Data-Intensive Remote Collaboration using Scalable Visualizations in Heterogeneous Display Spaces

BY

THOMAS MARRINAN
B.S. Computer Science, B.A. Graphic Design, Drake University, 2010

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Computer Science
in the Graduate College of the
University of Illinois at Chicago, 2016

Defense Committee:

- Andrew Johnson, Chair and Advisor
- Luc Renambot
- Angus Forbes
- Steve Jones, Department of Communication
- Jason Leigh, University of Hawaiʻi at Mānoa

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Andrew Johnson, who provided mentorship throughout my graduate career and challenged me as a student and a researcher. He also provided inspiration and support in my studies to help discover my true passion in Computer Science. A special thanks goes to Dr. Luc Renambot, whom I collaborated closely with on both research and development. I would also like to thank the rest of my dissertation committee for providing insight and feedback from the time I was developing a research question until the time it culminated with analyzing the results of evaluating studies.

Additionally, I must give a big thank you to the entire Electronic Visualization Laboratory. The students, faculty, and staff have been incredibly supportive. I feel lucky to have worked with each one of you, and I will always remain a member of the EVL family.

TM

CONTRIBUTION OF AUTHORS

Chapter 1 introduces an active area of research to which my dissertation makes its contributions. Chapter 2 provides a background about the technologies I used during my research, and includes portions of a published manuscript (Marrinan et al., *SAGE2: A New Approach for Data Intensive Collaboration Using Scalable Resolution Shared Displays*, 2014) for which I was the primary author. Chapter 3 is a literature review of related works that frame my dissertation within the scope of research in the field of computer supported cooperative work. Chapter 4 details my research methods for implementation and evaluation of a new technology, as well as outlines the main questions my dissertation sought to answer. Chapter 5 presents the results of user studies I conducted and provides insight into answering my research questions. Chapter 6 concludes my thesis by summarizing the knowledge that was gained and providing future area to expand upon my research. My advisor and chair, Andrew Johnson, along with the rest of my committee contributed valuable feedback during the editing of this document.

PREFACE

As a graduate student at the Electronic Visualization Laboratory, my research interests include visualization, human-computer interaction, and computer-supported cooperative work. The main research topic I aimed to contribute toward is how to improve collaboration across distance, especially when the technology is heterogeneous at various locations. Video and teleconferencing systems have been widely adopted in research and industry for communication between individuals or groups at various locations. The next generation of communication systems will enable real-time data-conferencing, where applications and their respective data are shared along with audio and video. Although some data-conferencing abilities have started being integrated into existing software, they are generally still in the early stages and do not allow for real-time collaboration on unrestricted data types.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF FIGURES

FIGURE **PAGE**

LIST OF FIGURES (Continued)

FIGURE **PAGE**

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

FIGURE **PAGE**

LIST OF ABBREVIATIONS

EVL	Electronic Visualization Laboratory
SAGE	Scalable Adaptive Graphics Environment
SAGE2	Scalable Amplified Group Environment
SRSD	Scalable Resolution Shared Display
CAVE	CAVE Automatic Virtual Environment
PDT	Partially Distributed Team
AJAX	Asynchronous JavaScript And XML
DOM	Document Object Model
Mpixel	Megapixel (1,000,000 pixels)

CHAPTER 1

INTRODUCTION

Commercial software, such as Google Hangouts, Skype, and WebEx [1-3], has made significant strides to enhance remote collaboration between single users. Other systems, such as Oblong's Mezzanine [4], have enabled groups to collaborate over distance, but only when the display environment is cloned in all locations. These configurations are not always practical in the real world since collaboration is often interdisciplinary, with each discipline having a unique work environment suited for its needs. I have used SAGE2, the Scalable Amplified Group Environment [5], as a platform to implement a scalable solution for viewing and interacting with arbitrary content, and developed and tested various techniques for synchronizing data across multiple sites. SAGE2 is the successor to SAGE, the Scalable Adaptive Graphics Environment, which is a middleware to display and interact with an assortment of data-intensive information from multiple sources on displays of arbitrary size [6-8]. In other words, SAGE2 allows collaborators to use any Scalable Resolution Shared Display (SRSD) from a single monitor to a large tiled display wall to act as a seamless window manager for applications such as a high-resolution image viewer, interactive mapping software, 3D model viewer, and multi-user notepad (Figure 1).

Unlike video and audio, data-conferencing with fully synchronized content may not always be ideal. Groups or individuals may be analyzing different portions of the same data, and therefore desire unsynchronized interactions in certain situations. I have conducted user studies comparing three techniques of data-conferencing. The first technique was *data-pushing*, sending unsynchronized documents between locations, along with two video streams – one of the collaborators and one of the SRSD. The second technique was *data-duplication*, where one section of the SRSD contained fully synchronized versions of all data-conferencing content and a second portion of the SRSD contained local unsynchronized copies of the data-conferencing content. The final technique was to use *advanced data-synchronization options*, where collaborators chose which aspects of each shared application were synchronized and which were controlled independently. These studies aimed to answer three main research questions:

1) Does providing continuously synchronized applications improve the quality of collaboration and awareness of the remote team?
2) Can synchronizing applications provide remote teams the same quality of collaboration as local teams?
3) Does the size of a shared display have an effect on the quality of collaboration and awareness of the remote team?

While I used SAGE2 to develop and test these principles, the applications of the knowledge gained can broadly apply to data-conferencing software in general.

Figure 1. A typical SAGE2 session, depicting multiple applications windowed on a SRSD. This figure illustrates the wide array of content supported by SAGE2, including images, videos, PDFs, 2D and 3D custom applications, and off-the-shelf applications from remote sources.

CHAPTER 2

BACKGROUND

Parts of this chapter were previously published as:

T. Marrinan, J. Aurisano, A. Nishimoto, K. Bharadwaj, V. Mateevitsi, L. Renambot, L. Long, A. Johnson, and J. Leigh, "SAGE2: A New Approach for Data Intensive Collaboration Using Scalable Resolution Shared Displays," in *Proceedings of the IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing* (CollaborateCom '14), 2014, pp 177-186.

Today, scientific and industrial data is collected, stored, and analyzed digitally, often in the cloud. The resolution of this data is continuously increasing with improvements to the instruments and sensors used for measuring and monitoring the physical world. Additionally, supercomputers are being leveraged to simulate natural phenomena, from global weather systems to chemical reactions at the atomic level, generating massive volumes of data. These troves of data are invaluable to researchers and businesses as they explore the raw information and evidence needed for new insights, discoveries, and innovations. However, making those insights is an increasingly complicated task as the scale and complexity of data continue to grow at unprecedented rates. Since big data problems frequently require the combined efforts of many individuals from disparate fields, the next generation of data intensive visualization and interaction environments will need to enable collaboration and group work.

To deal with the scale and complexity of data, the 2007 DOE Visualization and Knowledge Discovery workshop report [9] and the 2008 NSF Building Effective Virtual Organizations workshop report [10] recognized that new modalities for accessing more visual information were necessary. They both described large shared displays as the type of environments that are crucial for next-generation collaborative cyber-enabled exploration. Furthermore, there is now conclusive evidence that large display environments enable collaboration and significantly amplify the way users make sense of large-scale, complex data [11-19].

2.1 Ultra High-Resolution Shared Displays

Since large display environments present technical challenges and unique affordances, specialized software and middleware are needed to allow users to capitalize on the provided benefits. During the 1990’s, multi-projector systems were constructed to create visualization environments. The CAVE [20,21] was invented as a surround-screen virtual reality environment for small groups to use simultaneously. The CAVE is a cube-shaped room with projectors rendering stereo images on all six sides. One user at a time could be head-tracked to create user-centered perspective enabling complete immersion within the environment. PowerWall visualization systems [22] were developed to create high-resolution display environments. PowerWall systems consisted of an array of projectors pointed in the same direction to create a large seamless image on a wall. Their main purpose was for visualizing scientific simulations and facilitating small group collaboration.

In the early 2000’s, researchers began tiling flat panel displays in order to create ultra high-resolution display environments with enhanced color contrast. Middleware to display content and enable groups to interact with these systems were mainly developed following two distinct paths. The distributed rendering model [23] uses a cluster of consumer personal computers connected with a high-speed network to render an application across a set of tiled displays. The OptIPuter strategy [24] takes advantage of the exponential growth in network bandwidth and data storage to stream information to a tiled display wall from distributed remote computers, which do all the data retrieval and processing. The evolution of software solutions that enable scientists to display ultra high-resolution imagery are described in detail next for both the distributed rendering model and the OptIPuter strategy.

2.1.1 Distributed Rendering Systems

WireGL [25] represents one of the first distributed rendering systems designed for tiled display systems. WireGL virtualizes multiple graphics accelerators and streams OpenGL API commands to each node in a render cluster. This represents a shift from previous high-performance rendering systems, which required high-end multiprocessor servers. Chromium [26] represents an improvement on WireGL. Chromium also streams graphic API commands to cluster nodes, but focused on optimizing resource utilization. Rather than only considering display resolution, the Chromium framework load-balanced rendering tasks to distribute pixel tiles across the cluster. Equalizer [27] was later developed to further take advantage of each node's graphics accelerator. Equalizer runs parts of the graphical application in parallel on each display node, rather than running the whole application on the head node and streaming rendering calls to each node. This improves performance over systems such as Chromium and WireGL. The Cross Platform Cluster Graphics Library (CGLX) [28] also allows OpenGL applications to run on visualization clusters in order to be scalable to ultra high-resolutions. Performance test show that CGLX scales well with increasing resolution and scene complexity. The authors claim that CGLX is preferable to Equalizer due to the fact that Equalizer requires in depth knowledge about system configurations.

Traditionally, distributed rendering systems have been focused on high-performance visualization of a single ultra high-resolution interactive graphical application, often used in virtual reality. These systems have been built using low-level programming languages and APIs in order to maximize optimization and render speed. A few systems have recently been developed that advance these concepts and test new methods that achieve greater flexibility. Omegalib [29] is a framework to facilitate application development for hybrid reality environments. Omegalib is built on top of Equalizer and enables high-performance 2D and 3D graphics. This enables users to integrate information-rich analysis with virtual reality immersion. Omegalib is also configurable, allowing the visualization cluster to be partitioned, which enables multiple applications to run simultaneously. HTML5 canvas drawing [30] has recently been leveraged to utilize web-browsers for distributed rendering of web applications. A web server is responsible for distributing the rendering of native content drawn in a canvas element. Code is injected into the browser clients in order to parallelize rendering of ultra high-resolution content. The authors demonstrated that their system worked at resolutions up to 8240x4920 pixels over 16 tiled displays.

2.1.2 OptIPuter Systems

Vol-a-Tile [31] represents an early application that utilizes the OptIPuter paradigm for ultra high-resolution visualization of data-intensive content. Vol-a-Tile was designed to show large volumetric datasets on tiled display walls. It runs in parallel and uses MPI to communicate between display nodes of a visualization cluster. Data is streamed from a remote storage center to provide dynamic level-of-detail 3D images to the application. SAGE [6-8] is an open-source middleware that provides multiple users with a common operating environment, to access, display, and share an assortment of data-intensive information. The software allows each user to create a pointer on a tiled display wall by using their own personal device, or to directly approach the tiled display wall and interact through a multi-touch interface. In this manner, multiple users can simultaneously add and interact with content. SAGE displays uncompressed pixel streams from remote sources by utilizing high-speed networks to render content ranging from high definition images and videos to PDF documents and laptop screens. Montage [32] uses the application integration library provided by SAGE to render multiple web pages on a tiled display wall. Grouping and filtering techniques enabled users to make connections and discoveries difficult to achieve with standard resolution monitors. DisplayCluster [33] is a desktop-like windowing environment for cluster-driven tiled displays. DisplayCluster is similar to SAGE, but incorporates compression and decompression algorithms in the display nodes in order to reduce necessary bandwidth for raw pixel streaming. However, the compression algorithms can introduce visual artifacts. Therefore SAGE represented a system for scientific researchers who needed pixel perfect visualizations, and DisplayCluster represented a system for the array of researchers who did not have access to the high bandwidth networks.

Single machines have become powerful enough to control multiple monitors. Today, many businesses and research labs use a single PC to drive a tiled display wall. This has the benefit of not requiring specialized software and middleware to control ultra high-resolution content. However, these single machine tiled display walls still have limitations, since they are not infinitely scalable. Table I illustrates the most recent highest resolution display walls. All of these tiled display walls afford scientists with massive amounts of resolution that elucidate new discoveries in their data [34-40]. Tiled display walls with resolutions over 100 Mpixel, and certainly over 1000 Mpixel, are still not achievable using a single machine. Therefore it is still imperative to utilize graphics clusters with specialized software and middleware in order to display and interact with massive resolution visualizations. Additionally, standard operating systems that typically drive a single PC, such as Windows, Mac OS X, and Linux, are designed for a single user with a single set of input devices. The same specialized software and middleware that facilitate cluster-based graphical environments can also enable simultaneous multi-user interaction using an assortment of input devices. While technology has improved and resolution of visualization systems has increased, research surrounding collaboration supported by large shared displays remains a topic worth investigating.

TABLE I

This table depicts the tiled display walls that were the highest resolution system at the time they were built. These systems range from 100 Mpixel to 1500 Mpixel and are all driven by a graphics cluster.

Tiled Display System	Location	Year	Resolution
LamdaVision	University of Illinois at Chicago	2005	106 Mpixel
HIPerWall	University of California, Irvine	2005	205 Mpixel
Hyperwall-2	National Aeronautics and Space Administration	2008	256 Mpixel
HIPerSpace	University of California, San Diego	2008	287 Mpixel
Stallion	Texas Advanced Computing Center	2008	307 Mpixel
OptIPresence	University of California, San Diego	2009	316 Mpixel
Reality Deck	Stony Brook University	2012	1500 Mpixel

2.2 SAGE and SAGE2

In 2004, researchers at the EVL developed SAGE, the Scalable Adaptive Graphics Environment [6], as a middleware for SRSDs. SAGE is a window manager for tiled display walls that supports remote applications that stream rendered pixel buffers over high-speed, high-bandwidth, low-latency networks. The SAGE framework breaks content into pixel blocks to stream data only to the corresponding displays. Benchmark test show that SAGE achieves low latency and high throughput, thereby providing interactive frame rates [7]. Any OpenGL application can be made SAGE compatible with the use of the SAGE Application Integration Library (SAIL). SAGE also can replicate an application's visual output to multiple sites in order to facilitate remote collaboration [8]. At the time, SAGE, in conjunction with an SRSD, represented a new type of "digital lens" – an ultra high-resolution display that can effectively visualize large volumes of data in a collaborative environment. As a result, over one hundred and fifty sites have adopted SAGE around the world over the past decade. However, its architecture was based on a monolithic design that made it increasingly difficult to integrate new capabilities as user requirements grew.

Since SAGE was released, we have worked closely with users in academia, research, and industry to capture authentic collaborative workflows and motivate SAGE development. In addition to these direct observations of SAGE in use, we conducted a user survey in 2012 to capture feedback from 40 sites on how SAGE was used, its benefits, and features users wished SAGE had. Of the desired future features of SAGE, the most requested were:

1) Integration of multi-user applications
2) Enhanced real-time distance collaboration
3) A reduced barrier to entry

2.2.1 Collaborative Workflows Enabled by SAGE

Results from our user survey indicate that our user community is diverse, with 63% from academia, 20% from industry, and 17% government research labs. 70% of these sites reported that their SAGE installation is used for more than one project. Specific uses of SAGE include sharing telemedicine lectures, showing the output of large-scale scientific simulations, and running multiple ultra high-resolution applications simultaneously. Our survey also indicated an upward trend on adoption and usage of SRSD technologies. Currently 55% of sites have more than one SAGE installation, with 62% projecting to have more than one in the next four years. Also, currently 20% of sites have more than four SAGE installations, with 32.5% projecting to have more than four in the next four years. These installations vary in resolution, with many sites having SRSDs below 8 MPixels and many other sites having SRSDs above 100 MPixels.

In addition to receiving feedback through a user survey, we have worked closely with authentic SAGE users and observed its use. Theses observations have given us insight on several types of collaborative workflows enabled by SAGE that are difficult to achieve through other platforms. For instance, SAGE was often used for regular, large group meetings where group members presented short updates to their collaborators.

In these meetings, SAGE enabled *lightning collaboration*, where each collaborator could rapidly share content, such as images, videos, and PDFs, or share the screen from their laptop, in order to illustrate their progress. Transitions between presenters were quick due to each user being able to load content and create a pointer from their personal device, rather than taking turns using a master controller.

We also observed collaborators using SAGE in a *parallel investigation workflow*. In this workflow, several collaborators gathered together to investigate a problem using a SRSD. Each user had specific domain expertise that pertained to one aspect of a problem they were attempting to solve. Working on personal machines, these users each explored one or two online databases, or loaded data in an interactive application on their laptops to display its content. Seeing these disparate pieces of information together on a SRSD prompted new questions, which each researcher explored in parallel during the meeting, sharing results step-by-step in SAGE. This work was only possible because SAGE permitted multiple users to simultaneously share and control content on the SRSD.

SAGE was used in another type of collaborative work session, which we call *single-driver, multiple-navigators*. In this type of session, the driver would share an interactive application running on his/her laptop to explore a dataset and share findings. Navigators would ask questions, prompting the driver to perform specific operations on the dataset through his/her personal machine. SAGE enabled this collaborative work by allowing everyone to see the content simultaneously on a SRSD and by capturing real-time changes on the user's machine. In addition, navigators could point to content on the SRSD, using pointers created from their personal machines, to highlight aspects of the content.

Lastly, SAGE has been used for remote collaborative work sessions, with participants sharing multimedia content between SAGE environments with *remote gigabit data streams*. These modes of co-located and remote collaboration work remarkably well for a variety of problems, but only act as a step towards improved computer supported cooperative work.

2.2.2 Collaborative Roadblocks in SAGE

While the original SAGE has enabled collaborative workflows that were previously not possible, users have identified situations in which they desired more. In order to share and interact with content on the SRSD, users needed to download and install a client application. We noticed that in situations with frequent new users, the requirement to install this client application to engage in a SAGE session acted as a barrier to entry and limited participation for new users in *lighting collaboration* scenarios.

We observed that *parallel investigation workflows* were challenging in places where users wanted to interactively engage the content of their collaborators or synthesize results from multiple investigations, since content on shared screens remained controllable only by the single owner. We observed that this often resulted in duplicated work or diminished participation. We also noticed that while users could share large volumes of information, the lack of integration across content made it challenging for researchers to combine results from the parallel investigations.

In collaborative workflows involving *single-driver, multiple-navigators*, there was a challenge with collaborators needing to switch roles. Since everyone could not directly manipulate content within the shared application simultaneously, users would be forced to take turns as the driver. Occasionally, navigators needed to duplicate the work of the driver, so they could interact with the data on their personal devices. This delayed the collaborative session, and in some instances limited participation and work sharing.

Finally, we noted that remote collaboration frequently involves configurations of users that were not adequately supported by the original SAGE architecture. Remote single users (a user without access to a SAGE session or SRSD) were limited to viewing and sharing content by standard teleconferencing systems, which did not adequately allow for participation. In addition, while multiple SAGE sites could freely and independently reposition content to fit the needs of a co-located group, other sites were unable to gain insight on the position of individual items on remote SRSDs. This often resulted in miscommunication due to the lack of context (e.g. “see the image on the right”).

Our user survey supports our observations concerning collaborative roadblocks. Users listed over 25 applications that they wished were SAGE compatible for multi-user interaction, many of which were web-based, such as Google Maps. Users also identified enhanced remote collaboration through data sharing and native teleconferencing as a top desired priority.

