

Spatially-aware or Spatially-agnostic? Elicitation and Evaluation of User-Defined Cross-Device Interactions

Roman Rädle^{1*}, Hans-Christian Jetter^{2*}, Mario Schreiner¹, Zhihao Lu³,
Harald Reiterer¹, Yvonne Rogers³

¹HCI Group, University of Konstanz, Konstanz, Germany

²Intel ICRI Cities, University College London, London, United Kingdom

³UCL Interaction Centre, University College London, London, United Kingdom

{Roman.Raedle,Mario.Schreiner,Harald.Reiterer}@uni-konstanz.de

h.jetter@ucl.ac.uk, zhihao.lu.13@alumni.ucl.ac.uk, y.rogers@ucl.ac.uk

* First two authors contributed equally to this work

ABSTRACT

Cross-device interaction between multiple mobile devices is a popular field of research in HCI. However, the appropriate design of this interaction is still an open question, with competing approaches such as *spatially-aware* vs. *spatially-agnostic* techniques. In this paper, we present the results of a two-phase user study that explores this design space: In phase 1, we elicited gestures for typical mobile cross-device tasks from 4 focus groups (N=17). The results show that 71% of the elicited gestures were *spatially-aware* and that participants strongly associated cross-device tasks with interacting and thinking in space. In phase 2, we implemented one *spatially-agnostic* and two *spatially-aware* techniques from phase 1 and compared them in a controlled experiment (N=12). The results indicate that spatially-aware techniques are preferred by users and can decrease mental demand, effort, and frustration, but only when they are designed with great care. We conclude with a summary of findings to inform the design of future cross-device interactions.

Author Keywords

Cross-device interaction; user-defined gestures.

INTRODUCTION

Cross-device interaction between multiple mobile devices is an increasingly popular field of research in HCI [5,8,16,17,18,22,26]. It can be regarded as the latest incarnation of Weiser's vision of ubiquitous computing [35] in which user experiences truly begin to *cross* devices [8] and the co-located devices can be easily joined to create *ad-*

hoc device communities [13]. Ideally, users experience such a community as a single seamless and natural UI (or even a "symphony of devices" [8]) that is flexible in terms of use and is not restricted to a few possible configurations or predefined sequences of use [13,26].



Figure 1. Our experimental setup for multi-tablet cross-device interaction.

Applications of cross-device interaction have become increasingly diverse, ranging from collaborative photo sharing or brainstorming with smart phones [17,18] to multi-tablet active reading [4,5] and sensemaking [8]. User studies have shown that multi-tablet systems can be successfully used in the wild [5] and that users can effectively manage cross-device interactions with 5 to 10 devices [8]. However, many questions remain unanswered: How should cross-device interaction between mobile devices be designed so that they are easy to learn and easy to use? What role should increasingly popular technologies for sensing spatial configurations and detecting mid-air gestures play in their design? Should interactions follow a traditional, yet robust, non-spatial model, e.g., menu-based selection of devices [8]? Or should systems sense locations and use gestures to make cross-device interactions more like familiar non-digital interactions [7]?

These questions are especially important because our devices are still rather limited in terms of sensing their mutual spatial relations [8] without the use of expensive or custom-built sensing hardware such as rooms equipped with motion tracking systems [19]. However, such spatial information could enable interfaces to easily grow across nearby devices and annex them in natural ways [11,25], ideally as a byproduct of natural use in space (e.g., by putting tablets on a table, moving them around, placing them side-by-side, making pick-and-drop gestures between them [26]). Only very recently have new low-cost sensing solutions based on overhead depth cameras such as HuddleLamp [26] become available; these solutions can track not only the positions and movements of off-the-shelf mobile devices on tables but also above-the-table hand gestures. The availability of these new technologies calls for more research on *spatially-aware* vs. *spatially-agnostic* interactions, particularly since the necessary sensing technology was still considered futuristic and far beyond the state of the art just one year ago [8].

We therefore conducted a two-phase user study to better explore the design space for future cross-device interaction with multiple mobile devices. In phase 1 of the study, we conducted four elicitation sessions, each with a group of 4-5 participants. In each session, we elicited user-defined gestures for 19 typical cross-device tasks, ignoring current technological restrictions and leaving the question of *spatially-aware* vs. *spatially-agnostic* interaction to users. In phase 2, we used a low-cost tracking system to implement one *spatially-agnostic* and two *spatially-aware* techniques that were suggested in phase 1 and evaluated them in a controlled experiment with 12 participants.

In the following sections, we introduce previous and related work and report on our two-phase user study and its results. We conclude by summarizing our findings in the form of design recommendations for future cross-device systems.

BACKGROUND AND RELATED WORK

Today, performing tasks across multiple devices is often tedious [7], but new UIs promise to achieve a more natural use of multiple devices [4,8] through (1) new interaction techniques, (2) better sensors, and (3) user-defined gestures.

(1) Cross-Device Interaction Techniques

We differentiate between three main categories of cross-device interactions: *synchronous gestures*, *spatially-agnostic interactions*, and *spatially-aware interactions*.

Synchronous Gestures – Chen et al. propose a multi-tablet system for single-user reading activities that uses synchronous “conduit” interactions to move information between devices [4,5]. These interactions employ temporal simultaneity and sequences to express directed cross-device actions. For example, users might first designate a target by touching a device with their non-dominant hand, and then use their dominant hand (which offers more precision) to tap the item to be transferred. This approach is loosely

based on Hinckley’s and Rekimoto’s pioneering work that suggested synchronous gestures such as device bumping [12], pen-based stitching [11], or synchronous tapping [29] to exchange content or create multi-device tiled displays. Similarly, Lucero et al. used synchronous touch-based “pinching” across phones to create multi-device huddles in [17,18].

