Supporting Visual Exploration for Multiple Users in Large Display Environments

Sriram Karthik Badam[*]
University of Maryland
College Park, MD, USA

Fereshteh Amini[†]
University of Manitoba
Winnipeg, MB, Canada

Niklas Elmqvist[‡]
University of Maryland
College Park, MD, USA

Pourang Irani[§]
University of Manitoba
Winnipeg, MB, Canada

Figure 1: Two different types of operations performed using our multi-user lens technique for exploring multiscale visualizations. Selecting a region of interest (ROI) using a mid-air gesture (left), vs. merging two time-series plots when two users approach one another (right).

ABSTRACT

We present a design space exploration of interaction techniques for supporting multiple collaborators exploring data on a shared large display. Our proposed solution is based on users controlling individual lenses using both explicit gestures as well as proxemics: the spatial relations between people and physical artifacts such as their distance, orientation, and movement. We discuss different design considerations for implicit and explicit interactions through the lens, and evaluate the user experience to find a balance between the implicit and explicit interaction styles. Our findings indicate that users favor implicit interaction through proxemics for navigation and collaboration, but prefer using explicit mid-air gestures to perform actions that are perceived to be direct, such as terminating a lens composition. Based on these results, we propose a hybrid technique utilizing both proxemics and mid-air gestures, along with examples applying this technique to other datasets. Finally, we performed a usability evaluation of the hybrid technique and observed user performance improvements in the presence of both implicit and explicit interaction styles.

Keywords: Proxemics, gestures, visual exploration, collaborative sensemaking, user study, large displays, orientation, position.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Interaction styles; I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction techniques

1 INTRODUCTION

Collaborative sensemaking is one of the most promising applications [6] for so-called *smart environments* [40], where computing is woven into the physical space in the form of sensors, devices, and displays. Traditional mouse and keyboard interaction is unsuitable for such environments, due to the lack of fixed horizontal surfaces as well as the presence of multiple collaborators in the same space. Therefore, gestural interaction is often employed for controlling smart environments. However, gestures can be imprecise, difficult to learn, and may cause fatigue [14]. A potentially more natural style of interaction is based on *proxemics* [11]—the study of spatial relationships between people and artifacts—and was suggested by Jakobsen et al. [17] to control visualizations on large displays. However, their approach is not intended for multiple concurrent users in the same physical space, a situation that arguably is one of the most suitable (and common) for such environments.

In this paper, we explore the design space of multi-user visual exploration in large display environments in terms of both presentation (visual output) and interaction. For presentation, we propose a multi-view representation based on focus+context lenses to support the whole spectrum of collaborative coupling [37], from tight (lenses combined to support working together) to loose (lenses separate to support working independently). For interaction, we explore the balance between *explicit* input—where actions are used solely to interact with the system, such as gestural commands—and *implicit* input—where actions are proxemic, i.e. drawn from spatial and social cues to automatically trigger changes in the interface [19]. Towards this end, we design and evaluate our proposed interaction and presentation techniques using a formative evaluation with 12 participants—which influenced our design choices—as well as a summative evaluation with 18 additional participants.

We claim the following contributions: (1) several lens-based interaction and presentation techniques for multi-user sensemaking in large display environments, including both proxemic (implicit) and gestural (explicit) interaction styles; (2) results from a formative evaluation studying tradeoffs between using proxemic and gestural interaction; and (3) results from a summative evaluation on a hybrid technique combining both proxemics and gestures for pairs of analysts performing sensemaking collaboratively. During our design process, we also derive feedback mechanisms for the interactions.

2 BACKGROUND

There has been a recent surge in utilizing proxemics and gestural interaction in visual analytics and HCI. In this section, we present the related work on proxemics and gestures, general implicit and explicit interaction, and sensemaking in smart environments.

---

[*] e-mail: sbadam@umd.edu
[†] e-mail: amini@cs.umanitoba.ca
[‡] e-mail: elm@umd.edu
[§] e-mail: irani@cs.umanitoba.ca

2.1 Proxemics and Gestures

To achieve Mark Weiser's vision of ubiquitous computing [40], devices such as public displays and hand-held computers present in an ecosystem should be harnessed to create a context-aware system that captures knowledge about surrounding users and devices (actors). In this context, proxemics has been leveraged as a way to create connections between users and devices in an environment [10, 17, 23]. Greenberg et al. [10] discuss the use of proxemic attributes such as distance, orientation, identity, movement, and location, for building a structured implicit interaction model. This includes (1) using distance-based interaction zones that define the reaction of the display [39], and (2) orientation-based understanding of the user attention [33]. Peck et al. [29] used distance-based interaction zones to control the scale of interaction and found evidence of naturalness in physical navigation in front of a large display as the users then tend to associate visual scale (e.g., seeing overview/detail based on distance) with the interaction scale.

Beyond theories on leveraging proxemics, several experiments have found significant effects of user interaction on the proximity and orientation among users and devices during collaboration. Jakobsen and Hornbæk [16] observed that user activity clearly influences their physical navigation as they moved closer for sharing and collaborative analysis, while staying away for performing separate tasks. Jakobsen et al. [17] evaluated multiple design choices for mapping proxemics categories to single-user actions performed on a visualization such as selection, filtering, and navigation, by comparing the user experience against a baseline interaction with a gyroscopic mouse. They found that navigation over a map through movement, as well as view update and selection operations based on distance, came naturally to the users, while operations such as scaling and positioning based on location were not. Dostal et al. [7] developed the SpiderEyes toolkit for attention-aware and proximity-aware interfaces on wall-sized displays. However, the visualization scenarios presented in their paper are restricted to splitting the large-display space between users in a collaboration, and they do not fully support changing collaboration styles, where we often see both tightly and loosely coupled work [25, 38]. More recently, BodyLenses [21] introduced magic lenses for wall displays that are controlled by body interactions. This work details the design dimensions for lens-based presentation and interactions including implicit distance and movement mapping to change lens position and properties, as well as using explicit gestures for adjusting lens shapes. Our goal, however, is to evaluate the implict and explicit interaction styles for specific lens operations to build a hybrid interaction style that better supports parallel individual and collaborative work during visual analytics in large displays environments.

Recent innovations in capturing stereoscopic information using depth cameras has led to reliable use of gestural interaction in human-computer interaction [27]. Gestures can be either predefined or user-defined, and they can only be triggered by explicit motion that deviates from regular user interaction in an environment. From an gesture design perspective, Nancel et al. [27] studied the design of mid-air pan-and-zoom movements, including uni-bimanual interaction, linear vs. circular movements, and guidance for mid-air gestures to interact with the large displays from a distance. They found that linear gestures, involving a linear movement and clutching, were faster than circular gestures, and two-handed gestures were faster than one-handed gestures. Linear gestures also received better user feedback.

2.2 Implicit and Explicit Interaction in HCI

Our interaction techniques in this paper utilize both implicit and explicit actions performed by users in a collaborative setting. In this context, Ballendat et al. [2] describe an example of an interactive media player, which uses proxemics information to change the state of the media player as an implicit action, and also allow users to explicitly trigger events using direct touch. While implicit interaction conforms very well with state-based HCI systems such as a media player, their use in interactive visualization is—to the best of our knowledge—underexplored and limited [7, 17, 29].