2.2.3 SAGE2 Development

In 2013, I began development with a team of researchers on SAGE2 as a complete redesign and implementation of SAGE. We investigated a new paradigm, leveraging cloud-based and web-browser technologies to drive a SRSD and enable real-time communication and multi-user interaction. We discovered that this novel web-based platform can achieve the performance of stand-alone cluster software and middleware, while better enabling authentic, multi-user, interactive collaborative workflows.

Our goal in designing SAGE2 was to create the next generation system for facilitating data intensive co-located and remote collaboration using a SRSD. Interactive applications with simultaneous multi-user input, enhanced real-time communication, and a lower barrier to entry were three major priorities according to feedback from authentic users from the first generation of SAGE. To address these needs, we aimed to completely redesign and implement SAGE due to the fact that its aging architecture was not well suited to handle emerging technologies. Additionally, SAGE and other platforms for driving large display environments are built as custom standalone applications, which in our experience requires a technical expert with hours of training to install and support, limiting the adoption at new sites. Therefore, we decided to pursue a different approach that leverages cloud-based and web-browser technologies for their increasing collaborative power, flexibility, and ubiquity.

The power of web-browsers, and JavaScript programming [41], is increasing at an extraordinary rate. Browsers now support native two-dimensional and three-dimensional rendering through HTML5 and WebGL [42]. Hardware acceleration is leveraged for both rendering and CSS transforms. WebSocket communication has been standardized, enabling persistent two-way communication between server and client. Real-time peer-to-peer communication, using WebRTC [43], is currently under development with many features already integrated into most mainstream browsers. Additionally, browsers have adopted event handlers for many inputs, allowing interaction from devices such as a mouse, keyboard, and touchscreen.

Web-browsers have become a ubiquitous application found on any visual computing device. They are platform independent and do not require technical expertise to install. The web-browser also acts as a portal to the vast amounts of data stored and accessed through the cloud. Numerous APIs allow developers to retrieve static and dynamic data, perform data manipulation, and visualize and store the results. Web- based communication also enables real-time collaboration and data sharing through peer-to-peer connections.

These recent developments to web-based technologies have enabled high performance graphics and networking capabilities that were formerly only achievable with native applications. Additionally, the ubiquity of web technologies allows software to run on any modern operating system. Therefore, we have designed SAGE2 using the web-browser for rendering and user interaction, and the cloud-based infrastructure for data retrieval and real-time communication.

We decided to explore whether or not the new features of web browsers would allow them to perform as well as standalone applications in regard to high-performance graphics and networking. The SAGE2 architecture provides a proof of concept that applications can leverage the web browser runtime environment to achieve the necessary performance to drive SRSDs and support large volumes of data, while also benefiting from the cloud-based infrastructure to facilitate remote collaboration.

SAGE2 consists of several components: the server, various display clients, the audio client, various interaction clients, and the input client. Figure 2 illustrates how clients connect to the Server, which handles all inputs and outputs to maintain and synchronize content. The server is built upon Node.js [44], a platform for building network applications in JavaScript. Node.js is cross-platform and comes with a package manager that downloads most dependencies. This greatly reduces the time for installation and no longer requires a technical expert. In order to handle certain specific file formats, the SAGE2 server additionally depends on a few external applications. ImageMagick [45] is used for processing and converting images; FFmpeg [46] is used for processing and decoding videos; ExifTool [47] is used to extract metadata, such as creation time, dimensions, and file size, from various file types. The only two prerequisites for running SAGE2 are that Node.js, ImageMagick, FFmpeg, and ExifTool are installed on the machine that hosts the server, and up-to-date web browsers are installed on any machine running a client.

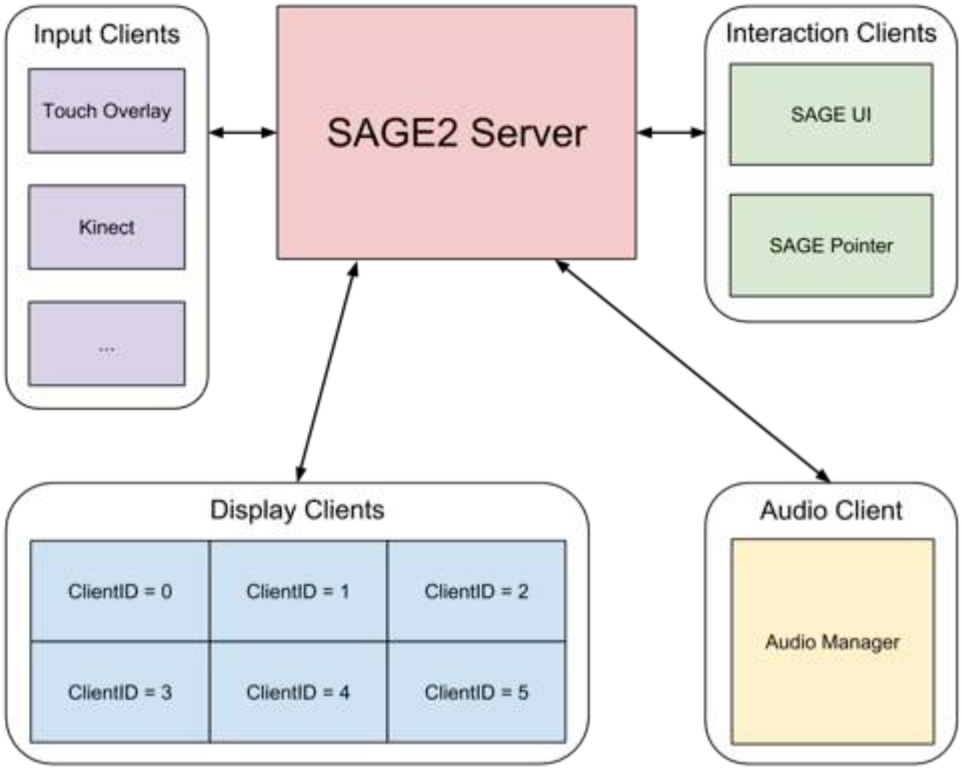


Figure 2. Illustration depicting SAGE2's architecture and communication scheme. The SAGE2 Server is a customized web server with clients accessed through visiting a URL in a web-browser.

2.2.4 SAGE2 Design Challenges

To explain how we achieved this novel approach, I will describe the process of overcoming numerous technical design challenges that were presented throughout the development process.

The first challenge was supporting cluster environments. One of the priorities for designing SAGE2 was hardware flexibility, so that it could be configured for scalable display resolutions. A SRSD can be run by one or multiple machines each connected to one or multiple monitors. Any of the SAGE2 components can run on the same machine or distributed across a cluster. Supporting a cluster environment presented multiple challenges. First, we needed to make all monitors appear as one seamless graphical environment, regardless of the number or configuration of display client machines. The SAGE2 display clients are instances of a web browser that connect to the SAGE2 Server by visiting a URL. To address this challenge we attach a parameter to the URL in order to identify each display client with a unique ID, mapping it to a specific row and column on the SRSD. Given its row and column, each display client shows its own viewport by offsetting content, based on its position on the grid. This results in users viewing content that spans multiple display clients that appears continuous and can be moved across the display seamlessly. Synchronizing audio with multiple video feeds across the SRSD presented another challenge. We determined that a separate audio client was necessary in order to mix all the audio sources. Like the SAGE2 display clients, the audio client runs in a web browser and gets initialized by connecting to the SAGE2 server by visiting a URL. When a video file is loaded into SAGE2, the audio client will output the sound and synchronize based on server commands, such as play / pause, update time, or mute. Another challenge with supporting cluster-based systems was how to properly handle distributed rendering with synchronized animations for interactive applications. To address this, each display client renders a portion of the overall application depending on the application's window size and position. Using WebSockets, the SAGE2 server handles animation synchronization by broadcasting instructions to all display clients to redraw. Each display client responds to the server when it has finished rendering its frame. Once the server receives responses from all display clients, it broadcasts the next redraw command. Since the display clients can be on completely separate machines, we cannot always achieve complete synchronization due to the fact that we do not have control over display refresh rates. However, this method enables applications to remain visually synchronized by never becoming more than one frame out of sync.

The second challenge we addressed was how to enable real-time multi-user interaction. Web pages rendered in a browser serve as both our display clients and our interaction clients. This allows users to interact with content on the SRSD by opening a browser and connecting to the SAGE2 server by visiting a URL. This greatly reduces the barrier to entry for new users since there is nothing to install, and they simply use an application that they are already familiar with. By visiting the SAGE2 interaction client page, a user can see an overview of the SRSD and can load any of the supported data types or applications. The user can also easily and quickly share local documents, show their screen, or add a pointer to the SRSD. Since the interactions clients are running on personal devices, there was a challenge in correlating interaction events with desired actions on the SRSD. Additionally, web browsers are not designed to handle simultaneous inputs from multiple distinct users. To address this, we capture events from the interaction client devices, such as a laptop or smartphone, and forward them to the SAGE2 server. The server then broadcasts these events to each display client. This enables remote devices to be used for interaction and removes the constraint of one pointer per machine. Another challenge we addressed was expanding the types of user interaction enabled by SAGE2. We wanted to support application interaction and window management in four major interaction zones for SRSDs: directly at the display using touch gestures, standing near the display using motion tracking or 6DOF devices, seated near the screen using a gyro-mouse or laptop pointer, or indirect control further away from the screen using the web interaction client [48]. To accomplish this, we utilize the Omicron input abstraction utility library [49]. Omicron is capable of receiving data from different types of input devices including touch overlays, motion tracking systems, game controllers, and speech recognition tools (Figure 3). Omicron then streams data from these heterogeneous devices in a uniform manner that the SAGE2 Server can interpret.

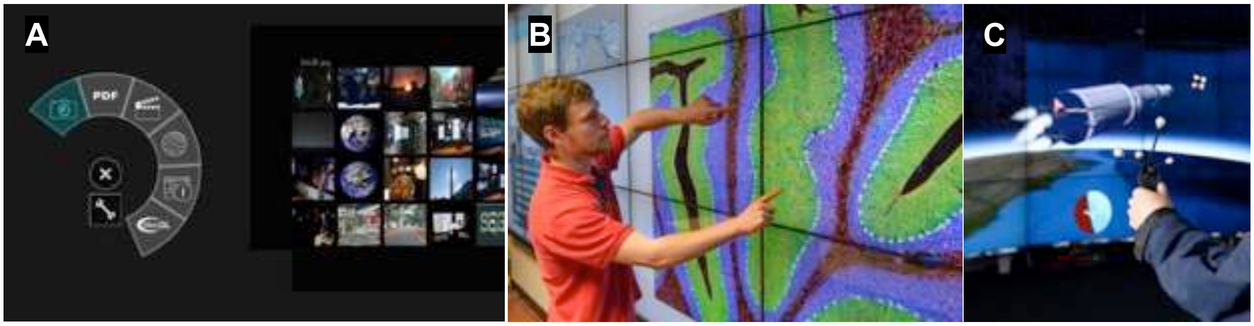


Figure 3. Multi-user interaction with expanded input modalities. Panel A shows a radial menu system that allows multiple users to simultaneously interact with a media browser. Panel B depicts a user interacting at a SRSD using a multi-touch overlay to perform a pinch-zoom gesture. Panel C shows a user interacting with a game controller that has been supplemented with motion tracking reflectors.

The final challenge we addressed was how to support multi-user applications. Interacting with ultra high-resolution displays from a distance makes precision pointing difficult. In standard desktop environments, resizing a window is commonly done by click-and-drag on a corner and moving a window by click-and-drag on a title bar. In order to remove the precision necessary to accomplish these tasks, we have enabled two interaction modes, window manipulation and application interaction, that a user can toggle between. When in window manipulation mode, events anywhere on an application will be performed on the window (click-and-drag to move, scroll to resize). When in application interaction mode, all events are forwarded to the interactive application to handle. Another challenge was to support loading custom web applications dynamically. Typically, web applications are only loaded once per browser tab and utilize the browser's default event handling system. In order to create multiple instances of a web application across the SRSD, we have encapsulated all web applications in a JavaScript class. Therefore, an application can be dynamically instantiated as many times as requested. We also forward SAGE2 events to these applications, so that they can respond to multi-user inputs, such as independent keyboards, pointing devices, and touches. A final challenge with supporting applications in SAGE2, was sharing them across sites. In order to share applications with remote sites and engage in collaborative interaction, we have designed all applications as objects accessible through a URL. This allows a reference to the application to be retrieved from any remote location, whether it is another SAGE2 site or a remote single-user on his/her laptop.

SAGE2 has a development team distributed between the University of Illinois at Chicago and the University of Hawaiʻi at Mānoa. Various developers contributed to the design and implementation of its features. Table II provides a snapshot of my contributions compared to the team at large.

TABLE II

Summary of my development contributions to SAGE2 in the scope of the whole project. P: Primary Contributor, C: Co-Contributor, S: Secondary Contributor.

Developer	Core	Interaction Client	Display UI	App Widgets	Security	Remote Collaboration	Pixel Stream	Basic Apps	Design Advice
Marrinan	C	P	P			P	P	C	C
Others	C	S	S	P	P		S	C	C

2.3 Early Experiences with SAGE2

SAGE2 is still under development, with its first beta release in November 2014. Since that time, a number of early adopters have installed SAGE2 and are using regularly. However, prior to my studies, it had not been formally evaluated. The rest of this section will discuss early observations of its use and outline three fictional use cases based on real-world scenarios that highlight how SAGE2 can uniquely handle the following collaboration sessions:

1) Imitate co-located collaboration with a physically distributed team
2) Collaboration between multiple teams working on different aspects of a related problem
3) Allowing a single remote user to join a collaborative session.

Developers interested in scalable visualizations tailored to multi-user interaction have commented on the ease of application creation. The wealth of third-party libraries for web applications provides developers with the capability to integrate advanced data processing and visualization features into custom applications. Also, SAGE2's multi-input event handler provides a unique API to design applications specifically meant for multiple simultaneous users.

SAGE2's web-based framework has reduced the barrier to entry for both system administrators at new installation sites and regular users wishing to interact with content on the SRSD. SAGE2 is cross-platform with very few dependencies. Additionally, the display clients can be running on any machine making it possible to create a large tiled display wall using multiple machines that are not clustered together. Therefore system administrators can install SAGE2 without hours of training or technical expertise, and they do not need to be fluent in maintaining a Linux cluster. Users can access a SAGE2 session and interact with content by simply visiting a URL in their web-browser. This eliminates the need for each user to download and install a specialized client application. The web-browser interface also provides users with a familiar environment that is accessible on any device.

2.3.1 Imitate Co-located Collaboration

It is often the case that a team is distributed at two or more physical locations, but wishes to work as though they were co-located. In this scenario, everyone is working on the same problem and looking at the same data. SAGE2 helps bridge the gap of physical distance by mirroring shared content at all sites, including live video and audio streams from commercial videoconferencing software. To illustrate how SAGE2 is used in such a situation, we will describe how it could be used during the judging of a research based photo competition conducted at a university. The panel of judges is distributed across two or three campuses and wishes to rank all submissions during a single joint session.

A SAGE2 session is started at one site, and the SRSD is mirrored at all other sites by simply visiting the same URL. All images along with their accompanying research description are uploaded to the SAGE2 Server and displayed on each SRSD. Additionally, the multi-user notepad application is launched, allowing any judge to jot down thoughts or comments at any time. During the first phase of judging, each image and its accompanying description are enlarged to full resolution and viewed one at a time. After a brief discussion, the image is placed in one of three groups – top contender, possible contender, not a contender. Once all images have been discussed, the second phase of judging begins. This stage, depicted in Figure 4, is more freeform with all judges working simultaneously. Spatial orientation of the images is used to compare and rank the images. A better image is moved up higher on the display – if an image is only a few pixels higher than another it is considered just barely of higher quality, whereas if it is positioned many pixels higher it is considered significantly superior. Any judge at any site can rearrange the images, add comments to the notepad, or communicate with the other judges. This process continues until a consensus is reached and the winners are determined.

Figure 4. Staged photograph illustrating judges at two separate campuses ranking research images for a competition. The content is mirrored at both locations and changes to one site are reflected in the other in real-time.

SAGE2 provides a number of unique features that enable this type of collaboration. Since SAGE2 provides a windowing environment on an ultra high-resolution SRSD, the judges are able to view multiple high-resolution images simultaneously. The multi-user paradigm allows multiple judges to reorder the windows or add comments to the notepad simultaneously, reducing completion time for the task. The videoconference allows judges at different locations to see and talk to each other as if they were present in the same room. This is incredibly important due to the fact that images are submitted from a variety of domains, and each judge has a unique expertise. Since the SRSD is mirrored at all locations, when a judge at one location interacts with the content, all other sites immediately reflect the modification. This gives collaborators at all sites a shared context from which to communicate.

2.3.2 Remote Collaborators Working on Different Pieces of a Related Problem

Another common scenario is when distributed teams work with the same or similar data and want to share related data and discoveries. In this case, each team may have a unique configuration for their SRSD, and not all documents and applications need to be present at each location. To illustrate how SAGE2 is used in this situation, we will describe how it could be used for disaster management planning in the city of Chicago.

Disaster management planning requires multiple groups to work together. In this example, we will illustrate the collaboration between Chicago city officials and researchers at Argonne National Lab. The city officials are responsible for crisis management and direct the disaster management planning. They are familiar with the city, its infrastructure, and its population. The researchers utilize their supercomputing facilities to simulate and visualize various disasters. Each team uses a wide array of content to analyze their data including interactive maps, PDFs, images, charts, and graphs.

Members of each team simultaneously use a multitude of applications to answer relevant questions. Panel A in Figure 5 shows city officials reviewing statistics, such as most densely populated area vs. day of week and number of public trains running vs. time of day. Panel B in Figure 5 shows the researchers running various disaster simulations, such as a tornado or a train derailment, and creating advanced visualizations. Both teams plot data on a map to see the spatial distribution and overall impact of these disasters. Once the city officials have determined a new scenario or one of the researchers have made an interesting discovery, they share their finding with the other team. Each team has independent control over all documents and applications on their SRSD. This allows important information to be shared between sites, while keeping less relevant data private.

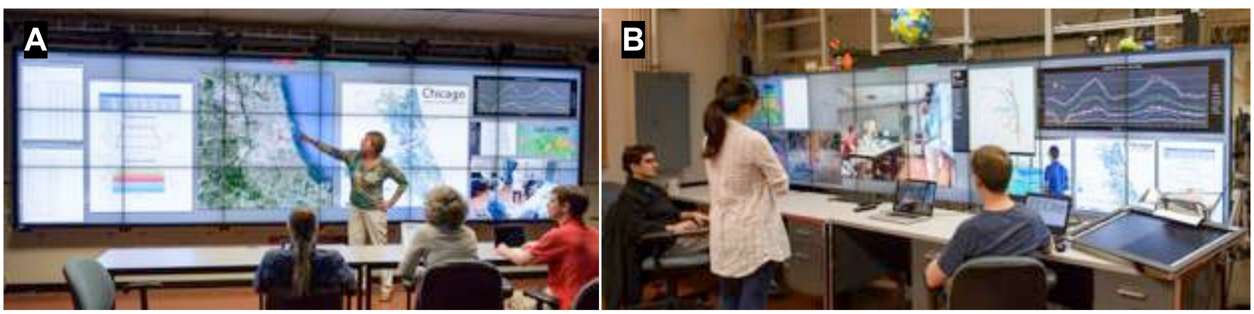


Figure 5. Staged photograph illustrating Chicago city officials and researchers at Argonne National Lab collaborating by sharing relevant information while maintaining separate focuses. Content is controlled independently, but can be shared across remote sites.