Spatially-Agnostic Interactions – In contrast, Hamilton and Wigdor’s multi-tablet system Conductor uses traditional menus with color-coded device names to select the tablets to share information with or to chain tasks across them [8]. Thus, in Nacenta et al.’s terminology [22], Conductor’s referential domain for selecting devices is “non-spatial”. However, Hamilton and Wigdor’s own experimentation and user study revealed that keeping track of multiple, often very similar devices can represent a surprisingly significant challenge, and also that users extensively used spatial configurations of tablets for categorical organization [8]. Therefore, it is possible that a non-spatial (or *spatially-agnostic*) interaction might diminish the benefits of cross-device interaction: It could create a mismatch between the spatial referential domain in which the user’s intention is expressed and the non-spatial way in which the interaction technique requires the user to select a destination device [22]. In other words, although menus are a robust and familiar way to select items that have no clear spatial relation, they might seem cumbersome for the purpose of selecting one out of many devices from a spatial configuration.

Spatially-Aware Interactions – An alternative involves spatially-aware interaction techniques that use real-world spatial configurations as the referential domain – for example, hyperdragging or pick-and-drop of objects between laptops and table surfaces [28] and flicking/throwing objects within AR settings [33], tabletops [27], or from phones towards large displays [6]. Throwing and flicking techniques are also frequently named by participants in gesture elicitation studies for multi-display interactions (e.g., [15,31]), including our own elicitation study in this paper. This popularity of spatial referential domains resonates with user studies that have observed how important space and spatial configurations are as meaningful cognitive resources during knowledge work in offices [14] or sensemaking on large screens [1]. However, in our elicitation study, participants also raised concerns about the accuracy of throwing and flicking and the danger of inadvertently sending content to the wrong device. This is particularly relevant in mobile cross-device interactions in which the target screens are relatively small and/or far away. More reliable spatially-aware techniques include world-in-miniature or radar views that allow target regions or devices to be selected in a top-down map-like representation of the environment [3,27,36]. User studies on tabletops have revealed that these approaches can be more accurate (albeit slower) than flicking [27].

(2) Sensing Space and Proximity

To achieve spatially-aware cross-device interactions, it is necessary to accurately track the current positions of all devices. One solution involves a camera-based motion capturing system [19], a typically expensive method requiring to instrument rooms, devices, and users; this is generally considered futuristic in the context of present-day mobile scenarios [8]. A low-cost alternative is using the front facing cameras of mobile devices to track positions by detecting fiducial markers on the ceiling [16]. Lucero et al. equip mobile phones with sensors for radio trilateration [17], and Marquardt et al. combine this with overhead depth cameras [20] to sense locations. This approach is further improved by Rädle et al.'s HuddleLamp, which enables tracking of uninstrumented tablets and phones using only an overhead depth camera [26]. Due to its low-cost and ready availability, we chose HuddleLamp for the experiment in phase 2 of our study.

(3) Elicitation of User-defined Gestures

Nielsen et al. propose gesture elicitation studies for the design of intuitive and ergonomic gestural interfaces and to avoid arbitrary gesture sets that are optimized for reliable recognition by technology rather than for ease of learning and use by humans [24]. Similarly, Nacenta et al. find that user-defined gestures are preferred by users and are also easier to remember [23]. Wobbrock et al. successfully elicited multi-touch gestures for typical tasks from non-technical users [37]; since then, many similar studies have explored connecting phones, public displays, and tabletops [15], for diagram editing with multi-touch and pen [10], for multi-display environments [31], for active tokens querying big data [32], or for skin input [34]. As we describe below, our work substantially differs from this research in two respects: First, we are not primarily interested in a single, ideally "optimal" gesture set, but rather in a wide variety of user suggestions and deep insights into users' underlying thinking and metaphors. Second, we do not stop at eliciting gestures, but additionally evaluate them in a controlled experiment to learn more about their cognitive and ergonomic properties during repeated use. To our knowledge, only [10,24,33] have pursued this approach, but in domains other than cross-device interaction.

PHASE 1: GESTURE ELICITATION STUDY

Phase 1 of our study sought to elicit user-defined cross-device gestures during the course of 4 focus groups. We explain why we decided on this rationale and why and how our methodology differs from traditional gesture elicitation.

Overview and Rationale

To learn more about users' ideas, preferences, expectations, and mental models for cross-device gestures, we decided to prompt users with typical cross-device tasks and then elicit suggestions from them for the corresponding cross-device interactions. In order to reduce bias and increase their creativity, we primed them with a video showing the latest

cross-device techniques; the participants were then asked to be imaginative and to ignore any technological restrictions they might know about. We also avoided commenting on the feasibility of their suggestions during the discussions.

Three main topics guided this first phase of our study:

1.) One great advantage of *synchronous gestures* such as SyncTap [29], conduit [4], bumping [12], stitching [11], and pinching [17] is that they only need built-in sensors. However, these gestures must be learned and executed by the users across screens in the correct sequence and timing. How "intuitive" are such synchronous gestures, and would users – rather than designers – suggest them themselves?

2.) *Spatially-agnostic interactions* such as device selection menus are familiar from GUIs and likely to be suggested by users (see "legacy bias" discussed in [21]). However, as described above, non-spatial selection is also potentially cumbersome [22]. Would users be aware of this limitation, and what would their suggestions and opinions be?

3.) *Spatially-aware interactions* such as throwing or flicking are popular suggestions in gesture elicitation studies (e.g., in [15,31]). They are fast and efficient [27], both because they employ an open-loop control paradigm [22] and because they make use of the user's natural understanding of space and physical movement. However, they are also less precise than other spatially-aware interactions (e.g., the top-down world-in-miniature or radar view representations in [3,27,36]) that offer closed-loop control [22] with higher accuracy but slower interactions [27]. We wanted to learn about users' preferences and opinions on this issue and to see whether they would identify speed and precision as important criteria.

Reducing Bias with Partners, Production, and Priming

To reduce bias, we did not introduce our three guiding topics of interactions to the users; we only applied them afterwards to categorize the users' suggestions, to analyze their verbal comments and opinions and their feedback on questionnaires. We also decided against a traditional gesture elicitation study (as in [37]), instead opting for an approach similar to Morris et al.'s proposal to reduce legacy bias through *partners*, *priming*, and *production* [21].

We used *partners* – i.e., focus groups of up to 5 partners in each session – to collect as many different suggestions, comments, and explanations as possible. By enabling partners to fruitfully build upon one another's ideas and asking them to decide on a single preferred interaction, we hoped to facilitate more reflection and discussion and to elicit more diverse opinions about the designs. Similar to [20,33], our sessions therefore contained an element of co-creation instead of pure elicitation. As a result, we received many novel and elaborate suggestions, including details on physical input and visual output.