In contrast, there has been significant amount of research into explicit interaction models using mouse, direct touch, direct manipulation, and also mid-air gestures. Malik et al. [22] present design choices for selection and transformation of views on a large display using single or multiple finger postures and gestures captured by a touchpad. Beaudouin-Lafon et al. [3] created different frameworks for managing gestural input and display modules in collaborative environments. They target explicit gestures using real and virtual instruments connected to graphical objects in a visualization. Shoemaker et al. [35] developed a body-centric model towards explicit interaction, by utilizing one's own body parts as a container and a control surface for interactions (e.g., selecting an interface object by touching one's hip). They used a virtual shadow to provide embodiment and a form of feedback for the body mappings. Beyond these styles, using mobile devices as a surrogate for interaction is common [8, 32] in mixed-modal environments, which have more than a single display and input device.

2.3 Large Displays & Multiple devices for Sensemaking

Large displays offer more space to think [1] and promise better collaboration between analysts [25]. Futher utilization of portable devices in such environments allows for spatialization of the sensemaking process and for supporting complex collaboration scenarios [25, 38]. However, as Chung et al. [5] mentioned, exploiting the physical space and embodiment, and supporting collaborative use, sharing, and organization mechanisms across large displays and other devices becomes important in these scenarios.

Isenberg et al.'s hybrid-image visualization [15] approach introduced distance-based visual encodings that graphically blend overview and detail views into a single representation on a large high resolution display. This encoding does not require interactivity as details become naturally perceivable within context at a close distance. These hybrid representations by nature promote movement without requiring any tracking mechanisms and are ideal for public spaces for progressive revealing of content. Together with interaction models designed for multi-user visual exploration on large displays (such as the ones evaluated in this paper), hybrid-image visualization can be an effective approach for supporting complex collaborative sensemaking scenarios [25, 38] with a reduced overhead on the system.

Finally, the VisPorter system [5] supports sensemaking across devices through the ability to share and integrate knowledge gained on private and public displays. They developed explicit gestures such as flicking (by touch) towards a device for sharing, while implicitly tracking the position information. However, designing these interactions for even larger displays (in terms of physical size) that afford interaction in 3D and physical navigation is a complex task. Our focus is on exploring and evaluating the design space of proxemic and mid-air gestural interactions in large display environments during sensemaking to find a balance between them.

3 SENSEMAKING IN MULTI-USER DISPLAY ENVIRONMENTS

Large displays support multiple concurrent users and have been shown to enhance productivity in co-located collaboration [6]. For complex sensemaking tasks, large displays and multi-monitor setups provide a large visual space to interact and organize information, and by extension, to think [1, 5, 10]. Due to being distributed in physical space, interaction models for these environments go beyond standard input devices—such as mouse and keyboard—and leverage the benefits of gestural and even full-body interaction. Unlike traditional input modalities, these new interaction models support both explicit and implicit modes of interaction:

- **Explicit interaction:** User action where the purpose **is** primarily to interact with a computer system.
- **Implicit interaction:** User action where the purpose is **not** primarily to interact with a computer system [19].

Explicit interaction is the traditional mode of interacting with computers, and includes actions such as mouse clicks and drags, keyboard presses, and touchscreen taps and swipes. Implicit interaction is a more novel approach, and it targets automatically using the body states and movements that are known or observed to exist when users interact in a physical space [20] to avoid learning explicit actions and reduce additional physical activity. Therefore, implicit interaction can be tightly coupled with utilizing proxemics attributes [11] of the users such as their position, posture, movement, orientation, and identity of users within a physical space to control a computer system [19, 21, 39]—also referred to as *proxemic interaction* [10]. The design space for proxemic interaction with visualizations—mapping proxemics dimensions to high-level tasks [13]—has been presented by Jakobsen et al. [17]. However, these designs do not fully extend to parallel individual and collaborative work due to inherent presentation conflicts. It is worth noting that implicit actions may go further than proxemics (e.g., using facial expressions). But we use these terms together since we focus on proxemic interactions that are treated as implicit (seen in previous work [10, 19, 21]) since they resemble common physical actions.

Explicit interaction through mid-air gestures [31, 41] also goes beyond traditional input methods. The design space for gestural interaction with large displays has been studied before [27]. However, it still depends on a vocabulary of *explicit actions* that have minimal or no meaning when performed outside the multi-device ecosystem. For this reason, gestural interaction is often not easily discoverable, requires significant training, and may need additional physical effort [14]. For example, direct-touch gestures such as tap and pinch, and mid-air gestures for zoom and pan used in recent works [27], convey the user intention quickly and unambiguously to the system, but they require the users to perform additional physical work to trigger the outcome on the visualization. Furthermore, the gesture vocabulary is also easily exhausted in the presence of a large number of interactions, which occurs commonly for realistic analytics tasks. On the other hand, explicit actions may be meaningful for operations where user confirmation is required.

We propose a combined presentation and interaction technique for multi-user visual exploration in large display environments (Figure 1). Drawing on design guidelines by Tang et al. [37] and Kister et al. [21], the **presentation** technique uses focus+context lenses owned by each collaborator. The lenses act as views for the users during parallel individual or loosely coupled work [21, 38], and can also be combined with other lenses for tightly coupled collaborative analytics. We refer to Kister et al.'s design space exploration [21] for lens design including placement, size, shape, and rendering. We focus on the **interaction** techniques—both implicit and explicit design alternatives (Figure 2)—for exploring data visualizations through the lenses. Here, we list the abstract lens operations and elaborate on possible options for their implicit and explicit designs followed by our choice for the preliminary implementation.

Initiate: Lenses visualize specific parts of a dataset, and the lens initiate operation involves selection of a region/part of an overview visualization. Selection operations in large display and mixed-modal environments are typically done through pointing and touch with hand [23] and other devices [32] in an explicit way. However, in contrast to selection of discrete objects (for instance, photos), selections in a visualization are more precise and granular. For selection, we define a gaze-controlled cursor highlighting the region-of-interest (similar to the cursor used by Peck et al. [29]).

- *Implicit Initiate:* When the user's gaze dwells on a region in the overview visualization, a lens is created at that location.
- *Explicit Initiate:* User creates a lens by navigating the cursor using hand and confirms the selection with a hand roll.

Scale: The ability to modify the size of a lens is needed in a large display environment since the distance from the screen affects the user's view. View scaling was previously performed by tracking the user's position and distance from the screen [12, 29], and by direct manipulation, popularly seen in movies such as Iron Man (2008) and Minority Report (2002).

- *Implicit Scale:* The size of the lens is directly mapped to the distance of the user from the display.
- *Explicit Scale:* The lens size is changed by hand movement pulling/pushing away the lens.

Move: Positioning the lens helps organize the workspace along with the scale operation. Move operations have been commonly done on large displays and across devices through explicit gestures such as flicking, drag-and-drop, grab-and-move [23, 24, 30].

- *Implicit Move:* The user's gaze or movement in the physical space controls the lens.
- *Explicit Move:* The user's hand directly moves the lens.

Zoom and Filter: Zooming and filtering the data within a view is a very common task in visualization [34]. These operations have been previously done through implicit actions such as leaning [12] and distance-based semantic zoom [17], and explicit actions such as pinch-to-zoom, linear, and circular gestures [27].

- *Implicit Zoom:* The user's distance and orientation is used to zoom and filter the lens content.
- *Explicit Zoom:* The lens content is zoomed using a mid-air hand zoom gesture—linear hand push/pull [27]—while standing still in front of the wall display.

Pan: This operation shifts the content of a lens (i.e., not its position) based on the data attributes. For example, this can be based on time for time-series data or based on spatial location for spatial data. Pan operations in large display spaces have been explored through movement of handheld devices in a two-dimensional space [28] and through mid-air gestures [27], which are both explicit.

