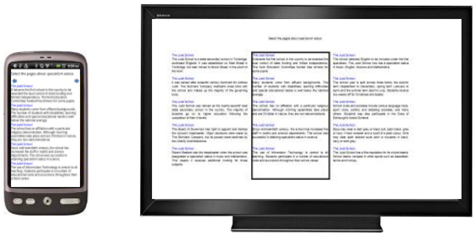

Figure 6. Text search task shown on mobile and large display

In *small* data size, participants had to select one page-result out of five displayed in a single column. In *large* data size, participants had to select three page-results containing the specified text among fifteen possible answers. In the mobile and hybrid conditions, the fifteen page-result texts were distributed in three pages. Two arrows at the top of the small screen indicated the presence of another page to the left, to the right or both (see Figure 6). Switching between pages was possible through the standard horizontal swipe gesture implemented in most modern touch devices. Paging was chosen over scrolling for three reasons: evidence shows that paging is better for text on small screens [26]; pilot tests showed that three pages were easier to navigate than a continuous vertical scroll; and, horizontal paging makes the large display configuration more equivalent in terms of control and visual configuration to the other two. Selection in the large display condition was achieved through a visible cursor controlled using the mobile device as buttonless trackpad. We considered alternative mechanisms based on block-selection (without a visible cursor) but we found that these were less familiar and performed worse in 3 participant pilot tests.

5.1 Results

In this experiment we excluded the data from four participants according to the *a priori* criteria.

5.1.1 Performance

An omnibus ANOVA with *configuration* and *data size* as main factors and participant as random factor revealed a strong effect of configuration ($F_{2,42} = 11.22$, $p < 0.001$, $\eta_p^2 = 0.35$), and data size ($F_{1,21} = 943.77$, $p < 0.001$, $\eta_p^2 = 0.98$) on completion time. Because no interaction was found between *configuration* and *data size* ($F_{2,42} = 1.42$, $p > 0.05$, $\eta_p^2 = 0.06$), we performed the post-hoc analysis of small and large data tasks together.

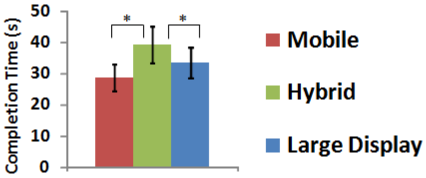

Figure 7. Text search: mean completion time (error bars indicate 95% confidence intervals)

For the text task, mobile was the fastest configuration (M=28.55s) followed by large display (M=33.26s, 16% slower) and hybrid (M=39.145s, 37% slower) configurations. Post-hoc tests corrected for multiple comparisons (Tukey's HSD) showed statistically significant differences between hybrid (the slowest) and the other conditions, but not between mobile and large display. Figure 7 shows a summary of the completion time data for the text task.

Error data was again analyzed using Friedman tests, which showed no statistically significant differences between configurations ($\chi^2(2) = 3.95$, $p > 0.05$). On average, the number of trials with errors was lowest in hybrid (M=2%), followed by mobile (M=4%) and large display (M=6%).

5.1.2 Subjective Evaluation

Responses to the questions were analyzed separately using Friedman tests. In this task we only found differences in performance ($\chi^2(2) = 6.21$, $p < 0.05$) as shown in Table 3.

Table 3. Text search: average ratings in the TLX questions.

Factor	χ² (2)	p	Mobile	Hybrid	Large D.
Physical Demand	4.48	>0.05	4.14	4.39	3.71
Mental Demand	4.73	>0.05	2.75	3.25	2.66
Temporal Demand	4.45	>0.05	3.86	3.68	2.98
Performance	6.21	<0.05*	8.34	7.45	8.41
Effort	4.86	>0.05	4.34	4.43	3.66
Frustration	1.69	>0.05	3.07	3.18	2.64

Table 4. Text search: configuration preference rankings.

	Best		Worst
Mobile	8	10	4
Hybrid	2	6	14
Large Display	12	6	4

In the overall ranking hybrid was the least preferred configuration (14/22 participants, 64%). The large display was the most preferred configuration (12/22, 54%) followed by mobile (8/22, 36%). A Friedman test of the rankings shows significant differences between the three configurations ($\chi^2(2) = 9.82$, $p < 0.01$). Table 4 shows the complete preference choices.