We also employed *production* and *priming* [21]. *Production* was promoted by requiring groups to produce at least 3

proposals for each task before choosing their favorite, in an attempt to move beyond a few simple, legacy-inspired techniques. *Priming* was implemented by showing each group an introductory video that depicted a variety of the latest cross-device interactions in order to reduce the group's bias towards legacy-inspired GUI interactions. In addition, users were encouraged to perform their suggestions with physical props such as tablets, pens, and paper so that they would think about the capabilities and affordances of mobile form factors instead of technological restrictions. Nevertheless, we still followed a strict formal procedure during elicitation with carefully selected materials, questionnaires, and tasks.

Task Set

An initial set of tasks was extracted from cross-device systems and elicitation studies in the literature [4,8,15,26,31,32]. This initial set of 22 tasks was intended to represent the most typical and relevant cross-device tasks. In a pilot study, we then identified redundant tasks and those tasks that were too complex for non-technical users to understand. After removing these tasks, the final set contained 19 tasks (see Table 1).

| T# | Function | Object | Source | Destination | Distance |
|----|-----------|----------------|---------------|-------------|----------|
| 1 | Move | File | Tablet | (Tablet) | In reach |
| 2 | Move | File | Tablet | (Tablet) | Far |
| 3 | Move | File | (Tablet) | Tablet | Far |
| 4 | Move | File | Phone | (Tablet) | In reach |
| 5 | Move | File | Phone | (Tablet) | Far |
| 6 | Move | File | (Tablet) | Phone | Far |
| 7 | Copy | File | Tablet | (Tablet) | In reach |
| 8 | Copy | File | Tablet | (Tablet) | Far |
| 9 | Copy | File | (Tablet) | Tablet | Far |
| 10 | Expand | View | Tablet | (Tablet) | In reach |
| 11 | Duplicate | Screen | Tablet | (Tablet) | In reach |
| 12 | Duplicate | Part of Screen | Tablet | (Tablet) | In reach |
| 13 | Duplicate | Screen | Tablet | (Tablet) | Far |
| 14 | Duplicate | Part of screen | Tablet | (Tablet) | Far |
| 15 | Open | File | Tablet | (Tablet) | In reach |
| 16 | Open | File | Phone | (Tablet) | In reach |
| 17 | Connect | Keyboard | - | Tablet | In reach |
| 18 | Copy | All files | (All tablets) | Tablet | Far |
| 19 | Copy | All files | (All tablets) | Phone | Far |

Table 1. Set of 19 tasks used for the elicitation study.

Tasks 1-9 represent typical cross-device object movements. Tasks 10-14 deal with stitching and duplicating screens. Tasks 15-19 are miscellaneous tasks such as pairing a wireless keyboard with a tablet or copying all files from all other devices to the personal device. In the table, "source" and "destination" define the devices involved and the direction. Parentheses mean that a remote device, e.g., "(Tablet)", is not held in the hands of the users and is not lying directly in front of them. A distance of "in reach" means that the device is within an arm's length, while "far" means that users have to stand up and walk to reach it.

During the study, we prompted each group with one task at a time. Each prompt was an animation that first showed the

starting point and then the outcome of the task. To avoid bias, the animation did not show any user interactions. For example, one prompt first showed two tablets lying on a table, with an object on the screen of the first tablet. Then it showed this object disappearing and appearing on the screen of the second tablet, without hinting at any possible user interaction.

Participants and Groups

We recruited 17 participants (7 female) through mailing lists and posters on a university campus. The age of participants ranged from 18 to 43 years (mean=26.4, SD=7.2). All participants had several years of experience with using a smart phone (mean=3.9, SD=2.7). 13 participants also had experience with using a tablet (mean=2.2, SD=1.1). After reviewing the pool of participants, we manually assigned them to four groups to avoid overly heterogeneous focus groups. Group A (n=4) were undergraduate students in computer science. Group B (n=4) were researchers in computer science or professional software developers. Group C (n=5) and Group D (n=4) were students from non-technical subjects such as comparative literature, anatomy, neuroscience, linguistics, architecture, and financial risk management.

Procedure

In each session, the group sat around a table with various switched-off tablets, sheets of paper, and marker pens (in case participants wanted to demonstrate or sketch their suggestions). After the introductory video was shown, the recording of the elicitation began, and the following procedure was repeated for each of the 19 tasks: First, the animated prompt for the task was shown on a projector or large screen. Second, the group was asked to think of corresponding interactions and discuss them with their fellow group members. The groups were asked to produce at least 3 different interactions. Third, the group chose one favorite interaction for the given task. Finally, each group member filled out a questionnaire with 7-point Likert scales on the understandability of the task, their personal agreement with the group's selected favorite, and how difficult it had been to propose an interaction. After the 19 tasks, there was a debriefing and a closing discussion. Participants were given post-test questionnaires with 7-point Likert scales to determine whether they had always understood what they had been asked to do and whether they had felt that they could express their ideas during the session. Each session lasted between 1.5 and 2 hours, and participants were compensated for their time with £20.

Results & Discussion

The questionnaires revealed that participants had no problems understanding what they needed to do (\bar{x} =6.76, SD=0.75) or expressing their ideas (\bar{x} =6.06, SD=1.25). The agreement and difficulty ratings of each task revealed high overall agreement (\bar{x} =5.99, SD=1.15) and low to neutral difficulty (\bar{x} =3.57, SD=1.67). Unsurprisingly, the tasks with

lowest agreement and highest difficulty were tasks 3, 9, 18, and 19, all of which involved retrieving objects from one or multiple sources outside one's reach. However, there were no indications of more general problems with the study's design, its social setting, or the difficulty of the task set.

Synchronous, Spatially-Aware, and Spatially-Agnostic

After the analysis and thematic coding of the video recordings, we took the favorite gestures from each group and all tasks (19 tasks \times 4 groups = 76 favorites) and categorized them: 12 favorites (15.8%) were *synchronous gestures* (e.g., bumping devices), 10 favorites (13.2%) were *spatially-agnostic interactions* (e.g., selecting a target device from a menu of devices by name), and 54 favorites (71.1%) were *spatially-aware interactions* (e.g., flicking an item to a remote device).