- *Implicit Pan:* The lens is panned based on the orientation of the user's head when dwelling on it.
- *Explicit Pan:* Hand movement (left and right)—hand swipe—is mapped to the pan operation on the lens content [27], when the user is standing still in front of the wall display.

Merge: Merging content is helpful for collaborative analysis, cross validation, and trend analysis across lenses. Merge also corresponds to the *relate* visualization task connecting different sets of visualizations [34]. Various styles of composite visualizations have been surveyed by Javed et al. [18], and these are design choices for merge in terms of visual representation. Recent work on shape-shifting wall displays [36] for individual and group activities explores the implicit and explicit controls for managing the arrangement of individual displays. They merge the individual displays implicitly by observing when users move close to them.

- *Implicit Merge:* Lenses are merged when two users are close and aligned towards each other [36].
- *Explicit Merge:* Lenses are merged when the users position them close to each other with their hands.

Split: Lens splitting can be mapped to a similar implicit action or explicit gesture as merge. Takashima et al. [36] supported this by identifying when users walk away or through a split gesture.

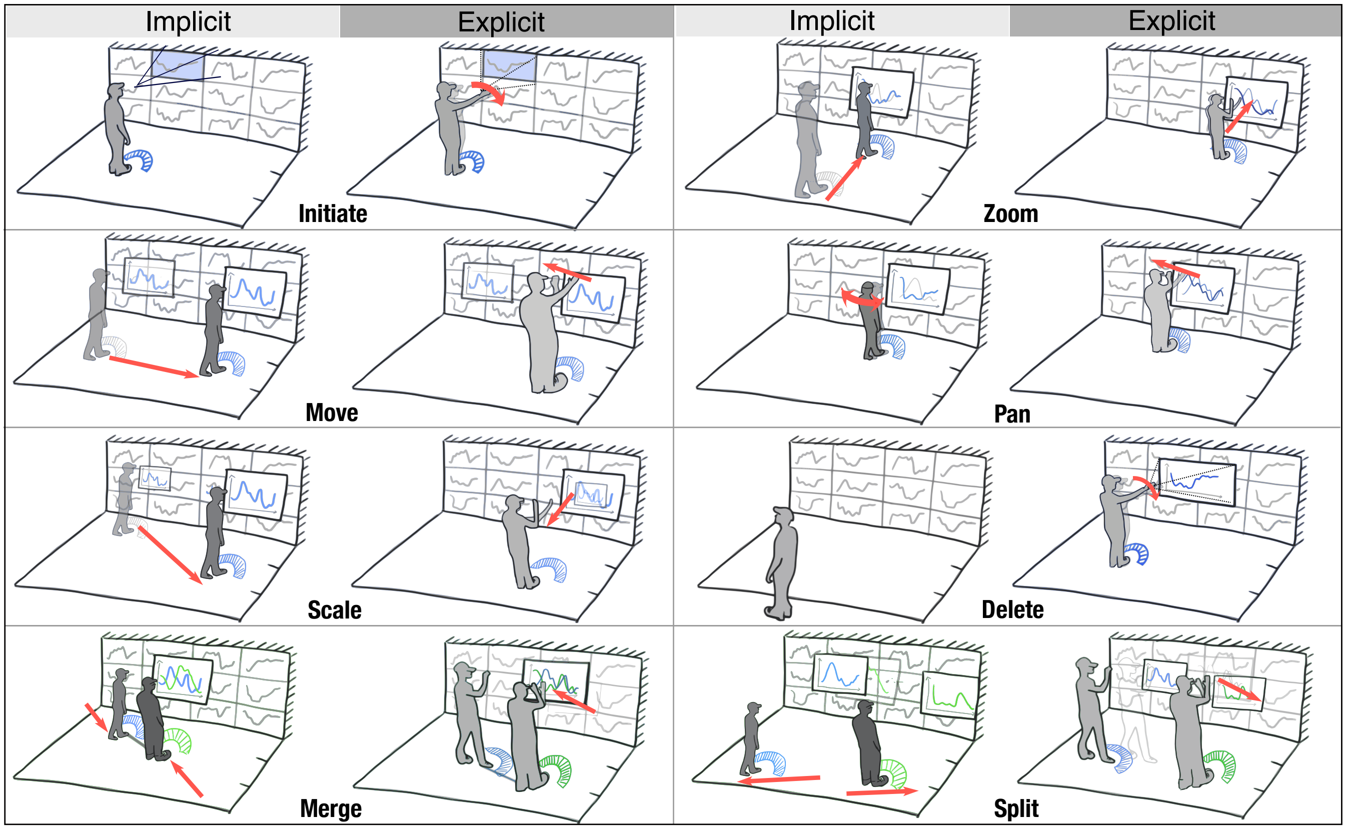


Figure 2: Visual summary of using proxemics and gestures to interact with focus+context lenses on a large multi-user display space.

- *Implicit Split:* Lenses are split when the users move away while facing away from each other.
- *Explicit Split:* Lenses are split when the users separate the lenses by pointing and dragging with their hands.

Delete: Being the inverse operation of initiation, it can be triggered based on similar metrics as initiation. Tracking attention is a popular way of understanding if a view is of interest [24].

- *Implicit Delete:* Lenses are deleted when the user moves out of her workspace while facing away from the lens.
- *Explicit Delete:* Lenses are deleted using a hand roll gesture while pointing to the lens.

4 IMPLEMENTATION

Our multi-user lens techniques involve multiple design choices for mapping proxemics and gestures to operations on a visualization. In this section, we give details about the hardware platforms, system architecture, and detection of implicit and explicit actions.

4.1 Hardware Platform

We used a semi-CAVE environment[1], including large wall and floor displays to develop the proposed multi-user lens technique. Both the wall and floor displays measure $160'' \times 90''$ with HD (1920×1080) resolution.[2] The displays were attached to a computer running Microsoft Windows 7 Vicon motion capture system with ten high-resolution cameras to track user proxemics and gestural input. Tracking is performed on the reflective infrared markers attached to the props worn by the participants: gloves for tracking the hand, slippers for tracking the foot, and baseball caps for tracking the user's head.

4.2 System Architecture

We opted for a web application environment to support multiple platforms (both desktop and mobile for future implementations). The system followed a client-server architecture with the client side using HTML, JavaScript, and CSS. The client manages the visualizations built using the D3 toolkit [4]. The server side of the application contains modules for handling connection and input from the Vicon system through the Vicon data stream SDK. It also includes several methods to process the Vicon input to detect actions, which are predefined interpretations of proxemic variables and gestures as discussed in the Design section. The communication between server and client sides of the application are handled through persistent connections from the clients. Once an action (e.g., a change in proxemic attribute value or a new gesture) is detected by the server, it is immediately pushed to the connected clients. This architecture, therefore, supports multiple clients at the same time.

4.3 Detecting Actions and Providing Feedback

Identifying proxemic interactions. Events based on proxemic actionsare calculated by taking input from the Vicon tracking system[3]. This input corresponds to the 3D position and orientation of each object (cap and slipper), which are further translated to get the distance and orientation with respect to the displays. These proxemic readings are then processed to identify conditions for various

---

[1] Visbox: `http://www.visbox.com/viscube-C2-HD.html`

[2] Note that this is a relatively low resolution compared to tiled display walls—where each display is HD—which are becoming common. However, this has no practical bearing on our techniques or evaluation results.