SAGE2 allows these teams to view and interact with various types of data simultaneously, such as interactive maps, PDFs, and images. More than one person at a given site can interact with any application simultaneously, which increases productivity and removes the need to switch who is in control. Each location can arrange the documents and applications independently in order to create a layout that is most effective for their team. Also, not all documents and applications are shared between sites, which helps reduce visual clutter, and ensures that only relevant content is available at each site.

2.3.3 Remote Single User Joining a Co-located Session

SAGE2 can also be leveraged when a single member of a team is traveling or working from home. In this case, everyone is working on the same problem and looking at the same data. The major difference between this scenario and the first is that the single remote member does not have access to a SRSD, but rather only has his/her laptop or tablet. To illustrate how SAGE2 works in this situation, we will describe how it could be used by a team collectively writing a co-authored publication.

A team of five is outlining a paper and reviewing previously published works. One of the members is attending a three-day conference, so the team schedules its meetings during lunch breaks. The other four members start a SAGE2 session and share numerous PDFs, a multi-user notepad, and the lead author's laptop screen. The remote member joins the session from her laptop, which gives an overview of the content and layout on the SRSD. She can get details of any specific shared application on demand by viewing it in a separate browser tab. Figure 6 depicts the team reading PDFs and taking notes on the related work, while the lead author integrates relevant content into a draft of the paper on his laptop. When the meeting concludes, the session can be saved, so the team can resume later without having to re-upload and reposition all relevant documents and applications.

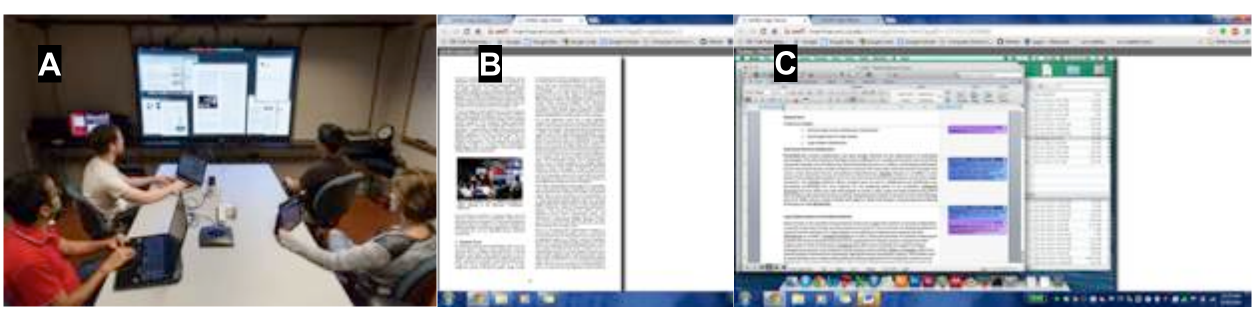


Figure 6. Panel A shows a team of authors collaboratively writing a paper. A remote single user is able to participate by viewing applications, such as PDF viewers and a shared laptop screen in separate web-browser tabs (Panels B and C).

SAGE2 allows a remote user to engage in a remote collaboration session without the need for special hardware or software. A browser allows the remote user to view and interact with shared content via an overview of the whole SRSD and detailed applications viewable one at a time. The co-located team and the remote user both receive immediate input from each other. SAGE2 allows both the co-located team and the remote individual to take advantage of the technology at their disposal, rather than forcing a group to fall back to the lowest common denominator. SAGE2 also allows the team to continue their work at a later time by saving the session.

CHAPTER 3

RELATED WORK

Collaboration is predicated on the idea that two or more people working together can achieve an outcome superior to what each individual could accomplish on their own. This thought falls in line with Gestalt psychology, which is known for the phrase *the whole is other than the sum of its parts*. People have grouped together and formed collaborations since the dawn of history to help society and pursue knowledge. Technological advances have long been a catalyst for collaborative productivity. The printing press enabled knowledge to be widely distributed and the telegraph allowed communication across vast distances. With the advent of computers in the middle of the 20th century, technology assisted collaboration has been developed and adopted at ever-increasing rates. The benefits of collaboration and desire for groups to work together have remained virtually the same throughout time; technology has just been a factor to improve the distance and effectiveness of those collaborations. The rest of this section will take an in depth look at collaboration software and hardware, user studies on collaborative systems, and surveys of user needs with collaboration software.

3.1 Collaborative Software and Hardware

Douglas Engelbart is considered the Father of Groupware for first envisioning collaborative computing in the early 1960's [50]. He developed the oN-Line System (NLS), which introduced many modern computing concepts such as the mouse and screen windowing [51]. In addition, the NLS had interactive on-line abilities to support more than one user with the ability to edit the same documents. In general, groupware can be designed to facilitate collaboration in three distinct ways: through communication, conferencing, or coordination [52]. Communication is an unstructured exchange of information, such as phone calls or instant messaging chats. Conferencing represents multiple users interacting while working on a common goal, such as conducting a brainstorming session. Coordination refers to separate but interdependent work between multiple users, such as developing a project timeline and budget.

3.1.1 Shared Workspaces

Shared workspaces, such as a classroom, laboratory, or conference room, have long been used for collaboration. Computing resources, groupware, and large displays have been integrated into these spaces in order to facilitate collaboration. The technology described in this section, as well as some of the multi-user large display systems mentioned in Chapter 2, and their ability to facilitate collaboration is summarized in Table III.

TABLE III

Comparison of groupware, highlighting which aspects of collaboration each addresses. While both Mezzanine and SAGE facilitate communication, conferencing, and coordination across distance, they only address specific constrained types of collaboration.

	Communication	**Conferencing**	**Coordination**	**Distance**
Google Hangouts	X	X		X
Skype	X	X		X
WebEx	X	X		X
Liveboard	X	X		X
LACOME		X		
CubIT	X			
IMPROMPTU	X	X	X	
Mezzanine	X	X	X	X
CollaBoard	X			X
ShareTable	X		X	X
SAGE	X	X	X	X
DisplayCluster	X	X	X	

In 1992, Elrod et al. [53] introduced Liveboard, which represents some of the earliest research on large shared displays in a collaboration space. A wireless pen is used as an interaction device in order to allow multiple users to interact with the system. Liveboard primarily combines the capabilities of a whiteboard and a slide show presentation tool. Liveboard also has networking capabilities to allow for remote collaboration. This feature was primarily used to collaboratively write or draw in a shared document. Liveboard helps facilitate communication and conferencing with its ability to present information in a shared space and network with remote sites, but lacks the ability for multiple users to coordinate on a combined task.

MacKenzie et al. [54] present LACOME, the Large Collaborative Meeting Environment. LACOME allows multiple users to publish their desktop screen to a large shared display. Users can user their keyboard and mouse to manipulate content on the large shared display. LACOME supports different interaction modes to allow for window manipulation at any position of a window rather than just its title-bar and borders. Other modes allow for content interaction and ‘snapping regions’, which automatically move the pointer to the boundary of a window. LACOME facilitates conferencing by allowing users to share information from their desktops, but does not provide infrastructure for enhancing communication or task coordination.

Rittenbruch et al. [55] present CubIT, a large-scale presentation and collaboration framework. CubIT allows users to upload and interact with media content in a public space. A web-based interface allows users to upload images and videos from personal devices. Users interact with content on the large shared display through a direct multi-touch interface. Users are tracked using RFID tags, which are provided to each student and employee where CubIT is located. This allows touch inputs to be associated with a given user. CubIT supports communication by allowing users to share content directly with another user. However, it does not facilitate conferencing or task coordination as it is a presentation framework for a public space.

Biehl et al. [56] introduced collaboration software, IMPROMPTU, for opportunistic group work using multiple display environments. The authors highlight the need for software to enhance collaboration in authentic real-world settings. Their software allows users to *show* or *share* off-the-shelf applications with other users or with a shared large display. Showing an application streams the pixel data, whereas sharing an application allows other users to interact with it. The authors conducted a real-world field test at Microsoft, deploying their system with two separate development teams. One team all worked in the same room with individual computers as well as a large shared projector. The other team worked in individual cubicles all in close proximity to one another, and they had a conference room with a large shared projector. Both teams were observed for three weeks while their behaviors and use of IMPROMPTU was recorded. The authors concluded that their software was useful from the results of their observations and from participant surveys. In particular, the results showed that teams leveraged the framework for short-term opportunistic collaborations throughout the study. IMPROMTPU is a system that helped facilitate communication when sharing information with another individual, conferencing when sharing information with a large shared display, and coordination when cooperatively working in a shared application. However, IMPROMPTU was only designed to facilitate collaboration with a co-located team.

Oblong Industries, Inc. has created a high-tech collaboration room – Mezzanine [4]. Mezzanine is a room with 3 tiled displays in the front of the room and three more displays on one side. The room also has two video cameras, one pointing into the room to use for video conferencing, and one pointed at a whiteboard to digitize non-digital assets. Two wands are also present to interact with content on the displays. Users can connect to Mezzanine with their laptop or mobile device and share their screen. Multiple Mezzanine sites can connect to enable remote collaboration. Mezzanine enhances standard telepresence systems found in many businesses by adding InfoPresence™ – multiple streams of information in a collaborative environment. This system is based on live video feeds and harnesses users' local devices to create and render content. However, content appears to be shared, since the reachthrough feature allows any users to interact with another computer (e.g. clicking "next slide" in a PowerPoint that is on somebody else's device). Mezzanine helps facilitate communication, conferencing, and coordination for both co-located teams and remote collaborators. However, Mezzanine limits the type of data used during a collaboration setting and requires a custom-built conference room at all participating sites.

3.1.2 User Interaction on Large Displays

User interaction on large shared displays has been another area of research critical enhancing collaboration. Cheng et al. [57] presents an interaction method that allows multiple collaborators to simultaneously view and interact with content on a large shared display by using tablets. Any collaborator can use their tablet to create a focused viewport of the shared display. Standard touch gestures pan and zoom this focused view so that a user is able to reach all areas of the large shared display without physically moving. A double-tap brings the whole shared display in miniature onto the tablet. It also shows bounding rectangles for all collaborators focused views. A user can double-tap any bounding rectangle to enter that collaborator's view. Interaction techniques like this that utilize mobile devices can help facilitate task coordination by making it easy for any collaborator to engage with and navigate through relevant data.

Müller-Tomfelde et al. [58] introduced an interaction technique for large displays, called Pseudo Direct Touch (PDT). This approach decoupled a touch screen from the display surface. Instead a transparent touch frame is placed at a distance from the display wall. This enables users to use a touch interface without needing to physically navigate across a large display or blocking content from other collaborators. A ray is cast from the user's eyes through the touch point. This acts like a direct touch interaction at the position on the large display where the ray intersects. The authors conducted a user study to test accuracy of this method and discovered that providing visual feedback of where the PDT occurred significantly improved performance. Interaction techniques such as this can help facilitate communication and conferencing. Collaborators can easily navigate and point to relevant data to ask questions or share findings with the team.

Ponto et al. [59] created an extension to CGLX that incorporates separate multi-touch devices to be used for interaction and visual feedback. Input events are synchronized with the display environment and scene information is streamed to the mobile device. Their approach scales to handle a large number of external devices in order to support collaborative analysis. Supporting multiple interaction devices from multiple simultaneous users helps enable collaborative conferencing, since each collaborator can interact and participate in the discussion.

3.1.3 Web-based Collaboration

Remote collaboration has been greatly affected by the advancement of web-based technologies. The web is being increasingly used to collaborate in a variety of scenarios such as teaching, corporate meetings, and manufacturing. Google Hangouts, Skype, and WebEx [1-3] are examples of successful commercial products designed to enable remote collaboration. Collaborators can utilize personal devices to communicate and work on shared applications with each other. However, groupware designed for a single PC does not support scalable visualizations or viewing multiple data-intensive applications simultaneously.

Going beyond web-based applications that enable remote collaboration, Lowet et al. [60] synchronizes full web-browsers across distance. The authors implement two browser synchronization techniques in order to turn single-user web applications into a multi-user experience. Using JavaScript and AJAX, they implemented *input synchronization* and *output synchronization* as two methods of co-browsing. Input synchronization injects all user input events recorded from one browser into all other browsers. Output synchronization listens to changes in the DOM tree of a reference browser. Changes are sent to each other browser, which update their DOM tree in order to remain synchronized. This provides a synchronized view of data across remote sites, which can facilitate conferencing and coordinating tasks.

3.1.4 Collaborative Hardware

Customized hardware is also being developed to better support remote collaboration. Arroyo et al. [61] augmented a small tabletop multi-touch screen with a second screen for videoconferencing in order to allow groups in two different locations to collaboratively explore museum exhibits. Two museums in different cities each share a portion of the same art collection. Users collaboratively play a game to explore the entire collection that is broken into three stages – an initial stage of negotiation followed by two stages of completing collaborative tasks. The game typically lasted 10 minutes with both sites viewing mirrored content and required users at each end to communicate and coordinate with one another. The authors observed that users initially exhibited a turn-taking behavior, but gradually utilized more simultaneous and independent interactions.

Kuechler et al. [62] present CollaBoard, a remote video and data conferencing system. Two people in different locations work together with a large monitor and camera. The display shows the data from a shared application with a superimposed life-sized video of the remote collaborator. Simultaneous interaction enables collaborative work on the shared application. This aims to imitate co-located collaboration by allowing users located at different sites to observe the gestures and expressions of each other. CollaBoard enabled communication over distance that simulated co-located communication, both verbal and visual.

Yarosh et al. [63] presents the ShareTable, which combines videoconferencing with a shared tabletop task space. The shared tabletop space uses a projector and a video camera in order to overlay real objects from both locations. This was deployed in four divorced households in order to enable the remote parent to interact with their child. The ShareTable allowed parents and children to engage in shared activities such as collectively drawing a picture. Surveys results showed that ShareTable was preferred to phone calls and easier to use than videoconferencing alone since it enabled the use of tangible objects. ShareTable enhanced digital communication and enabled coordinated tasks to be accomplished by allowing tangible object to be overlaid with projected images.

3.2 User Studies on Collaborative Systems

Afforded data and interaction techniques have been evaluated to determine their importance in collaborative situations. Recurring findings suggest that awareness of collaborators and distributing work between collaborators are two of the most important factors to enhancing communication, conferencing, and coordination. Systems that provide feedback on what each user is doing, allow multiple users to interact simultaneously in a variety of ways, and aid in sharing information between individuals have enhanced collaboration.

3.2.1 Identifying Factors that Impact Collaboration

Ocker et al. [64] studied how to work effectively in Partially Distributed Teams (PDTs). A PDT is made up of two or more subteams in different physical locations. Their study identified six key factors for successful long-term collaboration of a PDT – shared identification, trust, awareness, coordination, competence, and conflict. Some of these features are inherent in the individuals, such as trust and competence, though groupware systems can help improve them over a period of time. Awareness and coordination were identified as procedural aspects of team management, and therefore are imperative for successful groupware to address.

Kim et al. [65] studied how user interaction on a large multi-touch wall affected collaboration. Their results show that both asynchronous access and multiple types of input (eg. touch, keyboard, mouse) help support group work. This also increased verbal communication and attention within a group. Allowing multiple users to interact simultaneously with a variety of devices allowed collaborators to coordinate their efforts with the most effective input device.

Yamashita et al. [66] conducted a study on remote collaboration between small groups at a tabletop display. The authors study the effect of showing the remote collaborators' upper bodies on projected walls that surround the tabletop display. The participants of the study were instructed to assemble a toy rail kit. Two participants memorized a map of the proper rail assembly. The other two participants listened to instructions given by the first two in order to physically assemble the rail. Results show that viewing the upper body of remote collaborators significantly decreased task completion time. The authors believe a major factor in the success of their experiment was the fact that all participants were allowed to freely move around the tabletop display, rather than being constrained to a fixed location. This provided greater flexibility and persistent awareness of remote collaborators actions.

3.2.2 Evaluating Usefulness of Technological Features

Different aspects of remote collaboration systems have been evaluated in order to provide insight on their usefulness. Balakrishnan et al. [67] test the usefulness of visualizations in complex tasks involving remote collaborators. Four conditions were tested – spreadsheet only (control), individual unshared visualizations, view-only shared visualizations, and full-access shared visualizations. Two users investigated homicide data in attempt to identify a serial killer. Each user had a portion of the relevant data needed to properly complete the task. The two users were on separate computers and communicated via instant messenger. There user study showed that visualizations improved collaboration for complex tasks, with full-access shared visualizations achieving the highest success rate. 7.7% of spreadsheet only, 50.0% of individual unshared visualizations, 33.3% of view-only shared visualizations, and 58.0% of full-access shared visualizations correctly identified the serial killer. This demonstrates how visual applications can aid the collaborative process when dealing with data-intensive content.

Epps et al. [68] studied awareness of both co-located and remote collaborators. Two users were at a shared tabletop display and a third user used a machine in a nearby room. All users had their own pointer to use on a shared application. The users played a game under four conditions: all three users worked cooperatively, the co-located users competed against the remote user, one user at the tabletop and the remote user competed against the other user at the tabletop, and all three users competed against each other. Users could see and hear each other, using a videoconference for the remote user. Over all four conditions, pointers on the shared application provided more awareness of the other users than the videoconference. Awareness decreased across the four conditions showing that users are more aware of each other when working cooperatively. This study demonstrates the importance of providing awareness of remote collaborators beyond simply using existing videoconferencing tools. Synchronized interactions in a shared application aid collaborative tasks, especially when team members are working cooperatively.

3.3 Surveys of User Needs

Even with modern technology and research on user behavior, no groupware has solved all user needs. In the book Collaborative Web Search [69], Morris and Teevan examine the 'who', 'what', 'where', 'when', and 'why' of collaboration. They report that the 'who' may be symmetric (each person tries to accomplish the same task) or asymmetric (overall task is divided and each person takes on a specific role). The 'what' for remote collaboration is either to consult (ask for suggestion), brainstorm (generate ideas), or reference (find ready-prepared material from others). The 'where' illustrates that remote awareness is key as it removes the necessity to explicitly ask other group members routine information. Also, division of labor can help reduce redundancy of tasks. The 'when' can either be synchronous (same time) or asynchronous (different times). The 'why' is because people need to find an expert to help with a specific problem. This helps elucidate the problem space to which this dissertation plans to make its contribution. This dissertation will study asymmetric synchronous collaboration scenarios for interdisciplinary coordinated work across distance.

Evans and Chi [70] identified tactics used before during and after a collaborative session. Context framing and requirement refinement were done before the collaborative session, foraging and sensemaking were done during the collaborative session, and organization and distribution of information was done after the collaborative session. Therefore distributing information foraging and improving collective sensemaking are ideal outcomes for collaboration systems.

Rama and Bishop [71] surveyed groupware to discover commonalities prevalent in most systems. They discovered that groupware typically works with a closed group of users, makes users aware of other collaborators, and focuses on collaboration not just document sharing.

This echoes the studies that show the importance of remote collaborator awareness, while also showing the need to go beyond document sharing to reach a group's collaborative potential.

Dyck et al. [72] provides suggestions for groupware development based on advances in the gaming industry. Limiting bandwidth through compression, aggregation of information, and ensuring all messages are useful to the receiver are techniques used in online games that could help improve latency issues in groupware. Degrading gracefully is another issue addressed by gaming, where applications are designed for handling slower or interrupted connections. Finally the authors suggest that like gaming software, groupware should handle adaptive window sizes to effectively make use of the provided screen real estate. Among other things, this work suggests the need for collaborative systems to support scalable visualizations in order to effectively use screen the real estate of its display.