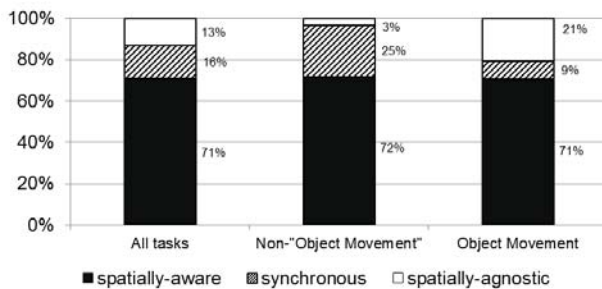


Figure 2. Categorization of favorites.

We checked whether the predominance of *spatially-aware interactions* was due to the many cross-display object movement tasks in the task set (tasks 1-9, 18, 19) by excluding them. For the non-“object movement” tasks (10-17), *spatially-aware interactions* were again the most popular at 71.9%, *synchronous gestures* were second at 25%, and *spatially-agnostic interactions* third at 3.1% (Figure 2). This clearly indicates how strongly participants associated cross-device tasks with interactions in space and how much their thinking and suggestions were of a spatial nature. Although they were already familiar with some of the *spatially-agnostic* or *synchronous* cross-device techniques that are popular today, *spatially-aware interactions* still seemed to be the most appealing, even for non-“object movement” tasks.

Synchronous gestures were most popular for task 10 “expanding view to other device” (4 of 4 groups) and task 17 “pairing keyboard and tablet” (2 of 4 groups). In general, participants seemed to use *synchronous gestures* when they wanted to refer to the device itself or the entire screen, but not for cross-device object movement tasks. For such object movement tasks, only 4 favorites (9.1%) were *synchronous gestures* (mostly similar to the “conduit” gesture [4]), while 20.5% were *spatially-agnostic* (Figure 2). Overall, *synchronous gestures* seemed to be most important in the context of expanding screens or pairing devices, but were not popular for object transfer tasks.

We were surprised that *spatially-agnostic interactions* were almost as popular as *synchronous gestures* for all tasks and twice as popular for object movement tasks (20.5%). They were particularly popular in Group C, which had a heterogeneous, non-technical background. While this could be seen merely as a case of strong legacy bias [21], it also hints at the ongoing importance and great practical relevance of more traditional menu-based interactions for cross-device interaction (e.g., as demonstrated in [8]). This also resonates with the surprisingly good performance of menu-based techniques that we observed later in phase 2.

Suggestions for Spatially-Aware Interactions

We further analyzed the 54 favorites involving spatially-aware interactions. 25 favorites (46.3%) entailed open-loop flicking/throwing gestures between devices. Participants discussed their potential limitations with regard to precision and control. For example, Group D addressed the problem of inadvertently sending content to the wrong person or device in a room with 4 or 5 other tablets. Group C also discussed the idea that imprecise flicking might result in content ending up on the wrong tablet, and Group A brought up the problem of how to flick content between two devices when a third device is situated between them.

Groups B and D suggested a *slingshot metaphor* instead of merely flicking/throwing. Inspired by games like Angry Birds, they suggested that direction and the force/distance of flicking could be better controlled when an item is first pulled back from its current position with a finger on the touch screen and then launched in the opposite direction when the finger is lifted. They also considered additional visual output during aiming, such as highlighting the prospective target device to allow more control; the slingshot would thus become an intermittent or closed-loop control paradigm rather than an open loop [22].

Another suggestion involved using visual proxies to represent remote devices on the local screen. This was suggested in 18 (33.3%) of the 54 spatially-aware favorites. For example, Group A suggested that all remote devices be represented as bubbles on the edges of the local screen. These bubbles would appear where the imaginary line between the center of the local device and the center of the remote device intersected with the local screen's boundaries and could be used as proxy targets for drag-and-drop or flicking to remote devices. Alternatively, Group C suggested an overhead map or radar view containing live representations of all devices at their current locations as proxy targets (similar to [3,27,36]).

Conclusions and Input for Phase 2

Given the significant role that both *spatially-aware* and *spatially-agnostic* interactions played in participants' suggestions, we decided to further explore such techniques in phase 2 of our study. We decided to implement the two spatially-aware techniques (*edge bubbles* and *radar view*) that users suggested to address the problem of insufficient

control in open-loop flicking/throwing. Furthermore, given that *spatially-agnostic* interactions played a greater role than *synchronous gestures* in object movement, we also decided to implement a non-spatial *menu*-based technique to compare the two types of approaches.

PHASE 2: INTERACTION TECHNIQUES & PROTOTYPE

For phase 2 of our study, we integrated the *edge bubbles*, *radar view*, and *menu* techniques from phase 1 in a prototype application for multi-tablet sensemaking. The application prototype enabled users to use tablets to search a database with a few hundred text documents for keywords and to read the documents found. Users could highlight parts of the document in different colors, annotate documents, and copy relevant parts of a document into a summary document. For the experiment, we focused on three cross-device operations between a local tablet and a remote tablet: 1.) duplicating the current view of the document on the local tablet onto a remote tablet, 2.) selecting a piece of text from the current document on the local tablet and copying it to a remote tablet, and 3.) selecting an object on the local tablet and moving it to a remote tablet. To enable a fair comparison, we ensured that these operations had equivalent functionality, so that all three were possible with all three interaction techniques. Based on the observations in [8], we also assigned a unique color to each tablet. This color was always visible on the edges of the screen to facilitate identification and selection.

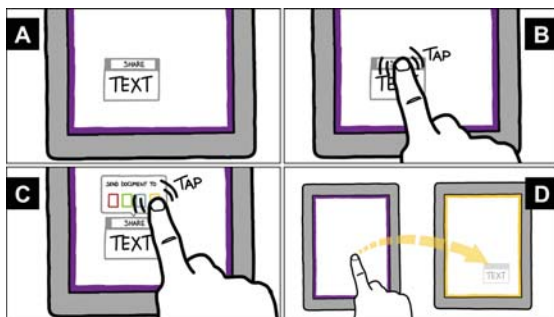


Figure 3. Menu cross-device interaction.