[3] Tracker: `http://www.vicon.com/Software/Tracker`

actions. For example, a gaze action would be triggered if the position and orientation of the users facing the display does not change for a pre-defined time period—three seconds in our implementation. For lens scale and move, the position of the moving users facing the display is directly mapped to the size/position of the lens. The lens scales uniformly based on the distance from the display—1× when close to the wall to 2× at the edge of the floor (similar for zoom). Distance between users is used to trigger lens merge when the users are close to each other (distance $\leq 0.75m$).

Identifying gestures. Gestural events are detected for a user's hand by tracking changes in the degrees of freedom of the glove such as roll, yaw, pitch, and position values from the data provided by the tracker. Each gesture detection procedure involved identifying the state of each object and the direction of movement within a space/time threshold to avoid false positives. For example, the hand gesture for lens initiation is detected when the user's hand (glove object) is horizontal pointing towards the display and has clockwise 50° change in the roll value within one second. Similarly, a pull gesture for lens zoom is detected when the user's hand is facing the wall display, and moves in the opposite direction towards the user by at least $0.2m$ in less than one second (mapped to 1.2× zoom). The swipe gesture working with similar thresholds but with a different direction of movement.

Providing feedback and visual cues. We use the floor display to provide feedback for actions. A halo is shown around the user on the floor to give feedback of the orientation. For lens merge, the two users are connected with a line on the floor display signifying a bond between them during group activity. When the users are very close, this feedback line becomes more opaque/solid compared to when they are far away. Once they move away from one another (or explicitly split), the line disappears. Finally, the floor space is divided into **interaction zones** [17]—close ($< 0.75m$), middle ($0.75 - 1.5m$), far ($> 1.5m$)—and visual cues are drawn for these regions. Finally, the participants are given feedback of the gestures identified using textual labels on the wall display.

Avoiding issues. The interaction zones on the floor are associated to specific actions to reduce conflicts between similar actions. For example, a lens zoom is triggered when the user is close and a lens scale is triggered if otherwise. To avoid constant visual updates when the users are stationary, the spatial attributes of the objects are rounded to closest $0.1m$ resolution for position and direction (orientation). Occlusion is avoided in this system by using multiple Vicon trackers around the room ceiling making sure that the users are in the field of view of at least two cameras at any time. Tracking information is rarely lost in this system; however, when lost the user is aware of this through the feedback mechanisms. Due to the high-precision nature of Vicon and ability to provide readings at 50fps, the latency of action detection is based on the time thresholds. These action definitions are verified through pilot studies.

5 FORMATIVE EVALUATION: IMPLICIT VS. EXPLICIT

To gain a better understanding of the usefulness of both implicit and explicit interaction styles, we conducted a formative user study on the physical affordances of our design ideas. The user study focused on collaborative visual analysis of multivariate time-series data, and required users to interact with the data through lenses to figure out patterns within and across variables. The dataset was sensor data measured by different types of sensors (e.g., moisture, temperature, and humidity) in a building over time. It contained 13,500 records for 30 sensors spanning over two weeks. The user interaction was not controlled or constrained, thus allowing the participants to freely interact and explore various features provided by the lens technique. Our focus through this qualitative inquiry is not only to observe which interactions style suits each lens operation but also to gain interesting and unexpected design opportunities that can expand the conceptual model of our lens.

5.1 Participants

We recruited 12 participants (5 female, 7 male) from our university's student population (6 groups of two). Participants signed up voluntarily and were rewarded $10 upon successful completion. All participants had experience with data analysis using visualization and more than 6 years of experience using computer systems. Only 7 participants had experience with mid-air gestural interactions.

5.2 Experimental Design and Procedure

The experiment followed a 2×6 within-participant design with Interaction Mode I (implicit, explicit) and Analytics Tasks A, resulting in 72 total trials (i.e., 12 per team). The participants were scheduled for one-hour sessions in groups of two. They were first introduced to the equipment used in the study, followed by a brief introduction of the study goal and tasks they were expected to perform. After signing their consent, the participants were asked to wear the props on their head, dominant hand and their feet. The investigator then described a list of gestures for the explicit mode and the proxemic attributes tracked by the system in the implicit mode. The participants were allowed to practice by testing each operation in the technique until they were comfortable. The participants were then quizzed to test their knowledge of the lens operations.

5.3 Tasks

Each team was given six tasks; three were low-level comprehension tasks involving finding specific values, trends, and extrema in the visualization, whereas three were high-level synthesis tasks involving identifying anomalies, comparing data, and correlating data. The tasks were designed to stimulate engagement, collaboration, and discussion during group analysis. The synthesis tasks (compare and correlate) require exploring multiple parts of the overview and encourage collaboration through lens merge operations. These categories were inspired by the work of Shneiderman [34] that summarizes the types of tasks performed on different data types.

Two variants of these six tasks were prepared, giving two task lists (TS1 and TS2), one for each interaction mode. The tasks within each set were randomly shuffled for each group to counter sequence effects. Each task required multiple implicit or explicit actions that were mapped to corresponding lens operations. The presentation order of interaction modes was fully counterbalanced.

5.4 Data Collection

All participants were encouraged to "think aloud" and announce their actions and decisions while interacting with the system to perform the tasks. During the session, the investigator took notes about important observed events and verbal comments made by the participants. The sessions were also video recorded (with consent).

After finishing the tasks with one mode of interaction, the users filled a five-point Likert scale questionnaire collecting their opinion about seven different factors (described in Results section). The sessions ended with a short interview as a way of triangulation.

5.5 Results

Inspired by grounded theory [9], we analyzed the participant data by open-coding the notes and transcribed interviews. Two researchers, who had observed and conducted sessions, developed two initial code-sets independently, and after reaching an agreement on a final code-set, one researcher proceeded to code the remaining session data. In what follows, we present the emerged themes, along with the corresponding participant group (G) identifications.

5.5.1 Implicit actions

Lens initiation was not implicit enough. While participants liked the idea of a smart environment that could guess when a lens needs to be initiated, they unanimously agreed that head dwell was not much efficient, as it led to false negatives (low discoverability) and false positives. Two participants said, "*Being interested in one chart does not mean I will only look at that one chart*" (G4), and in fact the chart of interest is usually decided by shifting focus between different charts in the overview. When participants knew the chart of interest, they had to consciously focus their line of sight to that chart region to create a lens and this was contrary to the implicit nature of the design. Five participants mentioned that the dwell time taken into consideration by the system maybe "too long". We conclude that deciding on an accurate dwell time is a non-trivial task, and gaze dwell may not be an appropriate action for this task.

Lens move and scale were liked. Perhaps the most interesting observation about these interactions was the description given by one participant as "unnoticed interaction" (G3). This participant (G3) further said, "*the interaction was so natural and intuitive that I almost did not notice!*". Intuitiveness was a common reason given by all participants who expressed positive opinions about these interactions. These results confirm the findings of Jakobsen et al. [17], and align with our motivation behind implicit actions.

Lens zoom and delete were perceived as "fun to do". Both zoom and delete required moving body across different proxemics regions. Three participants mentioned that they like the need for being active (G1, G2, G5) referring to the physical navigation that triggers the zoom. One participant described the zoom interaction as being "fun to do" (G2). Contrary to the observations of Jakobsen et al. [17] about similar zoom interactions, we did not observe awkward, slow, or uncertain movements from our participants. We can argue that this could be due to (1) mapping the movement speed directly to the zoom rate, and (2) the visual cues on the floor in the form of lines dividing the different proxemics regions.