White and Lutters [73] study cross-organizational partnerships to elucidate challenges in designing interdisciplinary groupware. Their findings show the importance of providing personal profiles for each collaborator and site characteristics to provide context when communicating with a collaborator. They also note that trust is an important aspect for successful collaboration. Cross-organizational groupware should incorporate some trust-building features into its design.

Hornecker et al. [74] study the effects of entry and access points for group interaction. Entry points are design characteristics that invite users to engage with the system. Access points are features that allow users to interact with the system. Entry points are often used to attract users to engage with access points. Their findings suggest that increasing the number of access points in groupware may reduce the dominance of a single user and facilitate more distributed collaboration.

Dimitracopoulou [75] provides a future research agenda based on current trends of collaborative systems. The author highlights four main goals for future collaborative systems. First, they should unify collaborative features currently distributed amongst many software packages and research projects. Second, they should provide analysis features for interactions and dialogs to gain insight on the types and quality of collaboration. Third, they should be flexible to be used in a variety of environments. Lastly, they should support ubiquitous computing devices. The last two goals help collaborative systems enhance group communication, conferencing, and cooperation by improving the ease of access and allowing all collaborators to participate.

Based on the collaborative systems that have previously been developed, the studies on effective groupware, and surveys of user needs, I formed my research goal to enhance communication, conferencing, and collaboration across distance between PDTs. I implemented multiple potential solutions that incorporate improving awareness of remote collaborators and go beyond document sharing to allow simultaneous interaction into synchronized applications. The next section describes these solutions as well as a comprehensive study to evaluate their effects on collaboration between PDTs.

CHAPTER 4

DATA SYNCHRONIZATION FOR HETEROGENEOUS DISPLAYS

Often complex problems require the combined efforts of multiple teams across varying disciplines. These authentic collaborative scenarios would require a flexible system capable of handling large volumes of data that can be customized to fit each domain's need. While custom built conference rooms, such as Mezzanine, may work for collaboration within a single organization, it is unlikely to be effective across varied groups, since each discipline has its own unique space and technology. A scalable software that is compatible with any SRSD and underlying operating system would allow any organization or research team to utilize their existing tools, rather than forcing all collaborative parties to agree to purchase and use a single technology. In order to facilitate remote collaboration, software should also reduce networking constraints such as limiting the number of other sites with which it can connect.

It is often the case that a team is distributed at two or more physical locations, but wishes to work as though they were co-located. In this scenario, everyone is working on the same problem and looking at the same data. Software that can mirror content at all sites, including live video and audio streams would help bridge the gap of physical distance. The addition of videoconferencing would allow users at all sites to view all collaborators in order to aid in awareness of the entire group.

Another common scenario is when PDTs work with the same or similar data and want to share related data and discoveries. In this case, each team may have a unique configuration for their SRSD, and not all documents and applications need to be present at each location. A portion of the data is shared between sites, while other content can remain private to each participating team. Evaluating the communication, conferencing, and coordination between groups using *data-pushing*, *data-duplication*, and *advanced data-synchronization options* in collaborations of this second scenario is my main dissertation contribution.

4.1 Technical Design of Infrastructure

The work previously done on SAGE2 made it a powerful and flexible system to better enable co-located and remote collaboration. However, there was still a need to develop and evaluate more advanced means of supporting groups working together across distance. In order to mirror content and have all interactions synchronized between multiple sites, SAGE2 required matching hardware configurations at all locations. For collaboration between heterogeneous SRSDs where groups are working on separate but related pieces of a problem, collaborators were limited to *data-pushing*, sending unsynchronized applications between locations. Multiple sites could not simultaneously interact with a shared application in this situation; each site could only interact with their own local copy. For example, if one site pushed a PDF to another site, the same document would be shown on each SRSD. However, the position and size of the document and which page of the PDF was being viewed would be completely independent

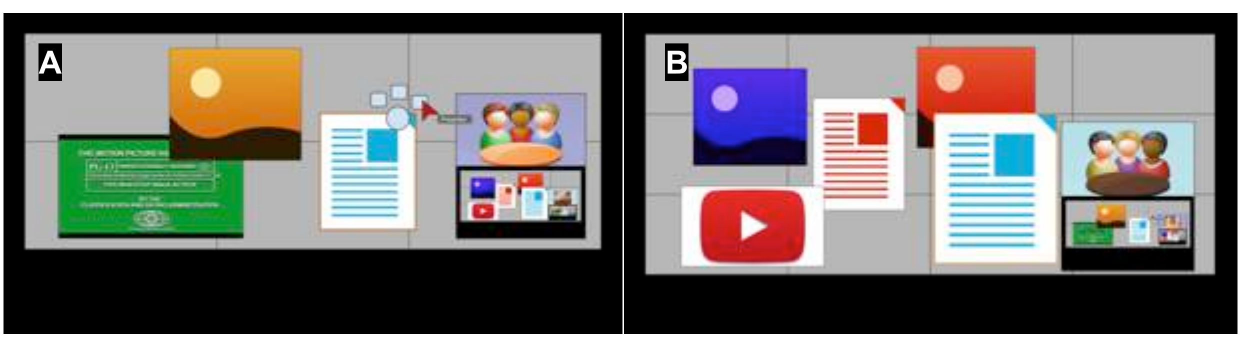


Figure 7. Illustration of data-pushing. Panel A shows one site, which has an assortment of content on its SRSD. A user pushes the blue document to the remote site. Panel B shows the second site, which has a separate assortment of content on its SRSD plus the blue document pushed from the first site.

between the two sites. In order to view what is happening at the remote site, I have supplemented this scenario by sharing two videoconferencing windows, using Google Hangouts, one camera showing the collaborators and one showing their SRSD (Figure 7).

I planned to improve coordination and user awareness by allowing users at different sites to collaborate on a shared application. I developed two new techniques to handle shared applications across remote SAGE2 sites.

4.1.1 Data-Duplication

The first new technique I developed was *data-duplication*, where one section of the SRSD acts as a shared portal that contains the fully synchronized version of all data-conferencing content and a second portion of the SRSD contains local unsynchronized copies of the data-conferencing content. The overall screen space is partitioned for each remote site in a current collaboration session. At any time, any user can move an application from the local space to a remote partition in order to share the application. The application will appear in the corresponding partition on the remote site's SRSD. Users at both locations are able to interact

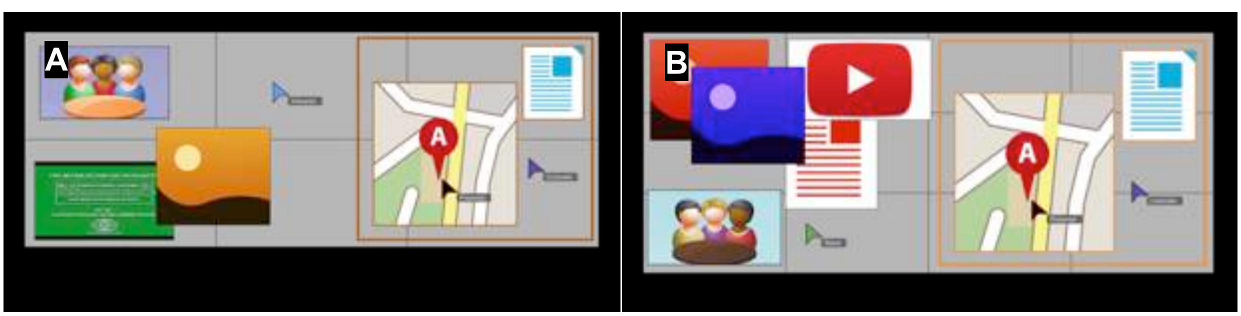


Figure 8. Illustration of data-duplication. Panel A shows one site, with the orange box on the right depicting the shared area with the remote site and the area outside the orange box depicting local applications. Panel B shows the second site, which also has an SRSD split into a local and shared partition. The shared application partition shows the same applications, with the same relative positions and sizes as the first site.

with the shared application. Both window management actions, such as window movement and resize, and interactions inside the application are fully synchronized across sites. Additionally, interaction icons for pointers and touch events in the shared partition will be displayed at both locations (Figure 8). This will gives appearance of both groups working on the same application simultaneously, and was designed to improve remote collaborator awareness.

Unlike the completely synchronized SRSDs that were described in section 2.3.1, this technique allows for both a shared and a private space for data. Additionally, remotely located SRSDs do not need to be homogenous. Any two SRSDs running SAGE2 could create a shared data portal regardless of size, resolution, or aspect ratio. Additionally, the shared portal can be freely moved and resized on each SRSD independently – only relative positions and sizes of applications inside the portal remain constant. This allows each SRSD to scale the shared portal to a size that is best for their technology.

Since applications are executing separately at each site, the application's state, which defines all properties necessary to fully reproduce the application, is observed at all sites. Any time an aspect of the state is modified, its value is streamed to each remote application. This ensures that the state of shared applications at all sites remain the same. Since each site has its own instance of the application, other aspects of the application can vary between sites, such frame rate. This gives users the appearance that they are interacting with the same application, while optimizing performance for the local hardware.

In order to accomplish this technique, I developed a data-duplication portal that is visible on both remote SRSDs. The portal is a rectangular area with a shared coordinate system based on the average size of the collaborating SRSDs. A unique matrix transform is applied in order to position and scale the portal on each SRSD, thereby allowing each group to place and size the portal for what best fits their needs. All pointers from the remote SRSD are added to the local SRSD so that collaborators can view both local and remote pointers within the portal. Whenever a pointer enters the portal, a message is sent to the remote site in order to make it visible, and whenever a pointer exits the portal, another message is sent to the remote site in order to make it invisible. Therefore remote pointers only appear within the shared area and do not interfere with the private content on the rest of the SRSD. Figure 9 depicts a *data-duplication* session with a shared portal on two heterogeneous SRSDs.

Figure 9. SAGE2 remote collaboration using the data-duplication technique. The yellow highlight in each photo shows the shared portal on each SRSD with synchronized applications and pointers. Panel A shows a team using a wide aspect ratio shared display. Panel B shows a team using a tall aspect ratio shared display.

4.1.2 Advanced Data-Synchronization Options

The second new technique I developed uses *advanced data-synchronization options*, where collaborators can choose which aspects of each shared application will be synchronized and which will be controlled independently. Users are given access to independent variables within an application's state and are able to choose whether or not to synchronize each individual property when sharing with a remote site (Figure 10). This technique was designed to allow teams at each location to coordinate their efforts based on individual needs versus whole group needs. For example, when watching a video across multiple sites it may be beneficial to synchronize the location of the play head and whether the video is playing or paused. However, different sites may wish to have independent control on the volume level. In order to keep users from manually selecting whether to synchronize each individual property, the state is ordered hierarchically. For example, the camera position in a 3D model viewing application can be the parent of x, y, and z positions. Therefore, to synchronize the camera position, a user would simply need to make one selection rather than three.

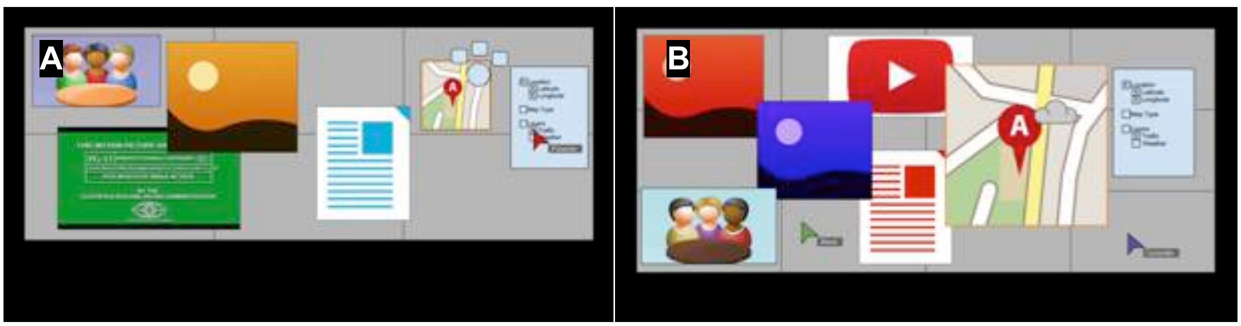


Figure 10. Illustration of using advanced data-synchronization options. Panel A shows one site, where a user is sharing a mapping application and selecting which properties are synchronized. Panel B shows the second site, which receives the shared application and displays the properties that are synchronized, such as latitude and longitude, and which are not, such as weather overlay.

Similar to *data-duplication*, applications will be executing separately at each site. Any time a synchronized aspect of the state is modified, its value will be streamed to each remote application. Since each site has its own instance of the application, unsynchronized aspects of the state never get streamed to the remote instances. This gives users the appearance that they are interacting with a shared application that allows the freedom to control certain properties locally.

In order to accomplish this technique, I have developed a graphical user interface that overlays a shared application depicting its state. Aspects of the application state are organized hierarchically. Properties that are not synchronized are visually grayed out, whereas synchronized properties are visualized in full color (Figure 11). Users can display or hide this interface and select or unselect any property at any time. Whenever a user at either site synchronizes or unsynchronizes a property, this option is streamed to the other site so that changes to this property are handled properly.

Figure 11. SAGE2 remote collaboration using the advanced data-synchronization options technique. Panel A shows the hierarchical application state – mapType is unsynchronized, whereas all other properties are synchronized. Panel B shows the shared application on SRSD with a mapType showing satellite imagery. Panel C shows the shared application on a second SRSD with a mapType showing road maps.

4.2 Hypothesis

Due to the previous research on remote user awareness with shared applications, local groupware, and large-scale visualizations and ultra high-resolution displays, I formed three hypotheses surrounding my main research questions about the data-conferencing techniques I had developed.

1) Both *data-duplication* and using *advanced data-synchronization options* would result in significantly improved collaboration compared to *data-pushing*.
2) Co-located sessions would produce the best collaboration (though perhaps not significantly better than *data-duplication* and using *advanced data-synchronization options*)
3) Using *advanced data-synchronization options* would be preferable to *data-duplication* for groups using a smaller SRSD due to the fact that limited screen space must be partitioned in the latter case.

4.3 Evaluation

In order to evaluate the effectiveness of these new remote collaboration techniques, I conducted an informal longitudinal study and a formal user study, both involving two distributed groups. In both studies one group used a large shared display with an ultra wide aspect ratio (around 25 Mpixels with a 16:3 aspect ratio) and the other group used a small shared display with a standard widescreen aspect ratio (around 8 Mpixels with a 16:9 aspect ratio).

4.3.1 Longitudinal User Study

For the informal longitudinal study I monitored meetings regarding the continued development of SAGE2 between the University of Illinois at Chicago and the University of Hawaiʻi at Mānoa for 14 weeks. During these meetings, members of the SAGE2 development team used SAGE2 to give progress reports and demonstrate new applications or improvements to the core features of SAGE2. Each meeting typically lasted about one hour. Participants would use their personal laptop to connect to the SAGE2 interaction client and load new content or interact with shared content on the SRSD. Since participants used their own laptops, constraints about the operating system and web browser were not enforced. Participants used Mac OS X, Windows, and Ubuntu and ran the SAGE2 interaction client in either Chrome or Firefox.

4.3.1.1 Longitudinal Study Data

The focus of the longitudinal study was to see how groups worked together across distance over time in an authentic work environment given a different set of tools for sharing data. In order to evaluate the effectiveness of collaboration, I decided to monitor participants' perception of collaboration by conducting surveys, and measure the quality of collaboration by cataloguing each session with audio / video recordings and user interaction logs.

The surveys asked participants to score three different aspects relating to the collaborative session on a scale of 1 to 10. First, participants were asked how easy the system was to use. This question was designed to measure how intuitive each data sharing technique was, and whether or not they became easier to use after gaining experience. Second, participants were asked how successful the system was at facilitating collaboration. This question was designed to measure participants' perceptions of the quality of collaboration enabled by each data sharing technique. Finally, participants were asked how much they liked the remote collaboration features. This question was designed to gain insight on how users felt most comfortable collaborating with a remote group. The first two questions were asked about both the local collaboration and the remote collaboration. The final question only pertained to the remote collaboration techniques.

The videos were recorded in order to be coded into collaboration modes that each group was engaged in. Collaboration is broken up into four distinct modes: not collaborating, communicating, conferencing, or coordinating. Watching the videos from each site would provide insight into what percentage of time the groups spent in each collaboration mode. All participant interaction with SAGE2 was logged in order to create a timeline and look for recurring interaction patterns that existed for each collaboration mode. The logs consisted of data for each participant on when they enabled/disabled their pointer, shared their screen, uploaded a new document, opened an application, moved/resized a window, interacted inside an application, and shared an application with the remote site.

4.3.1.2 Longitudinal Study Methods

Participants used the *data-pushing* technique for weeks 1-4, the *data-duplication* technique for weeks 5-8, the *advanced data-synchronization options* technique for weeks 9-12, and their choice for the final two weeks. I recorded window management interaction, and application interaction within SAGE2. I also conducted video and audio recordings of each session in order to code collaboration as either communication, conferencing, or coordination. After each study session, I administered a survey to all team members in both locations to determine how they perceived the usefulness of various collaborative features. Figure 12 depicts the setup of teams and displays at both the University of Illinois at Chicago and the University of Hawai'i at Mānoa.

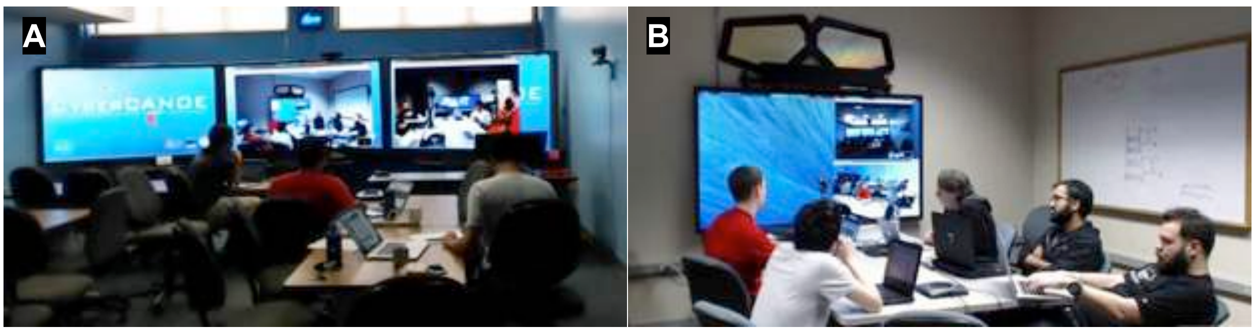


Figure 12. SAGE2 Development Team during their regular weekly meeting. Panel A shows the team at the University of Hawai'i at Mānoa using a large shared display. Panel B shows the team at the University of Illinois at Chicago using a small shared display.

The procedure for each session of the longitudinal user study was as follows:

- All subjects from the University of Illinois at Chicago gathered in a conference room in front of a smaller shared display – a 1.05m x 1.86m display with a 3840 x 2160 resolution (approximately 8 Mpixels). Subjects were seated approximately 1.0m - 2.5m from the display.

- All subjects from the University of Hawai'i gathered in a conference room in front of a larger shared display – a 1.05m x 5.58m display with a 11520 x 2160 resolution (approximately 25 Mpixels). Subjects were seated approximately 1.2m - 9.0m from the display.