Interaction Technique 1: Menu

In our experiment, the spatially-agnostic *menu* technique represented the many suggestions of traditional GUI techniques from phase 1. These were particularly popular in Group C and have also been used in recent publications [8]. First, the object to move or copy must be identified on the local tablet, and its “Share” button must be pressed (Figure 3A+B). This opens a context menu for selection of the destination tablet from a horizontal list of rectangles representing the remote tablets 1-4 by their color (Figure 3C). Please note that they are ordered by an internal ID number and *not* by their spatial location, since their locations are unknown in a spatially-agnostic technique. By tapping one of the rectangles, a remote tablet is selected as the target, and the object is moved or copied there (Figure 3D). Figure 3 shows an example of moving an object. To

copy text or duplicate a view, the necessary interactions are almost the same. In the first case, a “Copy” button appears next to the currently selected text; for duplication of the view, there is a “Share” button that is permanently shown in the bottom right-hand corner of the screen.

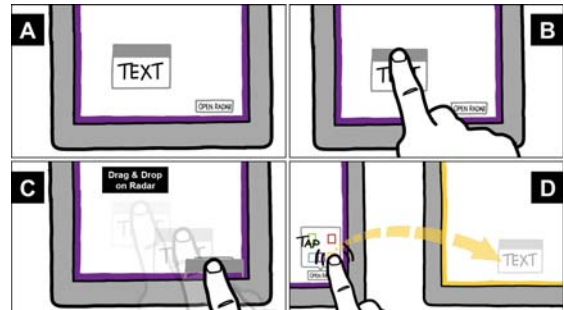


Figure 4. Radar view cross-device interaction.

Interaction Technique 2: Radar View

Following the suggestions from phase 1 (particularly from Group C for task 2), we included the *radar view*, a spatially-aware technique similar to [3,27,36] that displays a top-down map instead of just a list. The map shows color-coded rectangles as visual proxies for all devices at their current real-world locations from an overhead perspective and is updated in real-time when devices are moved (Figure 4C). To open the map and select a destination device, the text or object to move or copy is dragged and dropped onto the “Open Radar” button in the bottom right-hand corner of the screen. Tapping on one of the colored rectangles in the map then selects the corresponding remote tablet as the destination device and closes the *radar view*. To duplicate the current view, users need only press the “Open Radar” button and select a tablet without dragging text or an object onto it.

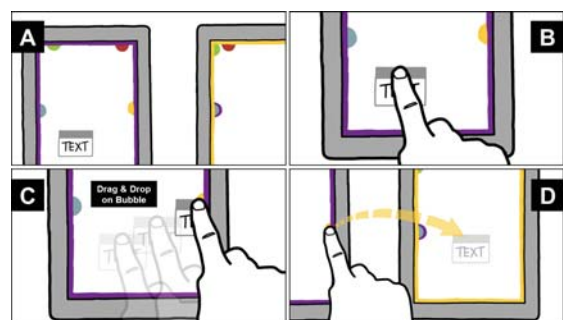


Figure 5. Edge bubbles cross-device interaction.

Interaction Technique 3: Edge Bubbles

Edge bubbles is another spatially-aware technique mainly based on suggestions made by Group A for tasks 1, 2, 5, 6, 8, and 9. Colored semi-circles around the edges of the screen serve as visual proxies for remote devices; similar to off-screen visualization techniques [2], they indicate direction in which the remote devices are located (Figure 5A). The distance to a device is mapped to the radius of its bubble. The locations of the bubbles are defined by imaginary lines between the center point of the local device

and the center points of the target devices in the real world. Each bubble is located where this imaginary line intersects with the edges of the local screen. The positions of the bubbles are updated in real-time and thus always reflect changes in the physical configuration of devices. Dragging and dropping an object onto one of the edge bubbles moves the object to the corresponding target device. Tapping an edge bubble duplicates the current view from the local device on the remote device or copies selected text to it.

PHASE 2: COMPARATIVE EVALUATION

Our comparative evaluation of *menu*, *radar view*, and *edge bubbles* was inspired by Nielsen et al., who suggest the evaluation of user-defined gestures in experiments to test them for cognitive and ergonomic quality [24]. Consequently, we tested these interaction techniques in order to 1.) learn more about ergonomic aspects (e.g., memory and stress) and 2.) to better understand the benefit of spatially-aware visual proxies of devices in comparison to spatially-agnostic menus. To this end, we designed the study as a controlled laboratory experiment with a within-subjects design and an independent variable *interaction technique* with three within-subjects factors: *menu*, *radar view*, and *edge bubbles*. The order of the three aforementioned tasks *Duplicate View*, *Copy Text*, and *Move Object* was kept constant, but the interaction techniques were systematically counterbalanced for each task using a balanced Latin Square. The dependent variables were the task completion time and subjective measures (ranking of techniques by preference, how much users liked a technique, mental demand, effort, and frustration level). Additionally, a questionnaire with two open-ended questions asked for ideas to improve the interaction techniques and suggestions for other cross-device interaction techniques. In order to achieve a high degree of external validity, we chose realistic tasks as part of a typical sensemaking procedure; as described above, the study prototype was a fully functional sensemaking application. To achieve higher internal validity, each task was repeated 48 times per condition.

Participants

12 participants (7 female, 5 male) were recruited to take part in the experiment. The mean age was 24.3 years ($SD = 2.5$, $min = 20$ years, $max = 28$ years). 11 participants were right-handed and 1 participant was left-handed. None of the participants had color vision deficiency and thus there were no problems with the color-coding employed. We only chose participants without a background in a computer science-related field. 8 participants were students from non-technical subjects such as economics or law, 2 were research assistants (in political science and physics), 1 was a kindergarten teacher, and 1 was an occupational therapist.

Apparatus

Figure 1 shows the physical setup of the experiment. As a working surface, we used a conventional office desk

(1.2×0.8m). Five Apple iPads (9.7" diagonal) were provided as tablet devices; as illustrated in Figure 1, these were situated in a U-shaped starting configuration that could be altered by participants in course of the study. The remote tablets to the left and right of the local tablet were comfortably within an arm's length and were therefore considered to be "in reach" according to Table 1. The other two tablets could only be reached by leaning forward and reaching out to them. The symmetric layout was chosen to account for the different handedness of participants. Each remote tablet was labelled with a number (Figure 1).