Lens zoom and pan were not accurate. In many cases, participants were observed to undo (or redo) zoom and pan actions because they had zoomed or panned the lens content too much or too little on their first attempt. Several comments were made during the interview about the need for more "control" during zoom and pan, and using "gestures" (G6) was a common solution suggested by the participants. One participant suggested "*having a way of turning this feature on and off*" (G3). Implicit actions for these operations lacked precision and accuracy on the first attempt, and when further attempts were made, their implicit nature is questioned.

Lens merge and split had mixed feedback. Lens merge was found to be the most interesting feature in implicit interaction mode. 75% of the participants described this feature as either "cool" (G3), "useful" (G2, G4), or "interesting" (G6). One participant remarked that "*this is similar to real world, where people have to get close to each other to share physical copies of documents*" (G6). Implicit lens split, on the other hand, received some conflicting results. While some participants still liked the feature, others expressed the desire for "keeping lenses merged" (G2, G3) while being able to "move around" (G2) referring to the need for both merged and separate work spaces during collaborative data analysis.

5.5.2 Explicit Interactions

Lens content manipulation gave users more control. A common theme observed across all explicit actions was that the participants liked having the direct control provided by these actions. One participant remarked, "*it is great to zoom in and out as much as you like*" (G6). Generally, participants learned the gestures related to lens content manipulation quickly. The participants who needed relatively more time to learn these gestures were observed to use the gestures more efficiently. In particular, it is worth highlighting a remark made by one participant about the initiate gesture (i.e., the hand roll gesture) where the gesture referred to as "natural" (G1) and "like drilling into the data" (G1). We also observed minimal false negatives for these gestures except when they were performed at a very high or very low speeds and ignored by the system. In all of these cases, the participants were able to identify the issue through the visual feedback (or in fact the lack of visual confirmation) provided for each gesture in the form of textual labels and they successfully corrected the gesture to achieve the desired results.

Lens move and scale gestures were demanding. Explicit lens move and scale did not receive good feedback from participants. Through observation, we noticed that participants usually did not put their hands down after completing a move or scale action. When asked, one participant mentioned that it was because they thought "*the lens would go back to where it was, if they let go of it*" (G4). Seven participants pointed out their hands were tired and experienced exhaustion when moving the lens. We rarely saw participants use the lens scale action. Three participants mentioned that this gesture "*required too much work for little return*" (G6).

5.5.3 Subjective Ratings

We also collected participant opinions about each action in both implicit and explicit modes through a post-session questionnaire. It included seven statements focusing on a different metric for each lens operation: Preference, Accuracy, Intuitiveness, Efficiency, Enjoyment, Collaboration, and Physical effort. Participants rated statements on a Likert-scale ranging from 1 (e.g., Not Preferred) to 5 (e.g., Preferred). For Physical Effort, the ratings ranged from 1 (Very hard to perform) to 5 (Very easy to perform). The mean scores are summarized in Figure 3. We performed Wilcoxon Signed Ranks tests at a significant level of $\alpha = .05$ to test for differences between 2 samples of participant ratings (implicit and explicit) for each statement and lens operation. For example, this meant comparing implicit and explicit lens zoom in terms of physical effort.

There was a significant difference between implicit and explicit lens initiation when it comes to perceived physical effort ($Z = -2.821, p = .005$). In fact, the majority of the metrics ranked significantly higher in the explicit mode for lens initiation.

Lens scale was ranked significantly higher in the implicit mode for both collaboration ($Z = -2.239, p = .025$) and physical effort metrics ($Z = -2.209, p = .027$). Also, the implicit lens move action was enjoyed more than explicit ($Z = -2.140, p = .032$).

Lens pan ranked significantly more intuitive ($Z = -2.360, p = .018$) when performed using explicit gestures. We did not, however, observe any other significant effects for zooming and panning.

Preference ($Z = -2.230, p = .026$) and accuracy ($Z = -2.877, p = .004$) ranked significantly higher for lens implicit merge operation. Lens split, on the other hand, required significantly less physical effort ($Z = -2.539, p = .011$) in the implicit mode, however, other ratings were in favor of the explicit mode. In fact, the participants preferred the explicit lens split operation significantly more ($Z = -2.235, p = .025$).

6 LESSONS LEARNED

The participant feedback and the subjective ratings indicated that lens initiation through head dwell was intrusive as it made the users consciously focus their line of sight to the visualization. On the other hand, the explicit counterpart of the same action received significantly better results for all metrics. A quick and unambiguious explicit gesture seems to be better for selection operations such as lens intitiation, since the participants also felt that the dwell time (for the implicit action) was too long and was responsible for the adverse user experience. The move and scale operations were liked in the implicit mode as they were deemed to be intuitive and did not have the delays seen in lens initiation.

Lens content manipulation through zoom and pan was not very accurate, but was seen as "fun." This can be due to the algorithm for implicit action mapping in lens zoom. When the user performs these implicit actions through body or head movement, the amplitude and frequency of this action should be mapped to the corresponding lens operation. For example, when the user tries to zoom in by getting closer to the display, the zoom events are triggered

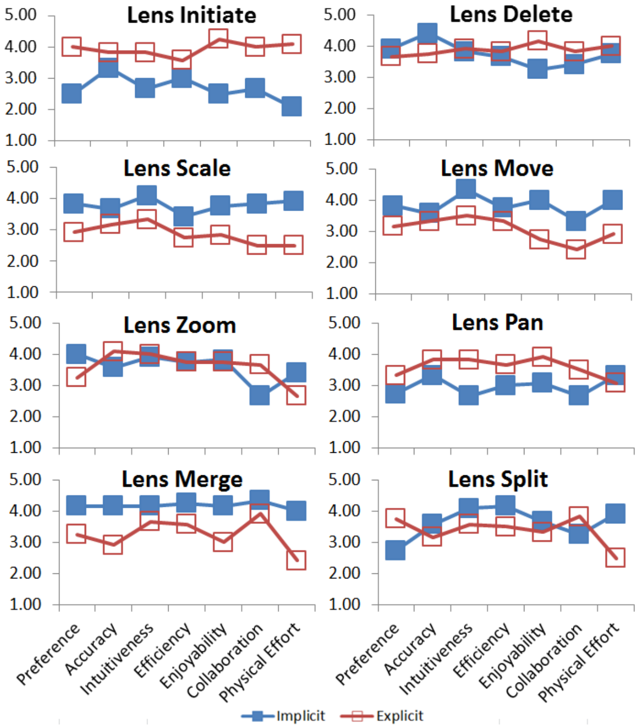


Figure 3: Mean values for Likert-scale questionnaire ratings.

Table 1: Overview of lens interactions in hybrid mode.

Context	Action	Interaction Details
Lens	Initiate	Hand orientation, Hand-roll gesture
	Delete	Hand orientation, Hand-roll gesture
	Scale	Body distance to the wall display
	Move	Body position and distance to display
	Store	Body orientation, Foot click gesture
Lens Content	Pan	Body orientation + hand swipe gesture
	Zoom	Body position/orient. + hand gesture
Multi-User	Merge	User distance + simultaneous gesture
	Split	User distance + simultaneous gesture

Figure 4: A hybrid merge interaction that uses proxemics to stack the lenses when the users are close (left) and changes the merge mode to content overlay when they perform a collaborative gesture (right).

at a frequency based on the user speed, and the amount of zoom is mapped to the distance moved by the user (amplitude). These parameters can affect user experience. On this note, even though these parameter were partly studied through pilot tests, participant feedback suggested that zoom and pan operations were not accurate for implicit actions and reflects a loss of user control.