Most participants commented that they found it easier to scan text on the large display, although some preferred the mobile display for this task because they were not accustomed to reading on large screens. In general, participants said that they found themselves shifting attention between displays in the hybrid configuration, even though they realized that it did not help them. Note that participants were explicitly told at the beginning of these trials to make use of either or both displays as they pleased.

5.2 Auxiliary Analyses

An omnibus ANOVA of *length of interaction* with configuration and data size as main factors and participant as random factor yielded main effects of data size ($F_{1,21} = 25.31$, $p < 0.001$, $\eta_p^2 = 0.55$), configuration ($F_{2,42} = 23.24$, $p < 0.001$, $\eta_p^2 = 0.52$), as well as an interaction between the two ($F_{2,42} = 16.39$, $p < 0.001$, $\eta_p^2 = 0.44$). A post-hoc analysis (Tukey's HSD) showed statistical differences in the interaction length between large display (M_{small} = 2947px, M_{large} = 24815px) and the other configurations (mobile: M_{small} = 16px, M_{large} = 1921px; hybrid: M_{small} = 12px, M_{large} = 1973px), but not between mobile and hybrid. Figure 8 shows the length of interaction with different configurations and data sizes for text search.

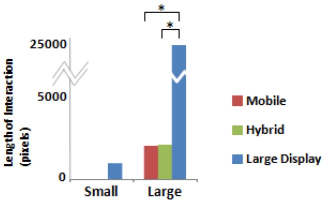

Figure 8. Text search: length of interaction. Hybrid and mobile LI for small data are too small to be visible.

The regression between gaze shifts and time was not significant in any data size ($F_{1,21} = 0.15$, $p > 0.05$; $F_{1,21} = 0.00$, $p > 0.05$).

5.3 Summary

In the text search task, configuration had a large impact on performance; mobile and large display were equivalent and hybrid was worst, regardless of whether the data fit the mobile screen or not. Although we did not find statistical evidence of a relationship between gaze shifts and completion time, we can only attribute the poorer performance of the hybrid condition to the availability of two possible foci of attention, regardless of the number of gaze shifts performed. Note that the length of interaction does not match the completion time results; even though length of interaction was at least an order of magnitude larger for the large display configuration, completion time was still equivalent to mobile, and 15% faster than hybrid.

The problems of the hybrid technique were reflected in the overall ranking of the technique by the participants. The tie in performance between mobile and large display is solved in favor of the latter in the subjective data.

6. EXPERIMENT 3: PHOTO SEARCH

The task for the photo search experiment is analogous to the text search task but with face photographs. A photo of a famous Hollywood female celebrity was displayed on top of the interface. Participants had to find and tap on the photos of the same person that appeared amongst photos from other celebrities. Photos were arranged on 3x6 grids per page (see Figure 9). In each trial participants had to find a different person. All photos shown were different, although some represented the same person (e.g., the photos for the correct answer). The photos had 106x159px (i.e. 7.8x5cm on the large display, 1.8x1.2cm on the mobile device).

In the *small data size* condition a single photo had to be selected among those in a 3x6 grid. For *large data*, there were three correct answers among three 3x6 grids. Paging and input were identical to those in Experiment 2.

Figure 9. Photo search task on mobile and large display.

6.1 Results

In this experiment we excluded the data from six participants according to the *a priori* criteria.

6.1.1 *Performance*

An omnibus ANOVA with *configuration* and *data size* as main factors and participant as random factor revealed a strong effect on completion time of configuration ($F_{2,38} = 4.29$, $p < 0.05$, $\eta_p^2 = 0.18$), and data size ($F_{1,19} = 933.58$, $p < 0.001$, $\eta_p^2 = 0.98$), as well as interaction between the two main factors ($F_{2,38} = 7.71$, $p < 0.001$, $\eta_p^2 = 0.29$). To follow up the interaction, we performed separate analysis on the large and small data conditions.

An ANOVA test on the small data trials showed that configuration had no significant effect on completion time ($F_{2,38} = 2.91$, $p > 0.05$, $\eta_p^2 = 0.13$). For the small data condition hybrid was the fastest configuration (M=8.89s), followed by large display (M=10.35s, 16% slower) and mobile (M=12.01s, 35% slower).

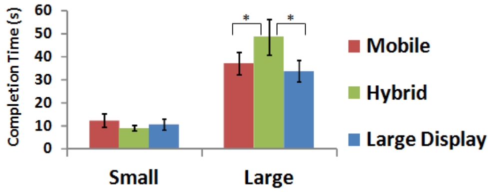

Figure 10. Photo search: completion time (error bars show 95% confidence interval).