To track the positions of the tablets for the spatially-aware interaction techniques *radar view* and *edge bubbles*, we used the HuddleLamp vision tracking of [26]. This system tracks devices with sub-centimeter precision at a rate of 25 frames per second. A Creative Senz3D RGB-D camera was set up at a height of 78cm, which provided a tracking region of 102cm×57cm. We preferred HuddleLamp over our lab's motion capturing system because it allowed us to avoid augmenting devices with passive markers that might distract participants during the tasks. In addition, we used actual tracking data to update the *radar view* and *edge bubbles* in real-time in order to expose users to the limited accuracy, reliability, and noise in real-world tracking settings.

The application running on each tablet was implemented in HTML5/JavaScript for Safari Mobile. Tablets were wirelessly connected to the tracking system to continuously receive location and orientation data for all tablets from a Web socket connection.

Task Design

The study consisted of the three tasks *Duplicate View*, *Copy Text*, and *Move Object*, and each task consisted of three conditions: *menu* (M), *radar view* (RV), and *edge bubbles* (EB). For each trial, users were prompted with a number between 1 and 4 to indicate the destination device. The trial was noted as successful when the target device was correct; otherwise, an error was noted.

In each condition, participants repeated the cross-device interaction 48 times (each remote tablet was the target device 12 times). The sequence of numbers was randomized to avoid learning effects. Participants were asked to perform the cross-device interaction quickly and without errors. In total, there were 12 participants × 3 tasks × 3 interaction techniques × 48 repetitions = 5184 trials, with 432 trials per participant. *Duplicate View* and *Copy Text* always used the center tablet as the source device and the different tablets 1-4 as destinations. It was also necessary to confirm the end of each trial by closing the duplicated view or deleting the copied text on the remote device. The *Move Object* task began with the center tablet as the source device and one of the tablets 1-4 as the destination device. The destination tablet was then used as the source device in the next trial, and so forth.

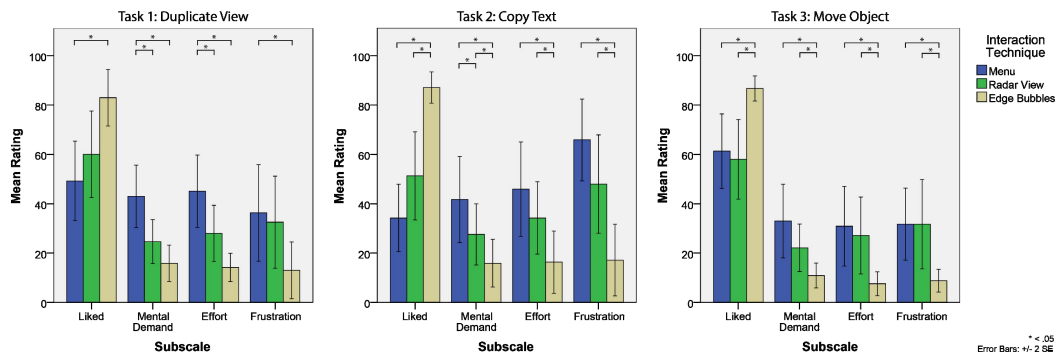


Figure 6. Subjective ratings *Liked*, *Mental demand*, *Effort*, *Frustration* for each task.

Procedure

After signing a consent form and filling out a demographic questionnaire, the participants were introduced to the first task in their assigned first condition. We did not include a training phase due to the simplicity of the task and the many repetitions. After participants had completed the task for the first condition, they were introduced to the next condition until all 3 conditions of this task were completed. After the task, participants were asked to rank the three interaction techniques in order from most favored to least favored. They also rated each interaction technique in a questionnaire with four subscales: *Liked* (scale from 0 to 100; 0: did not like it, 100: liked it), *Mental demand*, *Effort*, and *Frustration* (all subscales from NASA TLX [9]; 0: low, 100: high). Two open-ended questions at the end of the questionnaire asked for the reason for the ranking and for possible improvements to any of the interaction techniques. This procedure was repeated for each of the three tasks. Each session lasted about 1.5 hours, and participants were compensated for their time with €12.

PHASE 2: RESULTS & DISCUSSION

In the data analysis, Kendall's W coefficient (exact method) was used for the ranking of interaction techniques. The analysis of task completion time was conducted using repeated-measures ANOVAs with post-hoc pairwise comparisons (all Greenhouse-Geisser corrected). The subjective ratings were analyzed with Friedman's Tests and Wilcoxon Signed Rank Test was used for post-hoc comparisons. All post-hoc tests were Bonferroni corrected. Figure 6 shows the subscales *Liked*, *Mental demand*, *Effort*, and *Frustration* for tasks 1-3.

The ranking for tasks 1-3 was significant with a Kendall's $W_{T1} = .361$ ($\chi^2(2) = 8.67, p < .05$), $W_{T2} = .72$ ($\chi^2(2) = 17.17, p < .001$), and $W_{T3} = .65$ ($\chi^2(2) = 15.50, p < .001$). For all tasks, cross-device interactions were ranked in the following order: M_{EB} ($T_1: 1.3, T_2: 1.1, T_3: 1.1$), M_{RV} ($T_1: 2.2, T_2: 2.2, T_3: 2.3$), and M_M ($T_1: 2.5, T_2: 2.8, T_3: 2.6$) (values from 1 "most favored" to 3 "least favored"). The statistical analysis revealed that the order of the mean rankings from most favored to least favored for each task was consistently *edge bubbles*, *radar view*, and *menu*. The spatially-aware techniques proved to be favored by users even after many

repetitions, and the popularity of spatially-aware techniques during the elicitation study in phase 1 was clearly reflected in the results of phase 2. However, as we show in the following section, it is not possible to generalize this to all spatially-aware techniques.