Among collaborative operations, lens merge was well-received for the implicit setting, and it was praised to be similar to real world scenarios where people have to get close to each other to share. Lens split, on the other hand, had some issues, as users often faced a scenario where a split happened even when they did not intend for it. Lens delete and lens split are both inverse operations, and in the absence of an *undo* operation, these can affect the user experience adversely. User intention was discussed for multiple operations including zoom, pan, split and deletion, but group intention was not an aspect considered for merge and split, due to the simplicity of these operations. Further testing with complex collaborative operations is required to understand the best methods to identify group intent, which promises more conflict free collaborative operations.

7 HYBRID MULTI-USER LENS TECHNIQUE

Drawing from our lessons learned, we designed a hybrid interaction technique combining both implicit and explicit interactions. In the hybrid version, we included “lens store”, a new operation to allow storage of lenses on the floor display through a foot click gesture. User feedback indicated that being able to view the area behind a lens is necessary, since users often need to go back and forth between the lens and the overview during their exploration.

The hybrid version of our lens technique uses a mix of both implicit and explicit actions for lens initiate, scale, move, pan, zoom, merge, split, delete, and store. Table 1 summarizes the proxemic and gestural actions that were used. This does not mean that for every action designed under the hybrid technique, we have to necessarily include a mix of both proxemic and gestural interactions. However, when designing a set of such lens actions, we can now choose from a mixed pool of interactions based on both proxemics and gestures. For instance, we kept the lens move action to be performed by the implicit interaction based on user’s position relative to the wall display. One criticism from the previous study was that the implicit lens move action led to unwanted lens movements especially in collaborative scenarios when two users converse or when the users are very casual with their interactions, which lead to small movements of their heads. To avoid this, we introduced a region mapping for lens move, in which users can fixate the lens to avoid lens movement when they are close to the display. The lens merge and lens split operations in the hybrid version included both implicit and explicit aspects through a two-step process. During lens merge, the lenses would stack when users are close (resembling implicit merge), and switch to a different merge mode (e.g., content overlay) through a collaborative “hand raise” gesture (Figure 4) that captures group intent (inverse happens for lens split). These decisions were based on a design goal to allow alternative states for the proxemic interactions and trade-off the implicit nature of these interactions to some extent and achieve better user experience.

8 SUMMATIVE EVALUATION: HYBRID TECHNIQUE

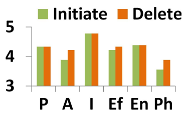
To evaluate our hybrid lens designs, we conducted a separate summative study using the same tasks as the formative evaluation and following the same methodology when collecting and analyzing the results. The goal of this evaluation was to learn about the usability aspects of hybrid actions—would users find them intuitive or confusing? Physically easy or difficult to perform? Accurate or not?

8.1 Method

Similar to the formative study, each session lasted just under an hour and involved a group of two participants performing the tasks collaboratively. We first recruited four participants for two pilot sessions. After making minor adjustments based on the pilot sessions, 18 additional participants (4 female and 14 male; 9 groups; G1-G9) who have not participated in the formative evaluation, were recruited for the actual trials. The participants were drawn from the student population at our university. There were five parts to the study: (1) training, (2) action-performance quiz, (3) six-task data analysis, (4) completing a 5-point Likert scale questionnaire, and finishing with a (5) semi-structured interview. In what follows, we report on the results of this study and discuss insights from our observations and user feedback. We did not include the "collaboration" item in the Likert scale questionnaire as we found that the participants had different perceptions of collaboration quality during the formative evaluation (since all the tasks were successfully completed). Instead, we chose to observe their collaboration style during the tasks, and inquire their individual opinions about their collaboration during the post-session interview. This left us with 6 dimensions (Preference (P), Accuracy (A), Intuitiveness (I), Efficiency (Ef), Enjoyment (En), and Physical Effort (Ph)).

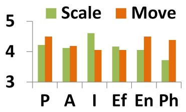
8.2 Results and Discussion

Overall, participants found the hybrid actions easy to learn. Most subjects (16 out of 18) were able to recreate the actions during the quiz at first try. However, during the actual experiment we observed that four participants were confused about pan and zoom as they forget the actions involved. In one group (G6), the teammate helped the participant remember the interaction, and in all four cases, participants were fully able to perform all actions correctly.

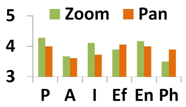


Lens Initiate and Delete. These lens operations both received high scores for all categories ($\mu > 4.2$), with the exception of physical effort ($\mu = 3.5$ and $\mu = 3.8$, respectively). Participants felt that these operations were hard to perform because the interaction required pointing to the specific chart in the overview while keeping the cursor inside the chart borders and performing a hand roll. This was easier with deletion as users deleted their lenses when far from the display at which their lenses were at the maximum size and easier to point at.

A potential solution for this issue is to use fisheye magnification to provide feedback. This would enlarge the chart the user is pointing at, thus, giving a visual feedback to reduce the ambiguity in selection. Participants found these actions very intuitive ($\mu = 4.7$); G1 described the action to be "like opening and closing doors."

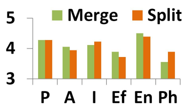


Lens Scale and Move. The lens scale and move actions, which included proxemic interactions, received high ratings across all categories in the questionnaire ($\mu \geq 4$). Participants enjoyed the implicit lens movement ($\mu = 4.5$), and mentioned, "*[it is] natural to want your lens close by at all times*" (G8). One participant said that lens movement "let him focus on the chart more" and called it "efficient" (G1). Subjects also found scaling lenses based on distance to the display to be intuitive ($\mu = 4.6$) and something one "almost does not notice" (G2). In our hybrid mode, we introduced the close region of the floor in which the lenses are automatically fixated. We observed that eight groups took advantage of this when discussing a chart.

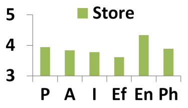


Lens Zoom and Pan. The lens zoom and pan actions received relatively low ratings for accuracy ($\mu = 3.6$ for both), efficiency ($\mu = 3.8$ and $\mu = 4$), and physical effort ($\mu = 3.5$ and $\mu = 3.8$). Both of these actions required explicit hand gestures, and we observed that participants misuse the interactions at the beginning of the trial session even though they had successfully gone through the exercise and quiz phases. Similar to other gestural interactions, we argue that the learning curves could be steep especially if the user has no prior experience with mid-air interactions.

Three participants mentioned that they did not find the pan interactions intuitive and expected them to be backwards. One participant compared this to how users switching from Mac to Windows or vice versa have a hard time adjusting to how the mouse pad scrolling works (as it is different between these systems). Prior experiences and work habits do affect the learnability of new gestures. One solution would be to introduce implicit components into these actions since they have a direct mapping. We tried to accomplish this by showing extra information on the charts automatically as annotations as the user enters the middle region [17], thus eliminating the need for some of the explicit content zooming. Participant feedback confirms the improvements targeted by this design.