An ANOVA test on the large data trials showed that configuration had a significant effect on completion time ($F_{2,38} = 12.33$, $p < 0.01$, $\eta_p^2 = 0.39$). For the large data condition, large display was the fastest configuration (M=33.68s), followed by mobile (M=37.19s, 10.4% slower) and hybrid (M=48.47s, 44% slower). Post-hoc tests corrected for multiple comparisons (Tukey's HSD) showed statistically significant differences between hybrid (the slowest) and the other conditions, but not between mobile and large display. Figure 10 shows completion times for photo search task.

Error data was strongly non-parametric and was analyzed using a Friedman test, which showed no significant differences between configurations ($\chi^2(2) = 1.73$, $p > 0.05$). On average, the number of trials with errors was lowest in hybrid (31%) followed by large display (32%) and mobile (36%) configurations.

6.1.2 *Subjective Evaluation*

We found significant differences among UI configurations on physical demand ($\chi^2(2) = 9.08$, $p < 0.05$) and mental demand ($\chi^2(2) = 7.25$, $p < 0.05$). These results along with the mean values of these ratings for different configurations are shown in Table 5.

For photo search, participants ranked large display as best (13/20, 65%) and ranked mobile as worst (10/20, 50%). Rank counts are shown in Table 6.

Table 5. Photo search: average ratings in the TLX questions.

Factor	χ^2 (2)	p	Mobile	Hybrid	Large D.
Physical Demand	9.08	<0.05*	5.03	4.85	3.9
Mental Demand	7.25	<0.05*	1.8	2.22	1.45
Temporal Demand	4.95	>0.05	3.85	3.37	3.18
Performance	4.82	>0.05	6.12	6.68	7.18
Effort	4.35	>0.05	4.5	3.9	3.25
Frustration	0.95	>0.05	2.93	2.38	2.92

Table 6. Photo search: configuration preference rankings.

	Best		Worst
Mobile	4	6	10
Hybrid	3	9	8
Large Display	13	5	2

6.2 Auxiliary Analyses

An omnibus ANOVA of *length of interaction* with configuration and data size as main factors and participant as random factor yielded effects of data size ($F_{1,19} = 101.49$, $p < 0.001$, $\eta_p^2 = 0.84$), configuration ($F_{2,38} = 28.51$, $p < 0.001$, $\eta_p^2 = 0.60$), as well as an interaction between the two ($F_{2,38} = 21.47$, $p < 0.001$, $\eta_p^2 = 0.53$). A post-hoc analysis (Tukey's HSD) showed statistical differences in interaction length between large display ($M_{small} = 872$px, $M_{large} = 4217$px) and the other configurations (mobile: $M_{small} = 11$px, $M_{large} = 1115$px; hybrid: $M_{small} = 14$px, $M_{large} = 1132$px), but not between mobile and hybrid. Figure 11 shows the mean length of interaction with small and large data.

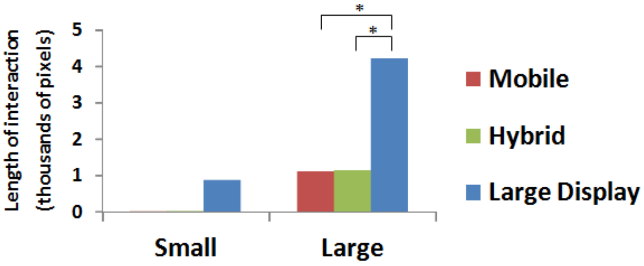

Figure 11. Photo search: length of interaction.

To investigate the relationship between gaze shifts and completion time, we ran a linear regression test for the data in the hybrid configuration which was not significant in either data size ($F_{1,19} = 2.34$, $p > 0.05$; $F_{1,19} = 1.59$, $p > 0.05$).

6.3 Summary

In the photo search task, mobile and large display performed equivalently, despite the large differences in interaction length. The large data condition brought out the differences in performance between techniques, which clearly showed that hybrid is the worst option. Paradoxically, the mobile configuration was rated by participants as inferior to hybrid for all workload questions, and ranked very similarly.

7. DISCUSSION

7.1 Distributing visual output has a cost

In principle, the hybrid configuration allows people to take advantage of two very different displays, which could allow users to make the most of both for different aspects of the task at hand. For example, a user may employ the large display to get an overview of the data space (which also enables physical navigation [1]), and the mobile display when direct input and a more flexible positioning of the device (e.g., bringing the display closer to the eyes) is beneficial [4].