- Google Hangouts was used for videoconferencing with two videos present on each site's shared display: one of the remote group (showing the people), and one of the remote group's shared display.

- Each subject had a personal device (laptop) to use during the study.

- Subjects took part in a meeting to discuss project updates, review code, test applications, etc.

  - Each subject connected to the interaction client of the SAGE2 software running on the local shared display using his or her personal device.
  - Subjects could upload documents (such as images, videos, and PDFs) from their personal device to the shared display.
  - Subjects could put a personalized mouse pointer on the shared display with a chosen identifier (nickname) and an assigned unique color.
  - Subjects could share content on their local shared display with the remote site's shared display.
  - All interaction with the shared display was logged and saved to a file.
- At the completion of the meeting, all subjects took a brief survey about their experience (using pen and paper). Surveys from the University of Illinois at Chicago were physically collected for analysis. Surveys from the University of Hawai'i were digitally scanned for analysis.

4.3.2 Formal User Study

For the formal user study, two groups of two participants represented teams from different domains that had a different set of knowledge, each required to achieve a common search and analysis goal. The fictional problem users were asked to solve was to find an ideal location to open a new coffee shop within a given city. The problem was split into two tasks – first, to come up with 2-4 potential locations based on a separate set of constraints for each team, and second, to determine the best location from the original selections based on additional information. Each team received separate "prior knowledge" in the form of additional information printed on a sheet of paper. This served as information known by one team that could not be directly shared digitally using SAGE2 in order to better represent authentic distance collaboration between experts in various domains. Each group repeated the task three times, once using each data sharing technique.

4.3.2.1 Formal Study Data

The focus of the formal study was to see how PDTs worked together to completed a task. In order to evaluate the effectiveness of collaboration, I decided to monitor participants' perception of collaboration by conducting post-use surveys, and measure the quality of collaboration by cataloguing each session with audio / video recordings and user interaction logs. Secondarily, I also decided to measure the effectiveness of completing the tasks by collecting data about completion time and task accuracy. It is important to note that secondary items of interest do not necessarily correlate with level of collaboration. For example, one person out of a group of four participants may dominate the entire task but still finish quickly and achieve accurate results. Conversely, another group of four participants may split work evenly and maintain ongoing communication throughout the task, but take longer to complete the task ensuring that they exhaust all possible options.

The surveys asked participants to score three different aspects relating to the group collaboration on a scale of 1 to 10. First, participants were asked how easy each data sharing technique, as well as the local collaboration features, was to use. This question was designed to measure how intuitive each data sharing technique was. Second, participants were asked how successful each data sharing technique, as well as the local collaboration features, was at facilitating collaboration. This question was designed to measure participants' perceptions of the quality of collaboration enabled by each data sharing technique. Finally, participants were asked how much they liked the remote collaboration features. This question was designed to gain insight on how users felt most comfortable collaborating with a remote group. In addition to scoring the three aspects of the data sharing techniques, participants were asked to rank the three data sharing techniques from best to worst.

The videos were recorded in order to be coded into collaboration modes that each group was engaged in. Collaboration is broken up into four distinct modes: not collaborating, communicating, conferencing, or coordinating. Watching the videos from each site would provide insight into what percentage of time the groups spent in each collaboration mode. All participant interaction with SAGE2 was logged in order to create a timeline and look for recurring interaction patterns that existed for each collaboration mode. The logs consisted of data for each participant on when they moved/resized a window, interacted inside an application, and shared an application with the remote site.

The videos were also used to time each session in order to determine how long the group took complete task 1 and task 2 of the study for each data sharing technique. Also, the locations that the group chose for both task 1 and task 2 were recorded to compare to ground truth answers in order to measure task accuracy for each data sharing technique.

4.3.2.2 Formal Study Methods

Team A was assigned to the large shared display and had constraints relating to other coffee shops, main streets, and highways. Team B was assigned to the small shared display and had constraints relating to donut shops, parking lots, and building roof color. These constraints were chosen since they are easily visible on a mapping application in either roadmap view or satellite view, not due to authentic relevance in determining an ideal location for a real coffee shop.

Ten fictional coffee shop locations and ten fictional donut shop locations were chosen within an approximately 16 km^2 area of a city in the United States of America. Real features on the map of the city were used as data about roads, parking lots, and building roofs. Fictional data was also generated to act as the “prior knowledge”. Four pieces of information were given to each team. These pieces of data related to each team's unique constraints and modified what was depicted in the mapping application, such as a parking lot being demolished or a new coffee shop opening. Appendix B provides a copy of all sheets of paper used as the prior knowledge. The locations of the coffee shops and donut shops were chosen such that there would be five locations in the city that would satisfy all constraints for both teams.

Fictional data was generated for the additional information required to complete the second task of choosing one final location to open a new coffee shop. Team A was given data about crime and storm damages in 16 areas of the city and Team B was given data about family income and business profits in 16 areas of the city. Figure 13 shows an example of one of these extra pieces of data. The values for these four variables were chosen such that there would be one optimal location area in the city for the final coffee shop. Appendix C provides all additional pieces of data used for task 2.

Figure 13. Example of the additional information given to teams so that they could select their final choice for a coffee shop location. This example depicts city areas and their average family incomes.

Data was generated for a total of six cities: Denver, Houston, Sacramento, Nashville, Philadelphia, and Phoenix. These cities were chosen since they were mid-sized urban areas that were not at the location were the study took place.

In order to complete both tasks, the participants were given a custom multi-user mapping application to use within SAGE2. The map could be toggled between two different views – roadmaps and satellite imagery. The map could also be panned and zoomed interactively. A semi-transparent blue rectangle depicted the approximately 16 km^2 searchable area within the city. Orange circles were used to denote the fictional locations of existing coffee shops. Purple circles were used to denote the fictional locations of existing donut shops. Participants were able to interactively add and remove two different colored markers – red and blue. There was no inherent meaning to the different colored markers; each group came to a consensus on how they wanted to encode data. Participants could interact with the red and blue markers by opening a context menu on the map. If the context menu was opened in free space, the user would have options to add a red marker, add a blue marker, or close the context menu. If the context menu was opened over an existing marker, the user would have options to change the marker color (switch between red and blue), remove the marker, or close the context menu. A panel on the side of the application could be used to toggle the visibility of any feature. Figure 14 depicts two users interacting with the custom mapping application used in this study. All interaction with the mapping application could be done using a mouse pointer. Left-click and drag was used to pan the map. The scroll wheel was used to zoom the map in and out. Right-click was used to open a new context menu. Left-click on the buttons in the side panel would trigger the toggling of visibility of the corresponding items.

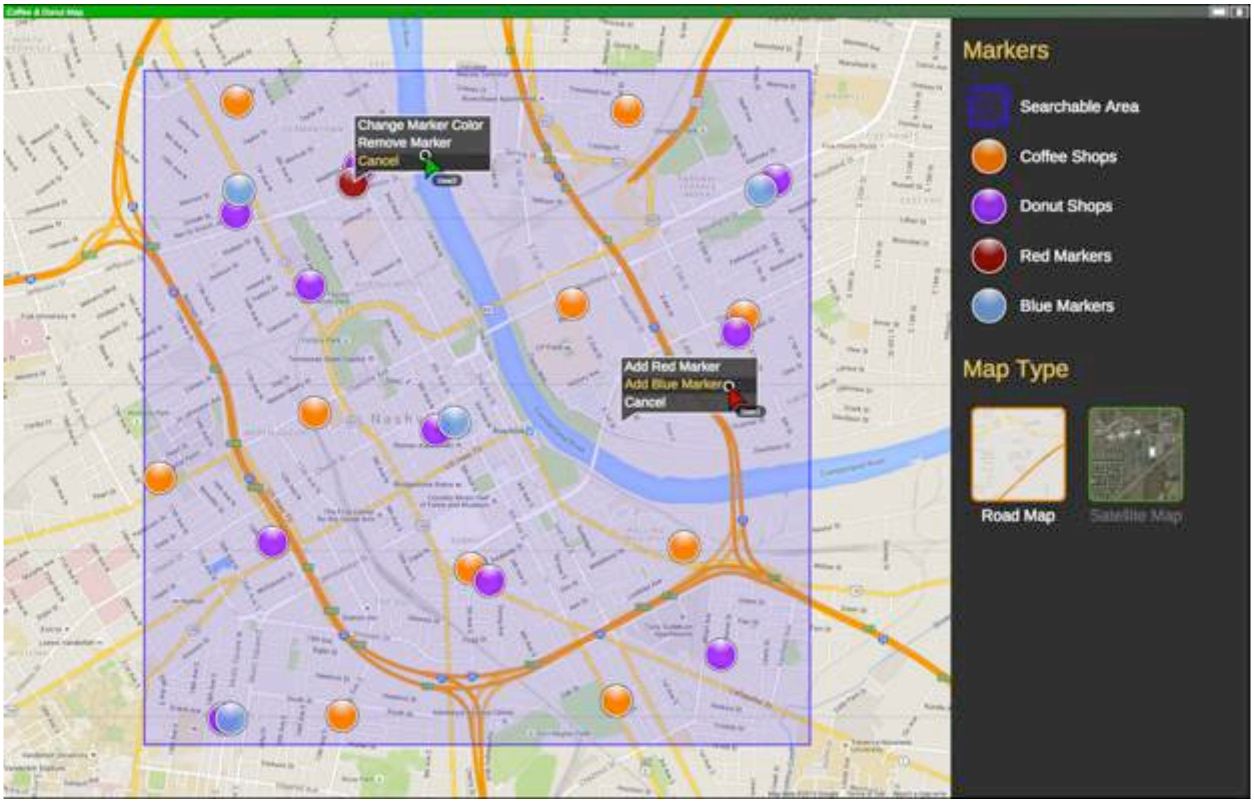


Figure 14. Mapping application used in the formal user study. Multiple users could interact simultaneously to manipulate the map, or add/remove markers.

Prior to starting the study, each group was allowed to interact with the mapping application, which had sample data using the city of Chicago with 3 coffee shop locations and 3 donut shop locations. This session lasted approximately 5 minutes, and enabled participants to get familiar with the controls of the mapping application. To complete the study, each group of subjects repeated the tasks three times, once for each data sharing technique. A brief training session on how to use the collaborative features of the upcoming data sharing technique preceded each set of tasks. Each time the type of data was the same, but in a different city with different locations of existing coffee shops and donut shops. The three conditions were:

1) Remote collaboration using *data-pushing*
2) Remote collaboration using *data-duplication*
3) Remote collaboration using *advanced data-synchronization options*

The order of these conditions was counter balanced across test groups as well as which city was assigned to which remote collaboration technique. I first conducted one pilot study in order to ensure the experimental process was smooth and consistent between trials. Following the pilot study, I conducted the formal user study with eight first-time groups, each with four participants. I then took three groups of four participants from the pool of people who'd already participated once and had them repeat the experiment in order to see if there was any change over time. Second time users did not participate with any member of their first time user group. Therefore, they were familiar with the collaboration software, but did not have predefined notions for their group's dynamics. Figure 15 depicts the setup of teams and displays at both sites.

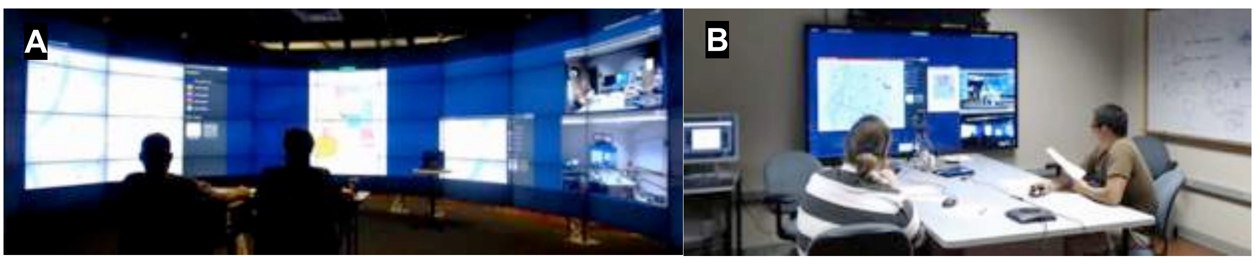


Figure 15. Partially distributed team working on finding a location to open a new coffee shop. Panel A shows the team using a large shared display. Panel B shows the team using a small shared display.

The procedure for each session of the formal user study was as follows:

- Team A, consisting of two subjects, gathered in a conference room in front of a larger shared display – a 1.72m x 9.17m display with a 12294 x 2304 resolution (approximately 28 Mpixels). Subjects were seated approximately 3.5m from the display, but were free to stand and walk closer if needed.

- Team B, consisting of two subjects, gathered in a conference room in front of a smaller shared display – a 1.05m x 1.86m display with a 3840 x 2160 resolution (approximately 8 Mpixels). Subjects were seated approximately 1.8m from the display, but were free to stand and walk closer if needed.

- Two videos were present on each site's shared display: one of the remote group (showing the people), and one of the remote group's shared display.

- Each subject had a computer mouse to interact with the shared display during the study.

- Subjects took part in a meeting to complete two tasks.

    - Task 1: find 2-4 suitable locations for a new coffee shop.

        - Team A constraints:

            - Cannot be near an existing coffee shop

            - Must be on a main street

            - Cannot be too close to a highway

        - Team B constraints:

            - Must be close to a donut shop

            - Cannot be on same block as a building with a black roof

            - Must be on same block as an outdoor parking lot

    - Task 2: Pick the best location for the new coffee shop.

        - Team A additional information:

            - Crime rates (by area)

            - Storm damage (by area)

        - Team B additional information

            - Average family income (by area)

            - Average business profits (by area)

    - Subjects had access to the following data to complete their tasks:

        - Map application

            - Location of all existing coffee shops and donut shops

            - Toggle between roadmap view and satellite view

            - Add / remove markers of two unique colors (for potential locations of new coffee shop, areas to avoid, etc. – however the teams decided how to best utilize the markers)

        - PDFs
            - Crime rates and storm damages (by area) – Team A only
            - Family income and business profits (by area) – Team B only
        - Paper documents (non-digital information that could not be shown in SAGE2)
            - Each site received unique additional information that served as “prior knowledge” heading into the task (e.g. a coffee or donut shop that is about to open or close, a new parking lot that was built after the satellite photo was taken)
    - Each site had a facilitator who was an expert with the software to perform high-level tasks (e.g. open another map, share a PDF with the other site, etc.) upon request by the subjects
    - All interaction with the shared display were logged and saved to a file.
- At the completion of all three techniques, each subject took a brief survey about their experience (using pen and paper).

4.3.3 Video Coding and Verification

For both studies, audio / video was recorded by using a screen recording software in order to capture the video conferencing window that shared the camera pointed at the team and their display. Videos from both teams were superimposed in the same frame, with the team using the large shared display in the upper right and the team using the small shared display in the lower left. Audio from the team using the large shared display was output to the right speaker, whereas audio from the team using the small shared display was output to the left speaker. This made it easy to view the entire group while also being able to distinguish conversations from each team independently. Videos were coded for collaboration mode at each location – either not collaborating, communicating, conferencing, or coordinating. Additionally, due to errors with the audio/video recording, there were a few stretches of time where the collaboration type was unknown. Table IV describes all types of errors that occurred during recording and how the video coder should handle the situation.

TABLE IV

Audio / video inconsistencies that occurred during recording of the user studies.

Issue	**Symptom**	**Resolution**
Audio recorded from wrong microphone	Soft, non-isolated audio (can hear both sides)	Attempt to hear as best as possible
Other window placed over video conferencing window	Video is partially or fully blocked	Attempt to determine collaboration from audio and video shown on other wall
GPU crashed, and later recovered	Video freezes and goes black for a period of time	Attempt to determine collaboration from audio
Video lost	No video for one team	None - code collaboration type as unknown

In order to determine collaboration mode, a set of strict definitions were given for each collaboration mode. A coder would mark down the time and mode of collaboration at any point in the video that the mode changed. Each team was coded separately, since they were not necessarily engaged in the same mode of collaboration the entire time. The definitions for the collaboration modes were as follows:

Unknown – audio / video not available.

Not collaborating – participants on one side are not collaborating. This could occur when participants are silently waiting, talking to each other about something off topic, or working independently without regard for the group.

Communicating – participants are communicating about the task one-on-one.

Conferencing – three or more participants are engaged in a discussion (both actively listening and talking count as being engaged).

Coordinating – the two teams are working in parallel after they have agreed to split the task.

If the two teams had split the task to work in parallel, the local teammates often would be engaged in silent work or communication with each other. This was still coded as coordination, since they were collaborating with the remote team at a higher level. However, if the two teams were working in parallel without having first conferenced about how they were going to split up the task, then they were actually working independent from one another. Occasionally, the facilitators would speak in the formal user study (when asked a question from a participant or needed to provide a point of clarification). This type of communication was omitted from analysis.

TABLE V

Combinations of what a team was doing (work), what a team was saying (talk), and what had happened previously in the collaboration (prior). This triplet was used to code the collaboration into one of four modes – not collaborating, communicating, conferencing, or coordinating.

Work	Talk	Prior	Collaboration Mode
None	None	No conference	Not collaborating
None	None	Conference: same task	Not collaborating
None	None	Conference: split task	Not collaborating
None	One-on-one	No conference	Communicating
None	One-on-one	Conference: same task	Communicating
None	One-on-one	Conference: split task	Coordinating
None	Group	No conference	Conferencing
None	Group	Conference: same task	Conferencing
None	Group	Conference: split task	Conferencing
Independent	None	No conference	Not collaborating
Independent	None	Conference: same task	Not collaborating
Independent	None	Conference: split task	Coordinating
Independent	One-on-one	No conference	Communicating
Independent	One-on-one	Conference: same task	Communicating
Independent	One-on-one	Conference: split task	Coordinating
Independent	Group	No conference	Conferencing
Independent	Group	Conference: same task	Conferencing
Independent	Group	Conference: split task	Conferencing
Parallel	None	No conference	Not collaborating
Parallel	None	Conference: same task	Not collaborating
Parallel	None	Conference: split task	Coordinating
Parallel	One-on-one	No conference	Communicating
Parallel	One-on-one	Conference: same task	Communicating
Parallel	One-on-one	Conference: split task	Coordinating
Parallel	Group	No conference	Conferencing
Parallel	Group	Conference: same task	Conferencing
Parallel	Group	Conference: split task	Conferencing

The guidelines for coding collaboration modes in the video analysis were based upon three important aspects – what each team was doing, what each team was saying, and what had happened previously. What each team was doing is split into three types of work: not working (none), working while the other team was not (independent), and working along with the other team (parallel). What a team was saying was split into three types of talk: not talking (none), talking with one other person (one-on-one), and talking to the whole group (group). What had happened previously in the collaborative session was split into three different types of prior actions: no prior group conferencing (no conference), a prior group conference where both teams agreed to work on the same task (conference: same task), and a prior group conference where the teams agreed to divide the work (conference: split task). Table V outlines the combinations of these three aspects and which collaboration mode it meant the team was in.

After coding all the videos of the formal user study for collaboration modes, a second investigator coded one trial (all three runs for both the team using the large shared display and the team using the small shared display) for inter-coder reliability verification. The trial chosen for verification was selected due to the fact that it contained no errors leading to unknown collaboration types, and that the participants engaged in all modes of collaboration at some point during the trial. For this, each coding was broken down to the collaboration mode assigned to each second of the video. Both percent agreement and Krippendorff's Alpha [76] were used to verify the encoding process. Percent agreement is the percentage of time the two coders agreed with their coded value. The overall percent agreement for the encoding of collaboration modes for the entire trial was 93.0%. Krippendorff's Alpha is a more stringent measurement, with values $\alpha \geq 0.800$ corresponding to a reliable coding, values $0.800 > \alpha \geq 0.600$ corresponding to a tentatively reliable coding, and values $\alpha < 0.600$ corresponding to unreliable coding. The overall value of Krippendorff's Alpha for the coding of collaboration modes for the entire trial was 0.897. Values for the six individual encodings within the trial are shown in Table VI.