Spatially-aware interactions are not always better

For all tasks 1-3, an ANOVA revealed a statistically significant difference between interaction techniques in terms of task completion time ($T_1: F_{1.88,20.69} = 22.69, p < .001, partial \eta^2 = .67$, $T_2: F_{1.45,15.98} = 56.04, p < .001, partial \eta^2 = .84$, $T_3: F_{1.82,20.04} = 24.95, p < .001, partial \eta^2 = .69$). All Friedman's Tests revealed statistically significant differences between interaction techniques for each of the subjective ratings *Liked*, *Mental demand*, *Effort*, and *Frustration* for all tasks 1-3 (Figure 6).

The spatially-aware *edge bubbles* outperformed *menu* in task times for tasks 1 & 2 and consistently scored higher than the non-spatial *menu* on the *Liked* subscale for tasks 1-3. Surprisingly, (also spatially-aware) *radar view* was outperformed by *menu* in terms of task time in task 3 and never scored significantly higher than *menu* on the *Liked* subscale. The differences between the two spatially-aware techniques are also visible in the higher *Mental Demand* and higher *Frustration* for *radar view* than for *edge bubbles* in tasks 2 & 3. Moreover, for all tasks, the task times for *radar view* were significantly higher than for *edge bubbles*. It seems that, despite its popularity among users, spatial awareness alone does not lead to enhanced user performance and better usability.

A potential explanation for why *edge bubbles* is clearly superior to *radar view* in our experiment is the cognitive load of mentally mapping the virtual proxy objects on the screen to their real-world counterparts. To use the *radar view*, users must locate the destination tablet on the map. This requires mentally switching from the natural egocentric view of the environment to a top-down view. This switch can be demanding, as we all know from using street maps or floor plans. Some mobile map applications try to facilitate this by automatically rotating the map so that it matches the egocentric orientation of a user. Research shows that such automatic rotation reduces users' mental load in comparison

to the “north-up” alignment [30]. Similarly, *edge bubbles* indicates the directions and distance of other tablets without the need to mentally switch between an egocentric and a non-egocentric view. This resonates with user comments that *edge bubbles* “[...] are very intuitive because proximity and direction are clear and natural [...]”; the bubbles were considered “immediately appealing” and the “most intuitive”.

The mental load of switching seems to diminish the benefits of the *radar view* in relation to the *menu*. This becomes evident in the absence of significant differences between the techniques in terms of *Frustration* in all tasks. Moreover, while the *radar view* helps users to identify devices faster when the spatial configuration is unknown or very dynamic, the performance of *menu* improves after users have internalized the mapping of colors to tablets over time. Using the *menu* only requires a sequential scanning of a one-dimensional list of colored objects, whereas the *radar view* still requires users to mentally switch between egocentric and top-down views. This explains why *menu* was faster than *radar view* in task 3 and the lack of differences between the interactions in *Mental Demand* for task 3 and *Effort* for tasks 2 & 3.

SUMMARY OF FINDINGS FROM PHASES 1 & 2

In the following section, we summarize our results and discussions from phases 1 and 2 in four findings that can inform the design of future cross-device interaction.

1.) Phase 1 has clearly revealed that users expect cross-device interactions to be *spatially-aware* (71.1% of all suggestions). In phase 2, the spatially-aware *edge bubbles* technique outperformed other techniques in a controlled experiment and was the most favored technique, even after many repetitions. We therefore recommend *spatially-aware* interactions for future mobile cross-device interactions, in particular because low-cost technologies such as [26] can now provide the necessary sensing.

2.) However, as shown in the experiment, *spatially-aware techniques* must be designed with care. The *edge bubbles* technique succeeded because of its directness of interaction and a spatial representation that did not require mentally switching between an egocentric and a top-down view. Top-down views such as radar views and maps [30] seem to introduce a cognitive load that can entirely diminish their advantages over simple *spatially-agnostic* menus.

3.) *Spatially-agnostic interactions* such as menus were popular for cross-device object movements (20.5% of suggestions), particularly for tasks involving one or multiple remote devices as sources. Their performance can be good or even equivalent to that of maps or radar views when the number of devices is small and their spatial configurations are not changing rapidly.

4.) *Synchronous gestures* were popular (25% of suggestions) for tasks concerned with expanding views or pairing devices, or whenever users wanted to refer to the device itself or its entire screen. They seem to be inherently suitable for pairing

tasks but were only suggested in very few cases for cross-device object movement (9.1%).

LIMITATIONS & FUTURE WORK

A limitation of our study is its focus on two-dimensional device configurations on a desk. The results are therefore not generalizable to other spatial configurations such as handheld devices or see-through tangible lenses. Additional work is needed to study cross-device techniques for more complex 3D device configurations.

Moreover, the different results for *edge bubbles* and *radar view* demonstrate how difficult it can be to generalize findings for all spatially-aware techniques. Different visual representations of space (e.g., egocentric or top-down) clearly had a great impact on user performance and usability. These differences must be investigated in further research in the future.

CONCLUSION

We have presented the results of a two-phase study exploring the design space of mobile cross-device interactions. First, we described our results from a gesture elicitation study in which 71% of the elicited cross-device interactions were *spatially-aware*. We discussed how participants strongly associated cross-device tasks with interacting and thinking in space. Based on the users’ suggestions, we implemented two spatially-aware interaction techniques and one spatially-agnostic technique, comparing them in a controlled experiment. The results showed that spatially-aware techniques, when designed with care, are preferred by users and can decrease their mental demand, effort, and frustration during mobile cross-device interactions.

REFERENCES

1. Andrews, C., Endert, A., and North, C. Space to Think: Large, High-Resolution Displays for Sensemaking. *In Proc. CHI '10*, ACM Press (2010), 55–64.
2. Baudisch, P. and Rosenholtz, R. Halo: a technique for visualizing off-screen objects. *In Proc. CHI '03*, ACM Press (2003), 481–488.
3. Biehl, J.T. and Bailey, B.P. ARIS: an interface for application relocation in an interactive space. *In Proc GI '04*, Canadian Human-Computer Communications Society (2004), 107–116.
4. Chen, N., Guimbretiere, F., and Sellen, A. Designing a multi-slate reading environment to support active reading activities. *ACM Transactions on Computer-Human Interaction* 19, 3 (2012), 1–35.
5. Chen, N., Guimbretière, F., and Sellen, A. Graduate student use of a multi-slate reading system. *In Proc. CHI '13*, ACM Press (2013), 1799–1808.
6. Dachsel, R. and Buchholz, R. Natural throw and tilt interaction between mobile phones and distant displays. *In Prof. CHI EA '09*, ACM Press (2009).