Lens Merge and Split. Lens merge and split both received mixed ratings and feedback from the participants. While participants found these actions physically demanding ($\mu = 3.5$ and $\mu = 3.8$) and rated both of this actions somewhat low for efficiency ($\mu = 3.8$ and $\mu = 3.7$), their ratings for the enjoyability dimension were high ($\mu = 4.5$ and $\mu = 4.3$) deeming both of these actions as fun to perform. Seven groups attributed this to the collaborative gesture, which was used for switching the chart type after an implicit merge, and in some cases even suggested alternative interactions for this purpose: "high five" (G3) and "hand shake" (G6). While we can conclude that the new hybrid design of these actions in which the merge and split are a two-step process is liked by the participants, similar studies have shown that collaborative actions involving touch might not always be welcomed in cases where the two collaborators do not know each other [26]. Participants also commented in several cases on how these interactions promote involvement and collaboration: "*merge was fun because it required us to work together*" (G7). The implicit first step interaction sometimes happened accidentally (3 cases) but welcomed and further pursued to the second step of explicit superimposition. In two groups (G1 and G8), we observed that team members started collaborating and helping out their partners only after the first instance of the merge action.



Lens Store. Participants were very excited to try the foot click gesture—tapping the floor with one's foot to temporarily store the lens. Two participants called this interaction "useful" (G3), while another (G1) mentioned he enjoyed using his foot to interact. In two cases (G6, G8), we observed that participants tried reading the lenses while on the floor, but soon brought back the lens to the wall display for better visibility. Several participants mentioned that it was "*easy to bring the lenses to the wall and back to the floor, so there was no need for using the lens on the floor*". One participant mentioned that "*it might be awkward to use foot for interaction in a formal meeting*" (G1). One interesting usage of this interaction was made by a participant (G7) to quickly move away his lens and let his team member create a lens in the close proximity (avoiding lens occlusion). We expect that better floor projection technologies will allow wider use of this relatively unexplored surface as an extra space for analytical activites.

9 APPLICATION EXAMPLES

Our multi-user lens techniques can be applied to a wide variety of visualization types beyond the application explored in our study; examples include spatiotemporal data (maps), social network data (graphs), and even multi-view visualizations. For example, a social network (node-link diagram) can easily be a very dense network for large groups. In this case, our technique can be applied to capture the links between people within a selected region of the social network and display/interact with them on the lens. The operations such as scale, move, zoom, pan, and merge can then be used to identify community structures within connected people.

Figure 5: Google maps example using our interaction technique.

As a more complex example, we applied the hybrid lens technique to a spatiotemporal dataset of locations of cellular devices (Figure 5). The spatial positions of cellular devices were shown as targets on an overview Google Map using markers. A gaze-controlled cursor is used to highlight the region (a rectangle in this case) in the viewing direction of a user. The cursor pans and scales similar to a lens and is used to select regions on the map. The region selected, through explicit action, is shown directly on the lens, and it can be further explored through (1) implicit move and scale to explore the entire oveview map with current lens content in sight, and (2) explicit zoom and pan operations to explore the contents of the lens in more detail. For collaboration, the multiple lenses containing the Google Maps and the markers are superimposed using implicit actions through Boolean operations for combining the spatial locations visualized in each lens. These operations include, (1) *add*: target location data in both lenses is combined and visualized in a single lens (similar to a set union); (2) *subtract*: unique targets in one of the lenses are shown in a lens; (3) *intersect*: common targets in the lenses are shown in the merged lens; and (4) *exclusive*: unique targets in the lenses are shown on the merged lens (i.e., an XOR operation). These merge modes can be switched through the collaborative gesture. These actions are similar to the ones proposed in our hybrid interaction models.

10 DESIGN IMPLICATIONS

The implicit-only and explicit-only versions had actions that were felt to be hard or not preferred (rating < 3) on multiple metrics. Examples include the physical effort involved implicit lens initiation, and enjoyability and efficiency of explicit lens scaling (Figure 3). On the other hand, the hybrid actions did not perform well only on the physical effort scale, with only the lens move operation in the hybrid requiring less physical effort ($\mu = 4.5$). This gives preliminary evidence that proxemic interactions can be less demanding when they are efficient and perceived to be natural (as was the case with lens move). Explicit actions (zoom and pan) were even more physically demanding as the participants were confused with their mappings. An adverse effect on physical effort for spatial zoom and pan interaction has been observed by Nancel et al. [27], in which case alternative interaction choices through a handheld device required less effort. Therefore, a future goal is to design and evaluate techniques for cross-device interaction utilizing device proxemics.

We also observed specific collaboration styles associated with our lenses even when group activity was not necessary. For example, to answer a specific value identification task, which could have been done by a single user, the participants collaborated in two distinct ways: (1) one member of the group would open a lens and perform pan and zoom actions to move to a specific date, while the other member stays close to the display and reads out the value, or (2) both create copies of a lens and cross-check their answers. Furthermore, for tasks requiring lens merge (comparing two different views into the dataset), the collaboration was seamless as the merge operation was in itself implicit and collaborative, and participants did not have to explicitly assign roles to each other (except figuring out who creates which lens). As a participant pointed out, this process was fun since it helped them collaborate efficiently. Since the technique is based on a focus+context lens that is owned by a user, there were also no interaction conflicts as the users would never interact with their partner's active lenses.

A major goal for our lens technique was to support a generic set of operations that can be performed across many visualization and HCI scenarios. For this reason, we chose operations that control both layout and content. In addition, the technique should support collaboration, which is one of the strongest features of such mixed-modal environments (i.e., environments with large wall, multi-screen displays, and handheld devices). This was the motivation for introducing the lens concept for presentation in the first place, as it inherently provides the freedom to support both individual and group activities. While such multi-view presentations are inspired by prior work [21, 37, 38], we see the major contribution of our work to be more on the interaction side in determining a balance between implicit and explicit interactions using proxemics and gestures for the visualization tasks. This, in turn, expands on existing work on proxemics for visualization [17, 21] and HCI [10].

Finally, there is a conflict in the definition of implicit interaction, defined by Ju et al. [19] as interactions that are implied rather than explicit input and output. However, over time our users become more aware of some implicit interactions, and in turn, they perform them consciously. Beyond this, proxemics cannot be equated to implicit interaction all the time (although the terms are used interchangeably [2, 10, 21]). This does not negate the fact that these body-centered interactions are very natural, direct, and fluid in nature, as our participants commented in the summative evaluation.

11 CONCLUSION AND FUTURE WORK

We presented a multi-user lens framework for sensemaking in large display environments. The technique can be controlled in two ways: (1) using implicit actions based on the position, orientation, and distance of a person in relation to other persons or the display, and (2) using explicit actions such as swiping a hand, or tapping with a foot. To find the most appropriate interaction mode—implicit or explicit—for different lens operations, we conducted a formative user evaluation involving six pairs of participants performing analytical tasks together on a time-series visualization. Our observations from this study were used to propose a hybrid model consisting of a mix of both implicit and explicit actions. We evaluated this new hybrid model in a follow-up qualitative evaluation using an additional nine pairs of participants. We observed that the users are able to benefit from seamless interactions within the hybrid model while feeling in control performing analytical tasks.

Future work includes (1) extending the techniques to complex operations for other visualizations and usage scenarios, (2) comparing and combining the techniques with 2D interaction on handheld surfaces, and (3) exploring floor feedback and foot actions.

ACKNOWLEDGMENTS

This work was partially supported by the U.S. National Science Foundation award IIS-1539534 as well as the Canadian NSERC Discovery program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

REFERENCES

[1] C. Andrews, A. Endert, and C. North. Space to think: Large, high-resolution displays for sensemaking. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 55–64, 2010.

[2] T. Ballendat, N. Marquardt, and S. Greenberg. Proxemic interaction: designing for a proximity and orientation-aware environment. In *Pro-*

ceedings of the ACM Conference on Interactive Tabletops and Surfaces, pages 121–130, 2010.