However, our experimental results indicate that distributing the visual output incurs significant performance overheads; hybrid showed the worst performance in the text and photo search and equivalent to worst in the map search. We now explore several factors which might explain this disadvantage including, attention shift, visual/input space mismatch and distraction.

Attention Shifts. In general, a user can only look at one display at a time. Shifts of attention require gaze shifts which force the user to quickly adapt to a display that is at a different distance [31], has a different resolution, and where visual objects are shown at a different size. Moreover, the visual content represented on the mobile display is often a subset of what a large display can show (with large data). Typically, this requires a user to reorient themselves within the new visual space after a transition, i.e., to find what they were looking at in the previous display. The necessary mapping operation between the two visual representations might have an added cognitive cost. We were able to observe a direct relationship between the number of gaze shifts and the completion time for the map task which suggests that each shift can cost an average of 1.8s for this specific task and context. Smaller (but comparable) time costs are shown in [27] for a lower-level multi-display targeting task (580ms). Part of the shifting cost might stem from the additional cognitive effort involved in deciding whether and when to switch displays. The design of multi-display applications to support particular tasks should carefully consider the inherent attention shift costs versus the projected interaction design benefits.

Visual/input space mismatch. Because the two displays show different views of the same information space, a mismatch can arise between the region being looked at on the large display and the information currently visible on the mobile display. Since input is only available on the mobile device, once users locate the appropriate element on the large display they need to explicitly align their mobile view of the information to the focus of attention on the large display. In other words, there is a duplication of the navigation task: physical navigation with body, head and eye movement; and virtual navigation through explicit interaction. We observed instances of this behavior in the video, and three participants explicitly commented on it. Researchers and designers considering this configuration should develop and employ techniques to aid the user in better aligning their mobile view with their locus of attention on the large display.

Distraction. It is also possible that the mere presence of a large display, sometimes with moving elements (the scope window frame) is a source of distraction that users cannot avoid looking at. This would explain why most participants used both displays in the hybrid configuration even when the data also fit the mobile display, and even though performance was poorer. This suggests the design of multi-display applications needs to carefully consider the overall distraction factor versus scope for improved interaction within and across tasks.

These experiments were not designed to tease out which of these cost elements is dominant for explaining the poorer performance of hybrid configurations across all tasks. However, they helped identify possible culprits, and provide evidence that, for these tasks, the overheads introduced by the availability of two sources of visual output overpower any advantages of having dual views.

7.2 Tasks and Configurations

The primary goal of this study was to investigate which configurations are best for which tasks. The large display configuration was best or equivalent to best in all tasks and data sizes. The mobile-only configuration was worst in the map task and equivalent to best in the photo and text tasks.

These results also help clarify the role of interaction in the distribution of the interface; although the separation of input and output has been extensively studied as a factor influencing performance we found that considerable differences in the amount of interaction for the photo and text tasks did not impact on the corresponding completion time differences. The large display configuration required amounts of interaction orders of magnitude above mobile and hybrid but still performed best or equivalent to best. This suggests that for tasks that do not require continuous navigation such as photo and text search, how we distribute visual information across displays is more relevant than the directness of the interface (i.e., direct vs. indirect input).

Surprisingly, the hybrid configuration was ranked more favorably by participants than how it corresponds to its performance or their own subjective workload ratings; hybrid was clearly ranked worst for the text task, but it came close second for the map task and was second in the photo task even though performance was substantially inferior than mobile. Although this anomaly requires further study, we speculate that users may like the freedom of choosing which display to use, regardless of performance.

7.3 Large display indirect vs. small dis. direct

Our results also provide a valuable comparison between the mobile and large display configurations. Large display was, on average, ranked above mobile in all three tasks, and participants consistently rated it as better performing and less demanding across all 18 Likert scales.

The possible explanations for these results are varied. First, it is better to have more simultaneous pixels without the need to interactively manipulate the viewport. This is in line with results from previous studies on vertical large display scenarios [1], and suggests that eye/head/body navigation is superior to interactive navigation. This is not merely a resolution issue: our maps, texts and photos all had the same resolution and layout regardless of the configuration. Second, it is possible that for a given fixed amount of visual information people prefer to interact with larger objects. Third, people might prefer the large display because it is more stable, and does not require looking down or holding the device at a certain angle or distance.