TABLE VI

Inter-coder reliability results for encoding an audio / video recording for the collaboration mode that each team is engaged in.

Technique / Display	Percent Agreement	Krippendorff's Alpha
Data-pushing / Large display	81.7%	0.700
Data-pushing / Small display	86.9%	0.802
Data-duplication / Large display	97.1%	0.946
Data-duplication / Small display	95.9%	0.917
Advanced data-synchronization options / Large display	98.0%	0.963
Advanced data-synchronization options / Small display	97.8%	0.960
Entire trial / Both displays	93.0%	0.897

CHAPTER 5

EFFECTS OF VARYING THE DATA SHARING TECHNIQUE BETWEEN REMOTE COLLABORATORS

In this chapter, the impact of using different data sharing techniques is investigated. First, the results of the longitudinal user study will be explored. This study focused on a real-world on-going collaboration that was only loosely constrained. Second, the results of the formal user study will be explored. This study focused on having a new group work together in order to accomplish a search and analysis task. Both scenarios have design implications for creating groupware that supports users at multiple remote locations.

5.1 Results of the Longitudinal User Study on SAGE2 Development

After each collaborative session, participants were asked fill out a survey about their experience. The first question asked participants about involvement in that particular meeting. This question served as a benchmark as I continued to develop the software, but did not have meaning in terms of user perception about collaboration based on data sharing technique. The remaining questions asked users to score the ease of use, successfulness of collaboration, and how much they liked the features on a scale from 1 to 10 (1 being worst and 10 being best). The full survey can be viewed in Appendix D. Average scores and their standard errors from the survey are summarized in Table VII.

TABLE VII

Overall results of the participant survey for the longitudinal user study. Answers were scored on a scale of 1 -10, with 1 being worst and 10 being best.

Survey Question	**Average Score**	**Number of Answers**
Results on Ease of Use for Group Interaction		
Local group	8.45 ± 0.14	75
Remote group using *data-pushing*	8.04 ± 0.28	28
Remote group using *data-duplication*	7.45 ± 0.33	22
Remote group using *advanced data-synchronization options*	8.56 ± 0.24	25
Results on How Successful the Tool was at Facilitating Collaboration		
Local collaboration	8.35 ± 0.15	75
Remote collaboration using *data-pushing*	7.96 ± 0.28	28
Remote collaboration using *data-duplication*	8.09 ± 0.27	22
Remote collaboration using *advanced data-synchronization options*	8.44 ± 0.25	25
Results on How Much Users Liked the Remote Collaboration Features		
Features of *data-pushing*	8.18 ± 0.26	28
Features of *data-duplication*	8.32 ± 0.27	22
Features of *advanced data-synchronization options*	8.60 ± 0.22	25

A two-tailed t-test was applied to compare the three data sharing techniques with each other as well as the local collaboration. The only significant differences were that both local interaction and remote interaction using *advanced data-synchronization options* were scored significantly higher for ease of use than remote interaction using data-duplication, with p-values < 0.05. All other differences between data sharing techniques were not significant. Ease of use comparisons are shown in Figure 16, Panel A. Success of facilitating collaboration comparisons are shown in Figure 16, Panel C. Comparisons on how much users liked the different data sharing techniques are shown in Figure 16, Panel E.

In addition to looking at the overall averages from the survey answers, I investigated to determine if there were any changes over time. Figure 16, Panels B, D, and F depicts the changes of survey answers over the four sessions that each data sharing technique was used.

There were no significant changes or patterns that emerged – scores fluctuated slightly in either directions.

Figure 16. Graphs of survey results from the longitudinal user study. Panel A shows overall reported ease of use. Panel B shows reported ease of use over time. Panel C shows how successful overall users found features at facilitating collaboration. Panel D shows how successful users found features at facilitating collaboration over time. Panel E shows how much overall users liked remote collaboration features. Panel F shows how much users liked remote collaboration features over time.

Although all sessions of the longitudinal user study were audio / video recorded, they were not formally evaluated for coding collaboration modes. The meetings on continued development of SAGE2 were primarily used as a forum for participants to deliver progress reports to the group, critique existing features, or brainstorm potential new features. Therefore remote sites were nearly always engaged in a group conference. This resulted in SAGE2 primarily being used as a tool for sharing the two videoconferencing windows on a SRSD, with other features rarely being used for local collaboration, let alone remote collaboration. Therefore, no further analysis was performed on the data collected during the longitudinal user study.

5.2 Results of the Formal User Study on Opening a New Coffee Shop

The formal user study was conducted with eleven trials, each with four participants. Participants ranged in age from 18 - 59, with 75% males and 25% females, and 63% native English speaker and 37% non-native English speakers. All participants were computer literate and had either completed a college degree or were in the process of obtaining one. This met the target audience of individuals who may be using data-conferencing software in the near future at their workplace.

5.2.1 Overall Results

This subsection covers the overall results of all trials. First I will cover participant user survey answers. Next I will cover the audio / video analysis for collaboration mode along with the user interaction logs. Finally, I will show the results from the secondary outcomes – completion time and accuracy.

5.2.1.1 Overall Survey Results

The survey for the formal user study consisted of twelve questions, plus the opportunity to leave extra feedback. For the first eleven questions, participants were asked to score certain aspects of collaboration on a scale from 1 to 10 (1 being worst and 10 being best). Similar to the longitudinal study, these questions were about the ease of use, successfulness of collaboration, and how much the participants liked the collaborative features. The twelfth question asked participants to rank the data sharing techniques from best to worst. The full survey can be viewed in Appendix E. Results from all 44 participants who answered survey questions 1-11 are summarized in Table VIII.

TABLE VIII

Overall results of the participant survey for the formal user study. Answers were scored on a scale of 1 - 10, with 1 being worst and 10 being best. There was a total of 44 participants that gave responses for this survey.

Survey Question	**Average Score**
Results on Ease of Use for Group Interaction	
Local group	8.84 ± 0.17
Remote group using *data-pushing*	7.34 ± 0.33
Remote group using *data-duplication*	8.14 ± 0.28
Remote group using *advanced data-synchronization options*	8.25 ± 0.20
Results on How Successful the Tool was at Facilitating Collaboration	
Local collaboration	8.53 ± 0.22
Remote collaboration using *data-pushing*	6.82 ± 0.34
Remote collaboration using *data-duplication*	7.95 ± 0.33
Remote collaboration using *advanced data-synchronization options*	8.57 ± 0.18
Results on How Much Users Liked the Remote Collaboration Features	
Features of *data-pushing*	6.91 ± 0.37
Features of *data-duplication*	7.86 ± 0.33
Features of *advanced data-synchronization options*	8.34 ± 0.21

A two-tailed t-test was applied to compare the three data sharing techniques with each other as well as the local collaboration. Users reported that both *data-duplication* and *advanced data-synchronization options* were significantly easier to use than *data-pushing*, with p-values < 0.05. While users reported that *advanced data-synchronization options* was easier to use than *data-duplication*, there was not a significant difference. When comparing remote collaboration to the collaboration of the local group, the users reported that the local collaboration features were significantly easier to use than all three data sharing techniques, with p-values < 0.05. Ease of use results are visualized in Figure 17, Panel A. Users also reported that both *data-duplication* and *advanced data-synchronization options* were significantly more successful at facilitating collaboration than *data-pushing*, with p-values < 0.05. While users reported that *advanced data-synchronization options* was more successful at facilitating collaboration than *data-duplication*, there was not a significant difference. The local collaboration features also were significantly more successful at facilitating collaboration than *data-pushing*, with a p-value < 0.05. However, there was no significant difference between how successful the local collaboration features were at facilitating collaboration and either *data-duplication* or *advanced data-synchronization options*. Success at facilitating collaboration results are visualized in Figure 17, Panel B. Finally, users reported that they like the remote collaboration features of both *advanced data-synchronization options* and *data-duplication* significantly more than *data-pushing*, with p-values < 0.05. While users reported that they like the remote collaboration features of *advanced data-synchronization options* more than *data-duplication*, there was not a significant difference. How much users liked remote collaboration features are visualized in Figure 17, Panel C.

Users also ranked the three data sharing techniques, with 1 being the best, and 3 being the worst. *Advanced data-synchronization options* was ranked highest with an average rank of 1.68 and a standard error of 0.12. Data-duplication was ranked second highest with an average rank of 1.80 and a standard error of 0.12. Data-pushing was ranked last with an average rank of 2.50 and a standard error of 0.09. These results are summarized in Table IX. A two-tailed t-test was applied to compare the ranking of the three data sharing techniques with each other. Both *advanced data-synchronization options* and *data-duplication* were ranked significantly better than *data-pushing*, with p-values < 0.05. However, there was no significant difference between

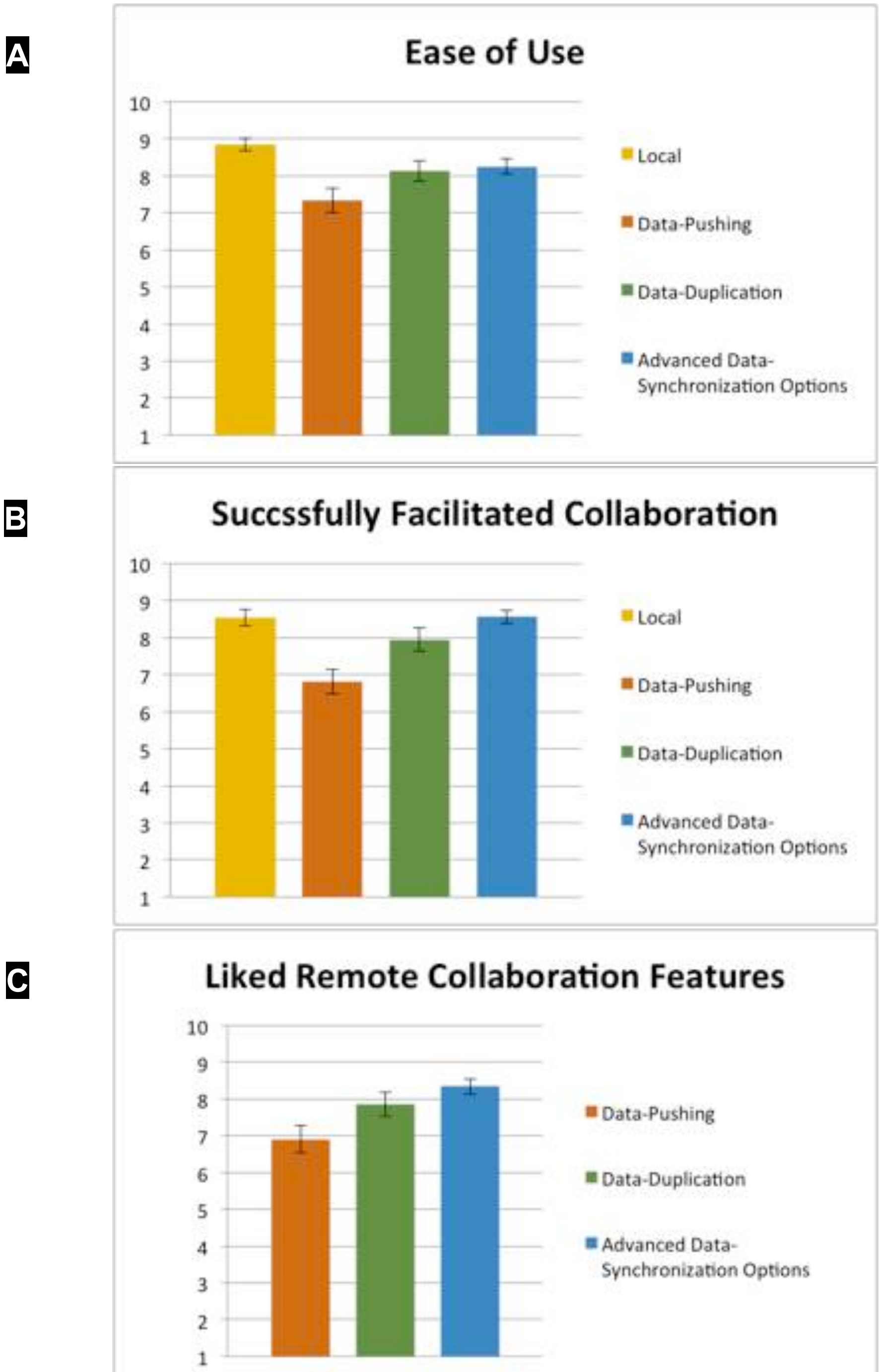


Figure 17. Graphs of overall survey results. Panel A shows reported Ease of Use. Panel B shows how successful users found features at facilitating collaboration. Panel C shows how much users liked the remote collaboration features.

the ranks of *advanced data-synchronization options* and *data-duplication*. These survey results back up Hypotheses 1 and 2 – that both continuous synchronization data sharing techniques would improve collaboration compared to *data-pushing*, and that local collaboration still is superior to remote collaboration (but not significantly so when compared to the continuous synchronization data sharing techniques).

TABLE IX

Average overall rank for the three data sharing techniques, with 1 being the best and 3 being the worst.

Data Sharing Technique	Rank (1-Best, 3-Worst)
Data-pushing	2.50 ± 0.09
Data-duplication	1.80 ± 0.12
Advanced data-synchronization options	1.68 ± 0.12

5.2.1.2 Overall Audio / Video and User Interaction Results

The videos containing the teams from both the large display and small display were analyzed and coded into sections of time based on the mode of collaboration that each side was partaking in. Each team was coded separately, since the entire group was not always engaged in the same mode of collaboration at the same time. Collaboration modes for each technique were normalized based on the length it took to complete the task, excluding any time there were technical errors causing an unknown collaboration mode, so that each trial carried an equal weight rather than teams who took longer having a larger impact on the averages.

On average, teams using the *data-pushing* technique spent 15.5% of their time not collaborating, 33.8% of their time communicating one-on-one, 33.2% of their time conferencing as a group, and 17.4% of their time coordinating their tasks. Teams using the *data-duplication* technique spent 10.4% of their time not collaborating, 31.5% of their time communicating one-on-one, 46.5% of their time conferencing as a group, and 11.5% of their time coordinating their tasks. Teams using the *advanced data-synchronization options* technique spent 6.4% of their time not collaborating, 24.7% of their time communicating one-on-one, 50.1% of their time conferencing as a group, and 18.8% of their time coordinating their tasks. Overall collaboration mode results are summarized in Table X and visualized in Figure 18.

TABLE X

Average overall percentage of time spent in each collaboration mode based on data sharing technique.

Data Sharing Technique	Not Collaborating	Communicating	Conferencing	Coordinating
Data-pushing	15.5% ± 2.8%	33.8% ± 3.9%	33.2% ± 2.7%	17.4% ± 4.0%
Data-duplication	10.4% ± 2.7%	31.5% ± 4.3%	46.5% ± 3.5%	11.5% ± 2.6%
Advanced data-synchronization options	6.4% ± 2.1%	24.7% ± 4.0%	50.1% ± 3.6%	18.8% ± 3.9%

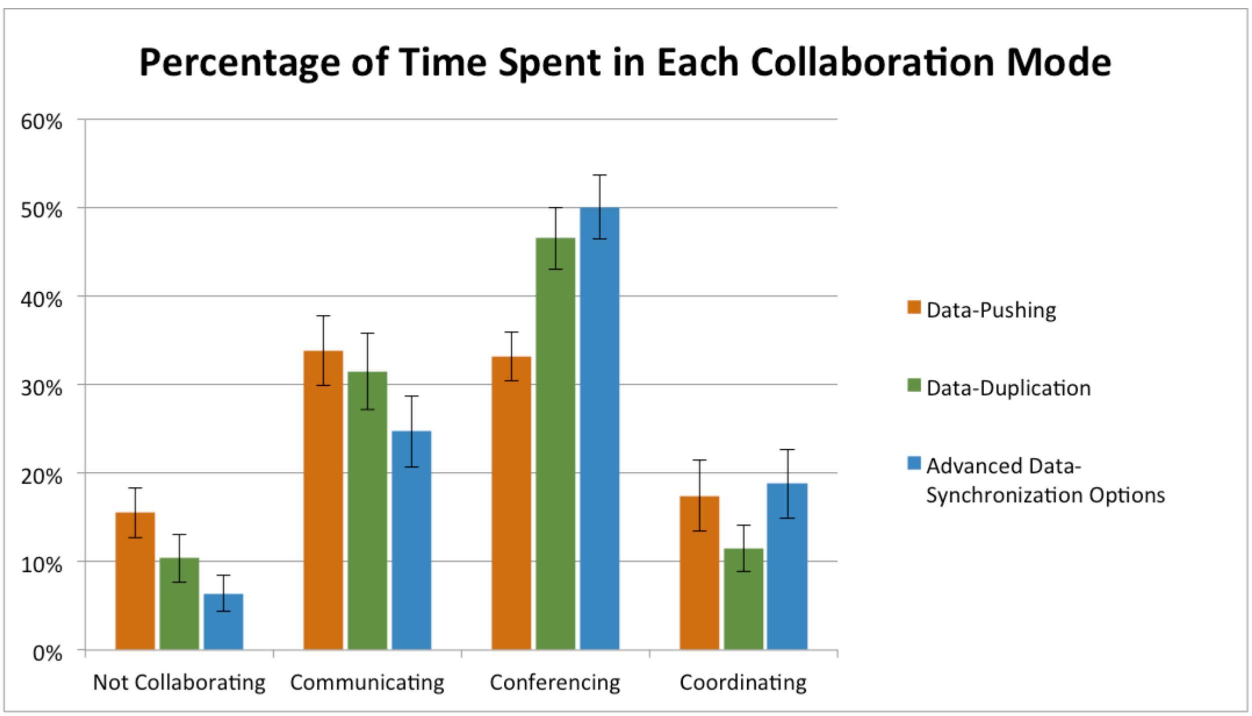


Figure 18. Average overall percentage of time spent in each collaboration mode based on data sharing technique.

A two-tailed t-test was applied to compare time spent in each mode of collaboration based on the three data sharing techniques. When comparing *data-duplication* to *data-pushing*, the only significant difference was that participants engaged in group conferences significantly more when using *data-duplication* than when using *data-pushing*, which had a p-value < 0.05. When comparing *advanced data-synchronization options* to *data-pushing*, participants were 'not collaborating' or 'communicating' significantly less when using *advanced data-synchronization options* than when using *data-pushing*. Also, participants were engaged in a group conference significantly more when using *advanced data-synchronization options* than when using *data-pushing*, all of which had a p-values < 0.05. When comparing *advanced data-synchronization options* to *data-duplication*, the only significant difference was that participants coordinated their efforts significantly more when using *advanced data-synchronization options* than when using *data-duplication*, which had a p-value < 0.05.

In order to further analyze collaboration modes, I compared the results for each trial to see how often the team using the large shared display and the team using the small shared display were in the same collaboration mode. For this analysis, it did not matter which collaboration mode the teams were in, but only if they were in the same mode as each other or not. The more often the two teams were in the same collaboration mode, the more inline the two teams were with each other. This indicated that the two teams had a better awareness of their remote collaborators.