[3] M. Beaudouin-Lafon, S. Huot, M. Nancel, W. Mackay, E. Pietriga, R. Primet, J. Wagner, O. Chapuis, C. Pillias, and J. R. Eagan. Multisurface interaction in the WILD room. *IEEE Computer*, 45(4):48–56, 2012.

[4] M. Bostock, V. Ogievetsky, and J. Heer. D3: Data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, 2011.

[5] H. Chung, C. North, J. Z. Self, S. Chu, and F. Quek. Visporter: facilitating information sharing for collaborative sensemaking on multiple displays. *Personal and Ubiquitous Computing*, 18(5):1169–1186, 2014.

[6] M. Czerwinski, G. Smith, T. Regan, B. Meyers, G. Robertson, and G. Starkweather. Toward characterizing the productivity benefits of very large displays. In *Proceedings of INTERACT*, pages 9–16, 2003.

[7] J. Dostal, U. Hinrichs, P. O. Kristensson, and A. Quigley. SpiderEyes: designing attention-and proximity-aware collaborative interfaces for wall-sized displays. In *Proceedings of the ACM Conference on Intelligent User Interfaces*, pages 143–152, 2014.

[8] E. R. Fisher, S. K. Badam, and N. Elmqvist. Designing peer-to-peer distributed user interfaces: Case studies on building distributed applications. *International Journal of Human-Computer Studies*, 72(1):100–110, 2014.

[9] B. G. Glaser and A. L. Strauss. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine Transaction, 1999.

[10] S. Greenberg, N. Marquardt, T. Ballendat, R. Diaz-Marino, and M. Wang. Proxemic interactions: the new ubicomp? *Interactions*, 18(1):42–50, 2011.

[11] E. T. Hall. *The Hidden Dimension*. Anchor Books, Garden City, NY, 1966.

[12] C. Harrison and A. K. Dey. Lean and zoom: proximity-aware user interface and content magnification. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 507–510, 2008.

[13] J. Heer and B. Shneiderman. A taxonomy of tools that support the fluent and flexible use of visualizations. *ACM Queue*, 10(2):1–26, 2012.

[14] J. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani. Consumed endurance: A metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 1063–1072, 2014.

[15] P. Isenberg, P. Dragicevic, W. Willett, A. Bezerianos, and J.-D. Fekete. Hybrid-image visualization for large viewing environments. *IEEE Transactions on Visualization & Computer Graphics*, 19(12):2346–2355, 2013.

[16] M. Jakobsen and K. Hornbæk. Proximity and physical navigation in collaborative work with a multi-touch wall-display. In *Extended Abstracts of the ACM Conference on Human Factors in Computing Systems*, pages 2519–2524, 2012.

[17] M. R. Jakobsen, Y. Sahlemariam Haile, S. Knudsen, and K. Hornbæk. Information visualization and proxemics: design opportunities and empirical findings. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2386–2395, 2013.

[18] W. Javed and N. Elmqvist. Exploring the design space of composite visualization. In *Proceedings of the IEEE Pacific Symposium on Visualization*, pages 1–8, 2012.

[19] W. Ju, B. A. Lee, and S. R. Klemmer. Range: exploring implicit interaction through electronic whiteboard design. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, pages 17–26, 2008.

[20] A. Kendon. Spacing and orientation in co-present interaction. In *Development of Multimodal Interfaces: Active Listening and Synchrony*, pages 1–15. Springer, 2010.

[21] U. Kister, P. Reipschläger, F. Matulic, and R. Dachselt. Bodylenses: Embodied magic lenses and personal territories for wall displays. In *Proceedings of the ACM International Conference on Interactive Tabletops & Surfaces*, pages 117–126, 2015.

[22] S. Malik, A. Ranjan, and R. Balakrishnan. Interacting with large displays from a distance with vision-tracked multi-finger gestural input. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, pages 43–52, 2005.

[23] N. Marquardt, T. Ballendat, S. Boring, S. Greenberg, and K. Hinckley. Gradual engagement: facilitating information exchange between digital devices as a function of proximity. In *Proceedings of the ACM Conference on Interactive Tabletops and Surfaces*, pages 31–40, 2012.

[24] N. Marquardt, R. Diaz-Marino, S. Boring, and S. Greenberg. The proximity toolkit: prototyping proxemic interactions in ubiquitous computing ecologies. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, pages 315–326, 2011.

[25] W. McGrath, B. Bowman, D. McCallum, J. D. Hincapié-Ramos, N. Elmqvist, and P. Irani. Branch-explore-merge: Facilitating real-time revision control in collaborative visual exploration. In *Proceedings of the ACM Conference on Interactive Tabletops and Surfaces*, pages 235–244, 2012.

[26] M. R. Morris, A. Huang, A. Paepcke, and T. Winograd. Cooperative gestures: multi-user gestural interactions for co-located groupware. In *Proceedings of the Conference on Human Factors in Computing Systems*, pages 1201–1210, 2006.

[27] M. Nancel, J. Wagner, E. Pietriga, O. Chapuis, and W. Mackay. Mid-air pan-and-zoom on wall-sized displays. In *Proc. of the ACM Conference on Human Factors in Computing Systems*, pages 177–186, 2011.

[28] A. Olwal and S. Feiner. Spatially aware handhelds for high-precision tangible interaction with large displays. In *Proc. of the ACM Conference on Tangible and Embedded Interaction*, pages 181–188, 2009.

[29] S. M. Peck, C. North, and D. Bowman. A multiscale interaction technique for large, high-resolution displays. In *IEEE Symposium on 3D User Interfaces*, pages 31–38, 2009.

[30] J. Rekimoto and M. Saitoh. Augmented surfaces: a spatially continuous work space for hybrid computing environments. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 378–385, 1999.

[31] Z. Ren, J. Meng, and J. Yuan. Depth camera based hand gesture recognition and its applications in human-computer-interaction. In *Proceedings of IEEE Conference on Information, Communications and Signal Processing*, pages 1–5, 2011.

[32] D. Schmidt, J. Seifert, E. Rukzio, and H. Gellersen. A cross-device interaction style for mobiles and surfaces. In *Proceedings of the ACM Conference on Designing Interactive Systems*, pages 318–327, 2012.

[33] J. S. Shell, T. Selker, and R. Vertegaal. Interacting with groups of computers. *Communications of the ACM*, 46(3):40–46, 2003.

[34] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, pages 336–343, 1996.

[35] G. Shoemaker, T. Tsukitani, Y. Kitamura, and K. S. Booth. Body-centric interaction techniques for very large wall displays. In *Proceedings of the ACM Nordic Conference on Human-Computer Interaction: Extending Boundaries*, pages 463–472, 2010.

[36] K. Takashima, S. Greenberg, E. Sharlin, and Y. Kitamura. A shape-shifting wall display that supports individual and group activities. Technical report, Department of Computer Science, University of Calgary, 2015.

[37] A. Tang, M. Tory, B. Po, P. Neumann, and S. Carpendale. Collaborative coupling over tabletop displays. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pages 1181–1190, 2006.

[38] M. Tobiasz, P. Isenberg, and S. Carpendale. Lark: Coordinating co-located collaboration with information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1065–1072, 2009.

[39] D. Vogel and R. Balakrishnan. Interactive public ambient displays: Transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, pages 137–146, 2004.

[40] M. Weiser. The computer for the 21st Century. *Scientific American*, 265(3):94–104, 1991.

[41] A. D. Wilson and H. Benko. Combining multiple depth cameras and projectors for interactions on, above and between surfaces. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, pages 273–282, 2010.