Importantly, we found no evidence suggesting that the loss of ‘directness’ between input and output is important in the discussed scenario, or that the ability of users to place the visual output of a mobile device anywhere in their field of view is beneficial for these tasks.

7.4 Limitations

Although these experiments cover a broad range of tasks that are common in mobile scenarios, there are other tasks such as text composition, photo manipulation or web browsing that might be affected differently across the various configurations. Exploration of these tasks will be valuable to expand and complete these results. Due to the limited nature of experimental design, we only tested two levels of data size; further experiments with data that requires viewport changes on a large display should be valuable in improving the understanding of how to support interaction with very-large-data in mobile scenarios.

The configurations chosen are representative of the single- and multi-display options that are already feasible in many scenarios (e.g., wherever there is a large public display and a communication channel between the mobile and the large display devices). However, by excluding configurations that require input sensing in the environment we constrained ourselves to scenarios that are already feasible with widely installed infrastructure. Other alternatives such as direct touch on the large display [24], the use of the mobile devices as ray-pointing input [27, 20], or more sophisticated combinations of small and large display content [4] fall outside the scope of this research and need to be addressed in the future. Although we did not explicitly consider the use of mobile projected large displays (e.g., [14, 17]) we believe that the results found in our study provide valuable guidance to the effects that different ways of distributing the interface might have on future mobile projector-based systems.

7.5 Lessons for Practitioners

Our study provides useful evidence for the design of mobile and multi-display distributed interfaces:

- Large displays can be used in mobile tasks to enhance user performance and preference, and reduce perceived workload.
- Showing same visual elements of an interface across displays should be avoided for single-user tasks.
- Tasks that require continuous navigation of the data space (e.g. map search) benefit most from the addition of a large display.

8. CONCLUSIONS

Supplementing a mobile device with large displays found in the environment has been widely considered as an opportunity to overcome some of the limitations of small mobile displays. However, there is very little evidence to guide designers on how to distribute the interface in these cases. In this paper we provide empirical evidence of how different distribution of input/output across mobile and large display affects user performance, subjective workload and subjective preferences in a wide range of common mobile tasks. We show that mobile-controlled large displays are generally the best option across the tasks, followed by not using a large display at all. Although we observed that a configuration with distributed visual content was worst or equal to worst across tasks, participants seemed to prefer it in many cases to the mobile-only baseline. We have also suggested several sources of overhead derived from splitting the interface across different displays, backed by empirical evidence, and provided advice for practitioners on how to apply our findings to design the next generation of multi-display environments.

9. REFERENCES

[1] Ball, R., North, C. Visual analytics: Realizing embodied interaction for visual analytics through large displays. *Computer Graphics* 31, 3 (June 2007), 380–400.

[2] Ballagas, R., Rohs, M., Sheridan, J., Borchers, J. The smart phone: A ubiquitous input device. *IEEE Per. Comp.* 5, 1 (J-M 2006), 70–77.

[3] Bi, X., Bae, S.-H., Balakrishnan, R. Effects of Interior Bezels of Tiled-Monitor Large Displays on Visual Search, Tunnel Steering, and Target Selection. *Proc. CHI* (2010), 65–74.

[4] Boring, S., Baur, D., Butz, A., Gustafson, S., Baudisch, P. Touch projector: mobile interaction through video. *Proc. CHI* (2010), 2287–2296.

[5] Cauchard J., Lchtefeld, M., Irani, P., Schning, J., Fraser, M., Kruger, A., Subramanian, S. Visual separation in mobile multi-display environments. *Proc. UIST* (2011), 451–460.

[6] Cheverst, K., Dix, A., Fitton, D., Kray, C., Rouncefield, M., Sas, C., Saslis-Lagoudakis, G., Sheridan, J. G. Exploring bluetooth based mobile phone interaction with the hermes photo display. *Proc. MobileHCI* (2005), 47–54.

[7] Chittaro, L. Visualizing information on mobile devices. *Computer 39* (March 2006), 40–45.

[8] Cockburn, A., Karlson, A., Bederson, B.B., A review of overview+detail, zooming, and focus+context interfaces. *Comput. Surv.* 41, 1 (Jan 2009), 1–31.

[9] Deller, M., Ebert, A. Modcontrol mobile phones as a versatile interaction device for large screen applications. *Proc. INTERACT* (2011), 289–296.