Similar to determining the amount of time the teams spent in each collaboration, the amount of time spent in the same collaboration mode for each trial has been normalized based on the length it took to complete the task, excluding any time there were technical errors causing unknown an collaboration mode, so that each trial carried an equal weight. Teams using *data-pushing* were in the same collaboration mode as their remote counterparts an average of 69.0% of the time, with a standard error of 4.6%. Teams using *data-duplication* were in the same collaboration mode as their remote counterparts an average of 77.4% of the time, with a standard error of 4.5%. Teams using *advanced data-synchronization options* were in the same collaboration mode as their remote counterparts an average of 87.7% of the time, with a standard error of 3.6%. Overall collaboration mode similarity results are summarized in Table XI and visualized in Figure 19.

TABLE XI

Average percentage of time that the team using the large shared display and the team using the small shared display were in the same collaboration mode as each other based on data sharing technique.

Data Sharing Technique	**Percentage of Time in Same Collaboration Mode**
Data-pushing	69.0% ± 4.6%
Data-duplication	77.4% ± 4.5%
Advanced data-synchronization options	87.7% ± 3.6%

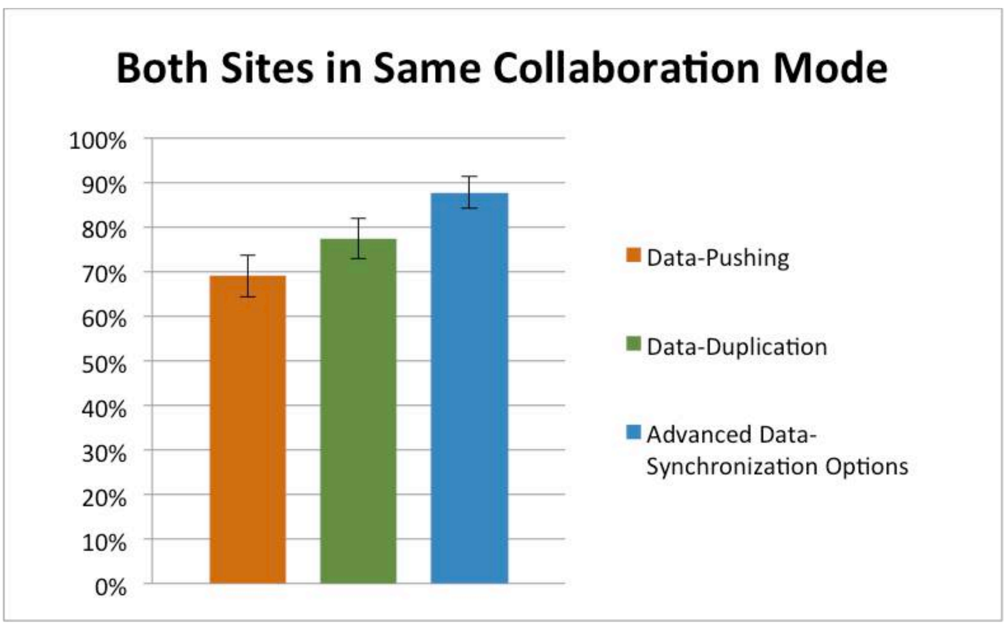


Figure 19. Average overall percentage of time that the team using the large shared display and the team using the small shared display spent in the same collaboration mode as each other based on data sharing technique. More time spent in the same collaboration mode indicates an increased awareness of the remote collaborators.

There is a clear ordering with the *data-pushing* teams being the least inline with each other up to the *advanced data-synchronization options* teams being the most inline with each other, with approximately a 10% increase in time spent in the same collaboration mode between each technique. However, when applying a two-tailed t-test to compare the three data sharing techniques, the difference is only significant between the *advanced data-synchronization options* technique and the *data-pushing* technique, with a p-value < 0.05. The collaboration mode results coded from the audio/video analysis further support Hypothesis 1 – that both continuous synchronization data sharing techniques improve collaboration compared to *data-pushing*.

Finally, I combined the collaboration mode coding from the videos with the user interaction logs in order to create a timeline view of each trial and investigate for patterns of interaction. In order to create a timeline view, I took the data from both teams and stacked them on top of each other. First there is the user interaction log data, where a tick mark is drawn for each type of interaction by each user at the corresponding time in the visualization. The width of the tick mark does not correspond to any data, since it only represents a single point in time. Second there is the collaboration mode data, where colored blocks represent chunks of time the team spent in each mode. The result shows how each team interacted while in each mode of collaboration as well as how they collaborated with each other. These visualizations can be created using absolute time (easy to compare multiple timelines) or normalized time (each timeline is same width regardless of task length). This elucidated a few interesting collaboration schemes that teams used.

First, some groups depicted a turn-taking pattern, where only one teams worked at a time and data was sent to the other team when the turn changed. An example of this is shown in Figure 20. This trial started with the team using the large shared display interacting with an application while the team using the small shared display was simply talking but not interacting with any application. After a period of time, the roles reversed, with the team using the large display no longer interacting with any application, and the team using the small display interacting with an application. In this instance, the roles switched two more times in total.

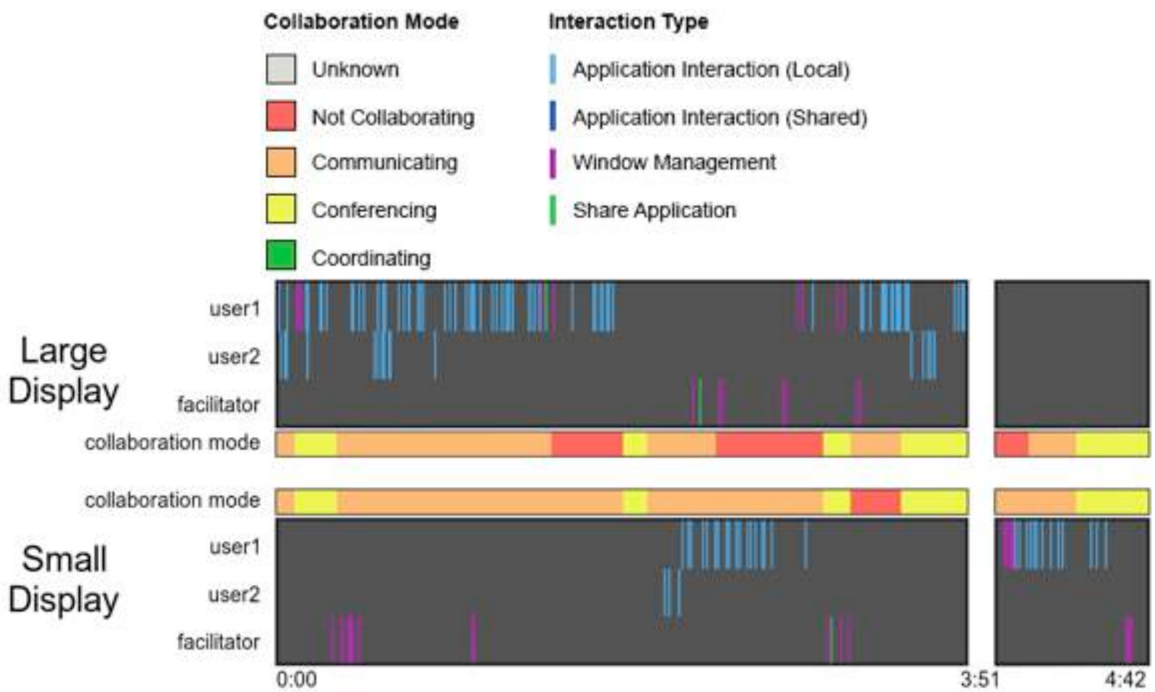


Figure 20. User interaction log and collaboration modes coded from the audio / video analysis depicting a turn-taking pattern by the two teams. The split time and break in the timeline visualization denotes the switch from working on task 1 to working on task 2.

Another pattern that occurred was having one team finish early and wait for the other team to catch up. An example is shown in Figure 21. After both teams were working for a while in this trial, the team using the large shared display finished their work and waited for the team using the small shared display. This becomes apparent when evaluating the fact that the team using the large shared display was no longer collaborating or interacting with any applications. Once the team using the small shared display finished, they started a group conference and the team using the large shared display started working again. This happened again at the very end of

Task 2, but with the team using the small shared display finishing first and no longer collaborating or interacting with any applications.

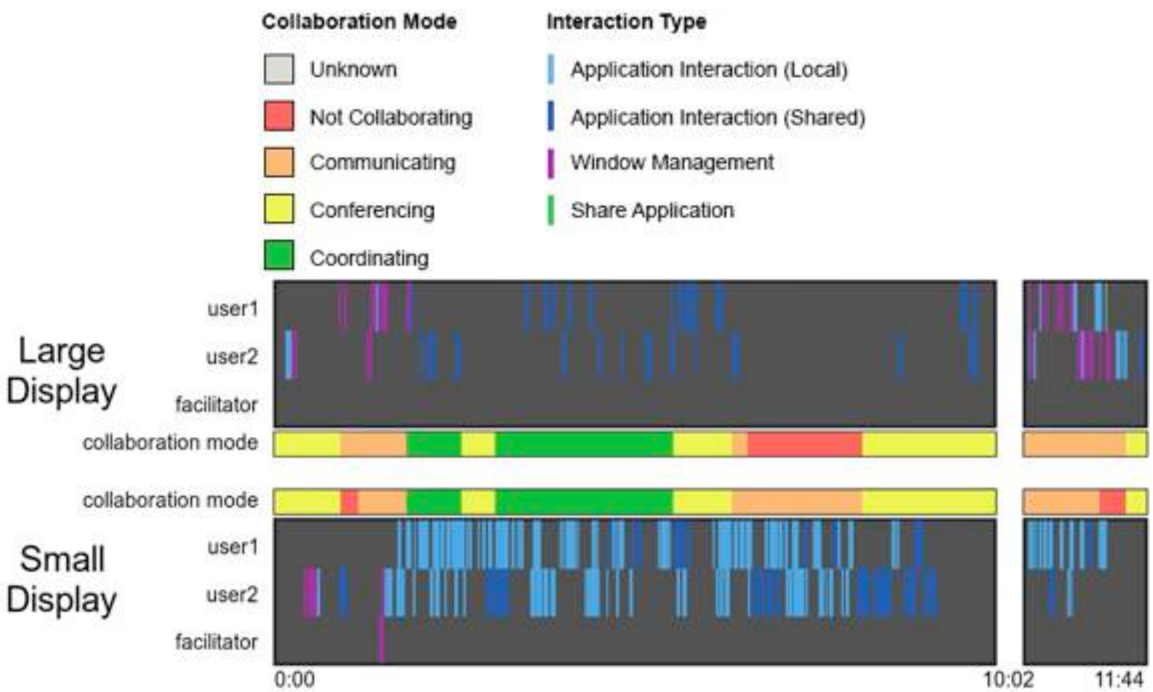


Figure 21. User interaction log and collaboration modes coded from the audio / video analysis depicting a finish and wait pattern. The split time and break in the timeline visualization denotes the switch from working on task 1 to working on task 2.

One final pattern that occurred was when the two teams would primarily each be working on a local application, then conference with each other, then each primarily be working in a shared application. An example is shown in Figure 22. The two teams started by conferencing followed by doing coordinated work on their own local applications. After a brief conference, the two teams each primarily worked in a shared application together.

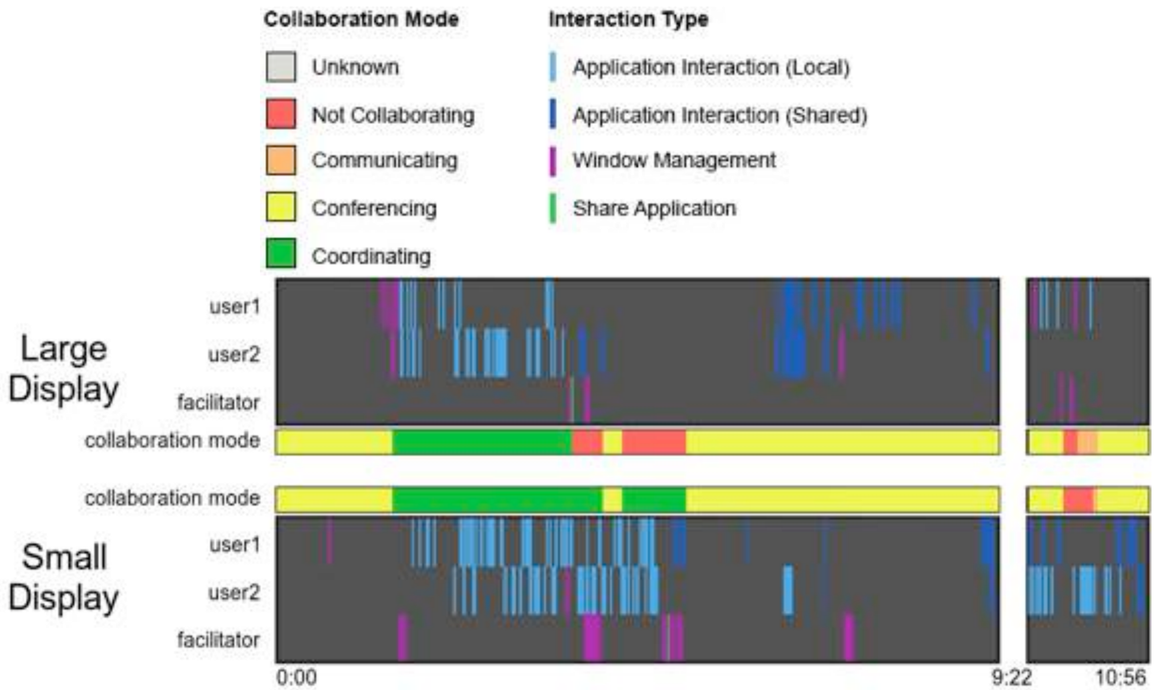


Figure 22. User interaction log and collaboration modes coded from the audio / video analysis depicting a pattern of both teams primarily working on local applications, followed by a conference, then both teams switching to primarily working in a shared application. The split time and break in the timeline visualization denotes the switch from working on task 1 to working on task 2.

See Appendix F for timeline visualizations of all trials.

5.2.1.3 Overall Completion Time and Task Accuracy Results

As a secondary analysis I looked at the results of task completion time and accuracy. These were deemed less important than the results discussed in the previous two subsections, since they do not necessarily directly correlate with the quality of group collaboration. When looking at completion time for both task 1 and task 2 between the three data sharing techniques, there are no significant differences. Groups using *data-pushing* completed task 1 in an average of 12 minutes and 29 seconds and task 2 in an average of 3 minutes and 9 seconds. Groups using *data-duplication* completed task 1 in an average of 13 minutes and 17 seconds and task 2 in an average of 2 minutes and 44 seconds. Groups using *advanced data-synchronization options* completed task 1 in an average of 12 minutes and 52 seconds and task 2 in an average of 2 minutes and 54 seconds. Completion time averages and there standard errors are summarized in Table XII and visualized in Figure 23.

TABLE XII

Average overall completion times (in minutes and seconds) for task 1 and task 2 of the formal user study based on data sharing technique.

Data Sharing Technique	**Completion Time (Task 1)**	**Completion Time (Task 2)**
Data-pushing	12:29 ± 1:43	3:09 ± 0:34
Data-duplication	13:17 ± 2:22	2:44 ± 0:32
Advanced data-synchronization options	12:52 ± 1:57	2:54 ± 0:33

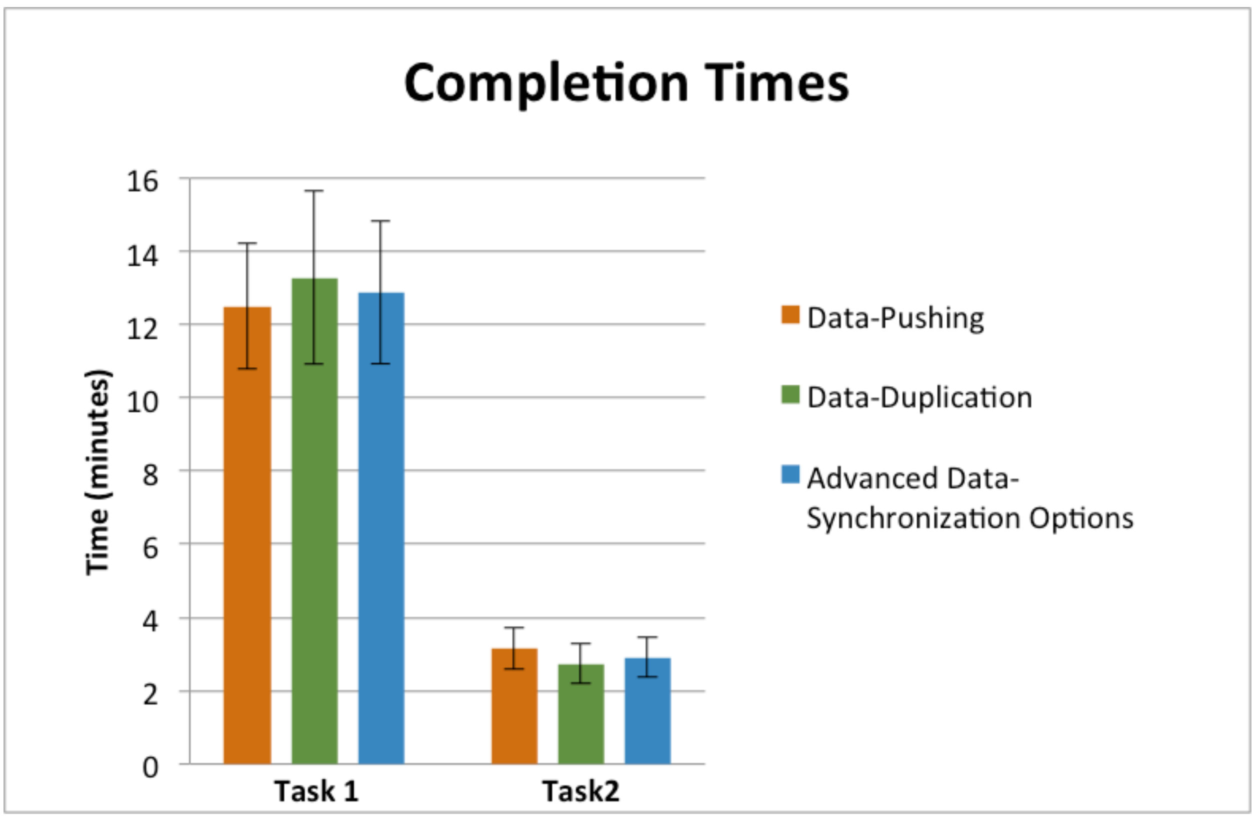


Figure 23. Average overall completion times for task 1 and task 2 of the formal user study based on data sharing technique.

Since the order of each data sharing technique was counter balanced, I also looked at the effects order had on completion time. The first technique the group used took an average of 16 minutes and 7 seconds to complete task 1 and average of 3 minutes and 35 seconds to complete task 2. The second technique the group used took an average of 12 minutes and 52 seconds to complete task 1 and average of 2 minutes and 29 seconds to complete task 2. The third technique the group used took an average of 9 minutes and 40 seconds to complete task 1 and average of 2 minutes and 44 seconds to complete task 2. Completion time averages and their standard errors are summarized in Table XIII and visualized in Figure 24.

TABLE XIII

Average overall completion times (in minutes and seconds) for task 1 and task 2 of the formal user study based on technique order.