7. Greenberg, S., Marquardt, N., Ballendat, T., Diaz-Marino, R., and Wang, M. Proxemic Interactions: The New Ubicomp? *interactions* 18, January (2011).
8. Hamilton, P. and Wigdor, D.J. Conductor: enabling and understanding cross-device interaction. *In Proc. CHI '14*, ACM Press (2014), 2773–2782.
9. Hart, S. and Staveland, L. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload*, North Holland Press.
10. Heydekorn, J., Frisch, M., and Dachselt, R. Evaluating a User-Elicited Gesture Set for Interactive Displays. *M&C 2011*, Oldenbourg Verlag (2011), 191–200.
11. Hinckley, K., Ramos, G., Guimbretiere, F., Baudisch, P., and Smith, M. Stitching: pen gestures that span multiple displays. *In Proc. AVI '04*, ACM Press (2004).
12. Hinckley, K. Synchronous gestures for multiple persons and computers. *In Proc. UIST '03*, ACM Press (2003).
13. Jetter, H.-C. and Reiterer, H. Self-Organizing User Interfaces: Envisioning the Future of Ubicomp UIs. *Workshop Blended Interaction (in conjunction with CHI '13)*, University of Konstanz (2013).
14. Kidd, A. The marks are on the knowledge worker. *In Proc. CHI '94*, ACM Press (1994), 186–191.
15. Kray, C., Nesbitt, D., Dawson, J., and Rohs, M. User-defined gestures for connecting mobile phones, public displays, and tabletops. *In Proc. MobileHCI '10*, ACM Press (2010), 239–248.
16. Li, M. and Kobbelt, L. Dynamic tiling display: building an interactive display surface using multiple mobile devices. *In Proc MUM '12*, ACM Press (2012).
17. Lucero, A., Holopainen, J., and Jokela, T. Pass-them-around: collaborative use of mobile phones for photo sharing. *In Proc CHI '11*, ACM Press (2011).
18. Lucero, A., Keränen, J., and Korhonen, H. Collaborative use of mobile phones for brainstorming. *In Proc. MobileHCI '10*, ACM Press (2010), 337–340.
19. Marquardt, N., Diaz-Marino, R., Boring, S., and Greenberg, S. The proximity toolkit: prototyping proxemic interactions in ubiquitous computing ecologies. *In Proc. UIST '11*, ACM Press (2011).
20. Marquardt, N., Hinckley, K., and Greenberg, S. Cross-device interaction via micro-mobility and f-formations. *In Proc. UIST '12*, ACM (2012), 13–22.
21. Morris, M.R., Danielescu, A., Drucker, S., et al. Reducing legacy bias in gesture elicitation studies. *interactions* 21, 3 (2014), 40–45.
22. Nacenta, M., Gutwin, C., Aliakseyeu, D., and Subramanian, S. There and Back Again: Cross-Display Object Movement in Multi-Display Environments. *Human-Computer Interaction* 24, 1 (2009).
23. Nacenta, M.A., Kamber, Y., Qiang, Y., and Kristensson, P.O. Memorability of pre-designed and user-defined gesture sets. *In Proc. CHI '13*, ACM Press (2013).
24. Nielsen, M., Störing, M., Moeslund, T.B., and Granum, E. A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI. In A. Camurri and G. Volpe, eds., *Gesture-Based Communication in Human-Computer Interaction*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, 409–420.
25. Pierce, J.S., Mahaney, H.E., and Abowd, G.D. *Opportunistic Annexing for Handheld Devices: Opportunities and Challenges*. Georgia Institute of Technology, 2003.
26. Rädle, R., Jetter, H., Marquardt, N., Reiterer, H., and Rogers, Y. HuddleLamp : Spatially-Aware Mobile Displays for Ad-hoc Around-the-Table Collaboration. *In Proc. ITS '14*, ACM (2014).
27. Reetz, A., Gutwin, C., Stach, T., Nacenta, M., and Subramanian, S. Superflick: a natural and efficient technique for long-distance object placement on digital tables. *In Proc. GI '06*, Canadian Information Processing Society (2006), 163–170.
28. Rekimoto, J. and Saitoh, M. Augmented surfaces: a spatially continuous work space for hybrid computing environments. *In Proc. CHI '99*, ACM Press (1999).
29. Rekimoto, J. SyncTap: synchronous user operation for spontaneous network connection. *Personal and Ubiquitous Computing* 8, 2 (2004), 126–134.
30. Seager, W. and Fraser, D.S. Comparing physical, automatic and manual map rotation for pedestrian navigation. *In Proc. CHI '07*, ACM Press (2007).
31. Seyed, T., Burns, C., Costa Sousa, M., Maurer, F., and Tang, A. Eliciting usable gestures for multi-display environments. *In Proc. ITS '12*, ACM Press (2012).
32. Valdes, C., Eastman, D., Grote, C., et al. Exploring the design space of gestural interaction with active tokens through user-defined gestures. *In Proc. CHI '14*, ACM Press (2014), 4107–4116.
33. Volda, S., Podlaseck, M., Kjeldsen, R., and Pinhanez, C. A study on the manipulation of 2D objects in a projector/camera-based augmented reality environment. *In Proc. CHI '05*, ACM Press (2005).
34. Weigel, M., Mehta, V., and Steimle, J. More than touch: understanding how people use skin as an input surface for mobile computing. *In Proc. CHI '14*, ACM Press.
35. Weiser, M. The Computer for the 21st Century. *Scientific American* 265, 3 (1991), 94–104.
36. Wigdor, D., Shen, C., Forlines, C., and Balakrishnan, R. Table-centric interactive spaces for real-time collaboration. *In Proc. AVI '06*, ACM Press (2006).
37. Wobbrock, J.O., Morris, M.R., and Wilson, A.D. User-defined gestures for surface computing. *In Proc. CHI '09*, ACM Press (2009), 1083–1092.