[10] Dix, A., Sas, C. Mobile Personal Devices Meet Situated Public Displays: Synergies and Opportunities. *Int. J. Ubi. Comp.*, 1, 1 (2010), 11–28.

[11] Finke, M., Kaviani, N., Wang, I., Tsao, V., Fels, S., Lea, R. Investigating distributed user interfaces across interactive large displays and mobile devices. *Proc. AVI* (2010), 413.

[12] Flesch, R. A new readability yardstick. *J. Appl. Psy.* 32, 3 (1948), 221–233.

[13] Gostner, R., Gellersen, H., Kray, C., Sas, C. Reading with mobile phone and large display. *Proc. Workshop on Designing and Evaluating Mobile Phone-Based Interaction with Public Displays* at CHI (2008).

[14] Greaves, A., Rukzio, E. Evaluation of Picture Browsing using a Projector Phone. *Proc. MobileHCI* (2008), 351–354.

[15] Greenberg, S., Boyle, M., Laberge, J. PDAs and shared public displays: Making personal information public, and public information personal. *Pers. & Ubi. Comp.* 3, 1 (1999), 54–64.

[16] Grudin, J. Partitioning digital worlds: focal and peripheral awareness in multiple monitor use. *Proc. CHI* (2001), 458–465.

[17] Hang, A., Rukzio, E., Greaves, A. Projector Phone: A Study of Using Mobile Phones with Integrated Projector for Interaction with Maps. *Proc. Mobile HCI* (2008), 207–216.

[18] Hosio S., Jurmu, M., Kukka, H., Riekki, J., Ojala, T. Supporting distributed private and public user interfaces in urban environments. *Proc. HotMobile* (2010), 25–30.

[19] Hutchings, R. Stasko, J., Czerwinski, M. Distributed display environments. *Interactions* 12, 6 (Nov. 2005), 50–53.

[20] Jota, R., Nacenta, M., Jorge, J., Carpendale, S., Greenberg, S. A comparison of ray pointing techniques for very large displays. *Proc. GI* (2010), 269–276.

[21] Luyten, K., Coninx, K. Distributed user interface elements to support Smart interaction spaces. *Proc. IEEE ISM* (2005), 277–286.

[22] Melchior, J., Grolaux, D., Vanderdonckt, J., Van Roy, P. A toolkit for peer-to-peer distributed user interfaces: concepts, implementation, and applications. *Proc. EICS* (2009), 69–78.

[23] McCallum, D.C., Irani, P. Arc-pad: absolute+relative cursor positioning for large displays with a mobile touchscreen. *Proc. UIST* (2009), 153–156.

[24] Nacenta, M., Gutwin, C., Aliakseyeu, D., Subramanian, S. There and back again: Cross-display object movement in multi-display environments. *JHCI* 24, 1-2 (2009), 170–229.

[25] Nacenta, M., Sakurai, S., Yamaguchi, T., Miki, Y., Itoh, Y., Kitamura, Y., Subramanian, S., Gutwin, C. E-conic: a perspective-aware interface for multi-display environments. *Proc. UIST* (2007), 279–288.

[26] Öquist, G., Lundin, K. Eye movement study of reading text on a mobile phone using paging, scrolling, leading, and RSVP. *Proc. MUM* (2007), 176–183.

[27] Rashid, U., Jarmo, K., Häkkilä, J., Quigley, A. Proximal and Distal Selection of Widgets: Designing Distributed UI for Mobile Interaction with Large Display. *Proc. MobileHCI* (2010), 495–498.

[28] Rekimoto, J. A multiple device approach for supporting whiteboard-based interactions. *Proc. CHI* (1998), 344–351.

[29] Remote Droid - http://remotedroid.net/

[30] Thevenin, D., Coutaz, J. Plasticity of user interfaces: Framework and research agenda,.*Proc. INTERACT* (1999), 110–117.

[31] Tan, D., Czerwinski, M. Effects of visual separation and physical discontinuities when distributing information across multiple displays. *Proc. INTERACT* (2003), 252–255.

[32] Terrenghi, L., Quigley, A., Dix, A. A taxonomy for and analysis of multi-person-display ecosystems. *Pers. & Ubi. Comp.* 13, 8, (Nov. 2009), 583–598.

[33] Yang, X., Mak, E., McCallum, D., Irani, P., Cao, X., Izadi, S. Lensmouse: Augmenting the mouse with an interactive touch display. *Proc. CHI* (2010), 2431–2440.