Technique Order	Completion Time (Task 1)	Completion Time (Task 2)
Technique 1	16:07 ± 2:05	3:35 ± 0:32
Technique 2	12:52 ± 2:04	2:29 ± 0:30
Technique 3	9:40 ± 1:17	2:44 ± 0:34

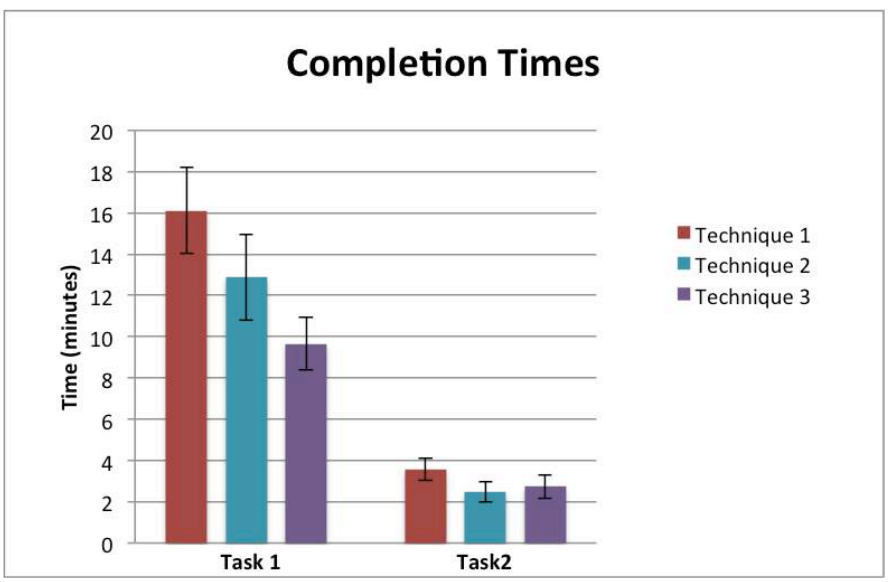


Figure 24. Average overall completion times for task 1 and task 2 of the formal user study based on data sharing technique order.

I used a two-tailed t-test to compare the completion times based on order and determine if there was a learning effect. For task 1, both the second and third technique were completed significantly faster than the first technique, with p-values < 0.05. While the third technique was completed faster on average than the second technique, there was no significant difference. For task 2, the second technique completed was significantly faster than the first technique, with a p-value < 0.05. There were no other significant differences between techniques. This suggests a short learning curve either for using the software or getting comfortable with the dynamics of a new collaboration group.

In order to assess accuracy, the each map was constructed to have five potential locations that fit all initial constraints from both teams. From those five locations, the additional information about crime, storm damages, family income, and business profit were constructed to have one optimized location. Exact location (specific latitude and longitude) was not a good measure of accuracy, since the groups were looking for general areas (i.e. any building on a given block) that met their constraints. Therefore the results were visualized as a heatmap, so

Figure 25. Example heatmap of task 1 results. Red pins show the five ground truth answers for potential locations that satisfied all constraints for both teams. Heated dots show frequency of group answers for finding 2-4 potential locations that satisfy all constraints. Panel A shows groups who used data-pushing. Panel B shows groups who used data-duplication. Panel C shows groups who used advanced data-synchronization options.

that they could be visually compared to the ground truth answers. Figure 25 shows an example of task 1 results for one city – groups using *data-pushing* in Panel A, *data-duplication* in Panel B, and *advanced data-synchronization options* in Panel C. A qualitative analysis shows that groups were generally able to correctly identify valid locations regardless of which data sharing technique they used.

Figure 26 shows an example of task 2 results for one city – groups using *data-pushing* in Panel A, *data-duplication* in Panel B, and *advanced data-synchronization options* in Panel C. While answers weren't always perfectly lined up with the ground truth, this had more to do with working from a starting point of 2-4 locations chosen in task 1, than an inability to correctly identify the best location. It was not infrequent that the best possible location was omitted from the list of potential locations simply because groups could only select 2-4. A qualitative analysis shows that groups were generally able to find the best location regardless of which data sharing technique they used.

Figure 26. Example heatmap of task 2 results. Red pins show the ground truth answer for the best location that satisfied all constraints for both teams. Heated dots show frequency of group answers for finding the best location that satisfies all constraints. Panel A shows groups who used data-pushing. Panel B shows groups who used data-duplication. Panel C shows groups who used advanced data-synchronization options.

See Appendix G for both heatmap visualizations of task 1 and task 2 accuracy results.

5.2.2 Effects of Display Size

In order to analyze the effects of display size, I separately analyzed the data from the team using the large shared display and the team using the small shared display. The large shared display was approximately 28 Mpixels and was designed to have more than enough screen real-estate to show all applications simultaneously without overlap. The small shared display was approximately 8 Mpixels and was designed to have a limited amount of space, requiring users organize multiple applications potentially resulting in overlap. First, I will present the differences in user perceptions from the survey results. Second, I will present the differences in time spent in each collaboration mode. Completion times and accuracy were not analyzed for effects of display size, since the entire group had to come to a consensus on their locations before each task ended. Therefore there are no differences in completion time or accuracy based on display size.

5.2.2.1 Display Size Dependent Survey Results

When analyzing the survey results from the participants who used the large shared display and comparing them to the participants who used the small shared display some interesting differences emerge. The first interesting pattern is that the participants using the small shared display gave higher scores on all answers than the participants using the large shared display. However, when using a two-tailed t-test to compare the differences, the only answers that were significantly different between the two teams were related to the ease of use and how much they like the *data-duplication* technique, both with p-values < 0.05. Since half the overall participants used the large shared display and the other half used the small shared display, each display size has responses from 22 participants. Average scores and their standard errors for the display size dependent survey results are summarized in Table XIV and visualized in Figure 27.

TABLE XIV

Display size dependent results of the participant survey for the formal user study. Answers were scored on a scale of 1 -10, with 1 being worst and 10 being best. There were 22 responses from participants using each display size.

Survey Question	**Average Score (Large Display)**	**Average Score (Small Display)**
Results on Ease of Use for Group Interaction		
Local group	8.68 ± 0.25	9.00 ± 0.23
Remote group using *data-pushing*	6.86 ± 0.52	7.82 ± 0.41
Remote group using *data-duplication*	7.50 ± 0.43	8.77 ± 0.32
Remote group using *advanced data-synchronization options*	8.09 ± 0.27	8.41 ± 0.31
Results on How Successful the Tool was at Facilitating Collaboration		
Local collaboration	8.32 ± 0.36	8.76 ± 0.25
Remote collaboration using *data-pushing*	6.55 ± 0.45	7.09 ± 0.05
Remote collaboration using *data-duplication*	7.50 ± 0.40	8.41 ± 0.50
Remote collaboration using *advanced data-synchronization options*	8.27 ± 0.26	8.86 ± 0.23
Results on How Much Users Liked the Remote Collaboration Features		
Features of *data-pushing*	6.64 ± 0.53	7.18 ± 0.54
Features of *data-duplication*	7.27 ± 0.49	8.45 ± 0.41
Features of *advanced data-synchronization options*	8.05 ± 0.30	8.64 ± 0.28

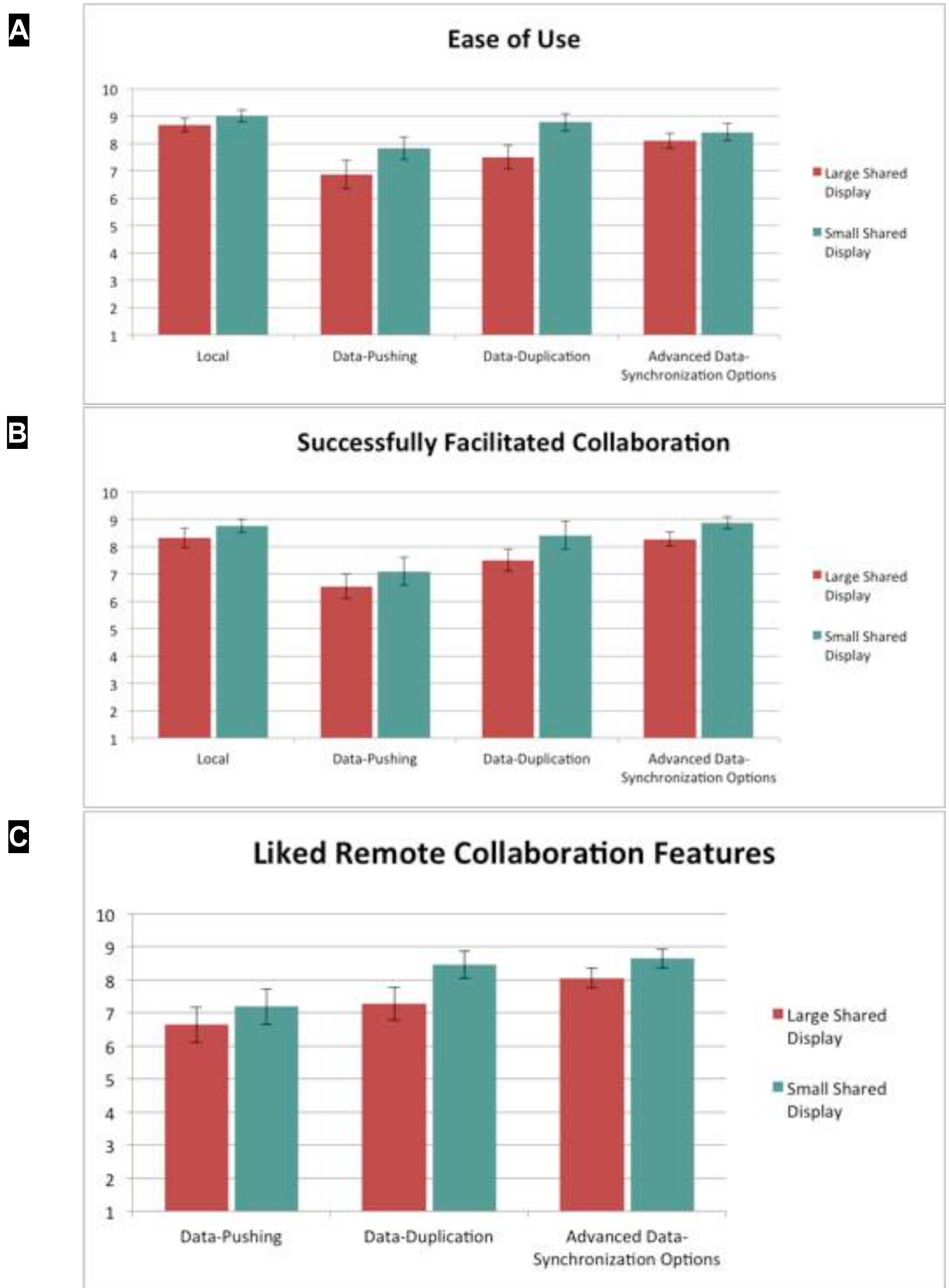


Figure 27. Graphs of display size dependent survey results depicting answers from both users of the large shared display and the small shared display. Panel A shows reported Ease of Use. Panel B shows how successful users found features at facilitating collaboration. Panel C shows how much users liked the remote collaboration features.

There were also slight discrepancies in how participants ranked the three data sharing techniques based on which size display they had used. Participants using the large shared display ranked *advanced data-synchronization options* best, followed closely by *data-duplication*. Participants using the small shared display ranked *data-duplication* best, followed closely by *advanced data-synchronization options*. These results are summarized in Table XV. However, when using a two-tailed t-test to compare the differences, there were no significant differences in ranks based on display size, with all p-values > 0.05. These survey results actually contradict Hypothesis 3 – that participants using a smaller SRSD would prefer *advanced data-synchronization options* to *data-duplication* due to limited screen real estate.

TABLE XV

Average display size dependent rank for the three data sharing techniques, with 1 being the best and 3 being the worst.

Data Sharing Technique	**Rank [Large Display] (1-best, 3-worst)**	**Rank [Small Display] (1-best, 3-worst)**
Data-pushing	2.45 ± 0.13	2.55 ± 0.14
Data-duplication	1.95 ± 0.17	1.64 ± 0.17
Advanced data-synchronization options	1.59 ± 0.18	1.77 ± 0.16

5.2.2.2 Display Size Dependent Audio / Video Results

When analyzing the coded collaboration modes from the audio / video recordings, one interesting difference stands out between the participants who used the large shared display and the participants who used the small shared display. Teams using the large shared display were in a ‘not collaborating’ state more often and ‘communicating’ or ‘coordinating’ less often. This difference exists with all data sharing techniques. However, when using a two-tailed t-test to compare the difference, the only significant difference is with the ‘not collaborating’ mode when using the *advanced data-synchronization options*, with a p-value < 0.05. Display size dependent collaboration mode results are summarized in Table XVI and visualized in Figure 28. The display size dependent collaboration mode results actually do back up Hypothesis 3 – that participants using a smaller SRSD engaged in higher modes of collaboration more frequently when using *advanced data-synchronization options* than when using *data-duplication*.

TABLE XVI

Average display size dependent percentage of time spent in each collaboration mode based on data sharing technique.

Data Sharing Technique	Not Collaborating	Communicating	Conferencing	Coordinating
(Large Display) Data-pushing	21.0% ± 4.1%	29.7% ± 4.5%	33.3% ± 3.9%	16.0% ± 6.3%
(Small Display) Data-pushing	10.1% ± 3.2%	38.0% ± 6.4%	33.1% ± 4.0%	18.9% ± 5.2%
(Large Display) Data-duplication	12.2% ± 3.4%	32.1% ± 6.8%	45.8% ± 5.4%	9.8% ± 3.0%
(Small Display) Data-duplication	8.6% ± 4.4%	30.9% ± 5.7%	47.2% ± 4.8%	13.2% ± 4.4%
(Large Display) Advanced data-synchronization options	10.9% ± 3.6%	21.8% ± 4.6%	50.3% ± 5.1%	17.0% ± 5.2%
(Small Display) Advanced data-synchronization options	2.0% ± 0.9%	27.6% ± 6.8%	49.8% ± 5.3%	20.6% ± 6.1%

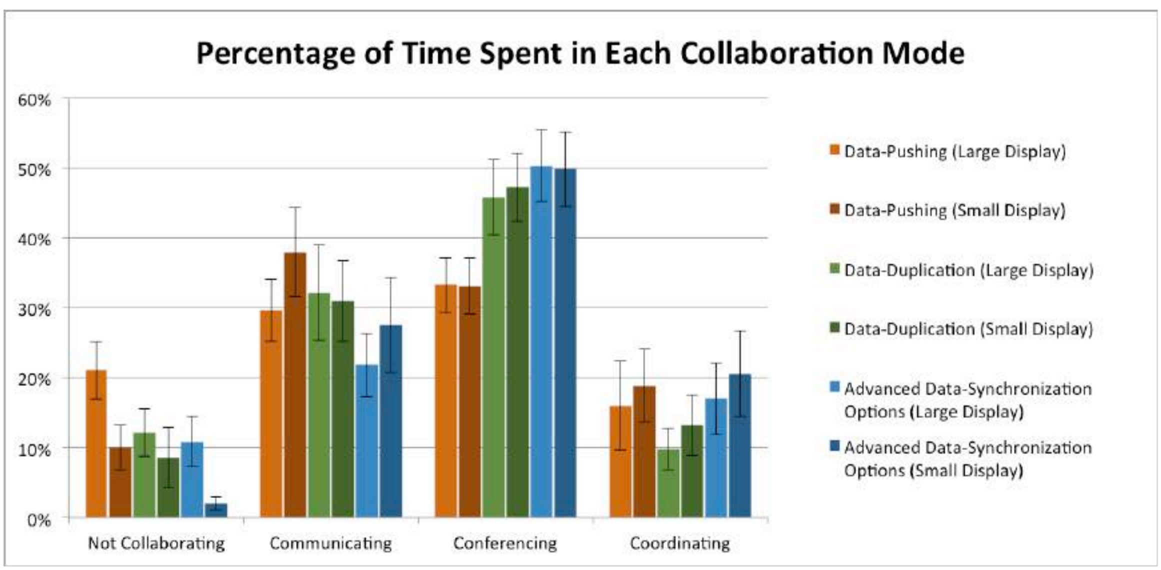


Figure 28. Average display size dependent percentage of time spent in each collaboration mode based on data sharing technique.

5.2.2.3 Differences in Display Size

There are a few important differences in how participants perceived the data sharing techniques based on whether they were using a large shared display or a small shared display. First, the participants using the large shared display found the *advanced data-synchronization options* easiest to use on average, whereas the participants using the small shared display found *data-duplication* easiest to use on average. Second, participants using the small shared display scored the *data-duplication* technique significantly higher than participants using the large shared display for both ease of use and how much they like the collaboration features, both with p-values < 0.05.

By analyzing the audio / video data coded for collaboration modes, there is one major difference between teams using a large shared display and teams using a small shared display. Teams using a large shared display were 'not collaborating' more frequently than the teams using a small shared display. While at first glance this may appear to favor using a small shared display, this result actually appeared due to some of the common patterns of collaboration that groups used, such as turn taking and finish and wait. The teams using a large shared display were able to finish their piece of a task faster than the teams using a small shared display. Therefore, teams using a large shared display had to wait longer on average, thus leaving them in a 'not collaborating' state more often than the teams using a small shared display.

5.2.3 Effects of Task Experience

In order to analyze the effects of task experience, I separately analyzed the data from the eight trials with first time users and the three trials with second time users. First, I will present the differences in user perceptions from the survey results. Second, I will present the differences in time spent in each collaboration mode. Finally, I will present the differences in task completion times. Accuracy was not analyzed for effects of experience, since all groups regardless of experience level accomplished the tasks with high accuracy.

5.2.3.1 Task Experience Dependent Survey Results

When analyzing the survey results from first time participants and comparing them to the participants who were using the system and completing the task a second time, the trends nearly reverse. First time participants felt that the *advanced data-synchronization options* technique was easiest to use, most successful at facilitating collaboration, and most liked, followed in each instance by the *data-duplication* technique then the *data-pushing* technique. Second time users, on the other hand, found *data-duplication* easiest to use, followed by *data-pushing* then *advanced data-synchronization options*. Also, second time users found *data-pushing* most successful at facilitating collaboration and most liked, followed by *data-duplication*

TABLE XVII

Task experience dependent results of the participant survey for the formal user study. Answers were scored on a scale of 1 -10, with 1 being worst and 10 being best. There were 32 responses from first time users and 12 responses from second time users.

Survey Question	**Average Score (First Time)**	**Average Score (Second Time)**
Results on Ease of Use for Group Interaction		
Local group	8.78 ± 0.20	9.00 ± 0.30
Remote group using *data-pushing*	6.91 ± 0.42	8.50 ± 0.36
Remote group using *data-duplication*	7.94 ± 0.36	8.67 ± 0.36
Remote group using *advanced data-synchronization options*	8.38 ± 0.22	7.92 ± 0.45
Results on How Successful the Tool was at Facilitating Collaboration		
Local collaboration	8.22 ± 0.38	8.67 ± 0.36
Remote collaboration using *data-pushing*	6.16 ± 0.38	8.58 ± 0.42
Remote collaboration using *data-duplication*	7.72 ± 0.42	8.58 ± 0.40
Remote collaboration using *advanced data-synchronization options*	8.75 ± 0.19	8.08 ± 0.38
Results on How Much Users Liked the Remote Collaboration Features		
Features of *data-pushing*	6.28 ± 0.44	8.58 ± 0.45
Features of *data-duplication*	7.63 ± 0.41	8.50 ± 0.48
Features of *advanced data-synchronization options*	8.50 ± 0.22	7.92 ± 0.48

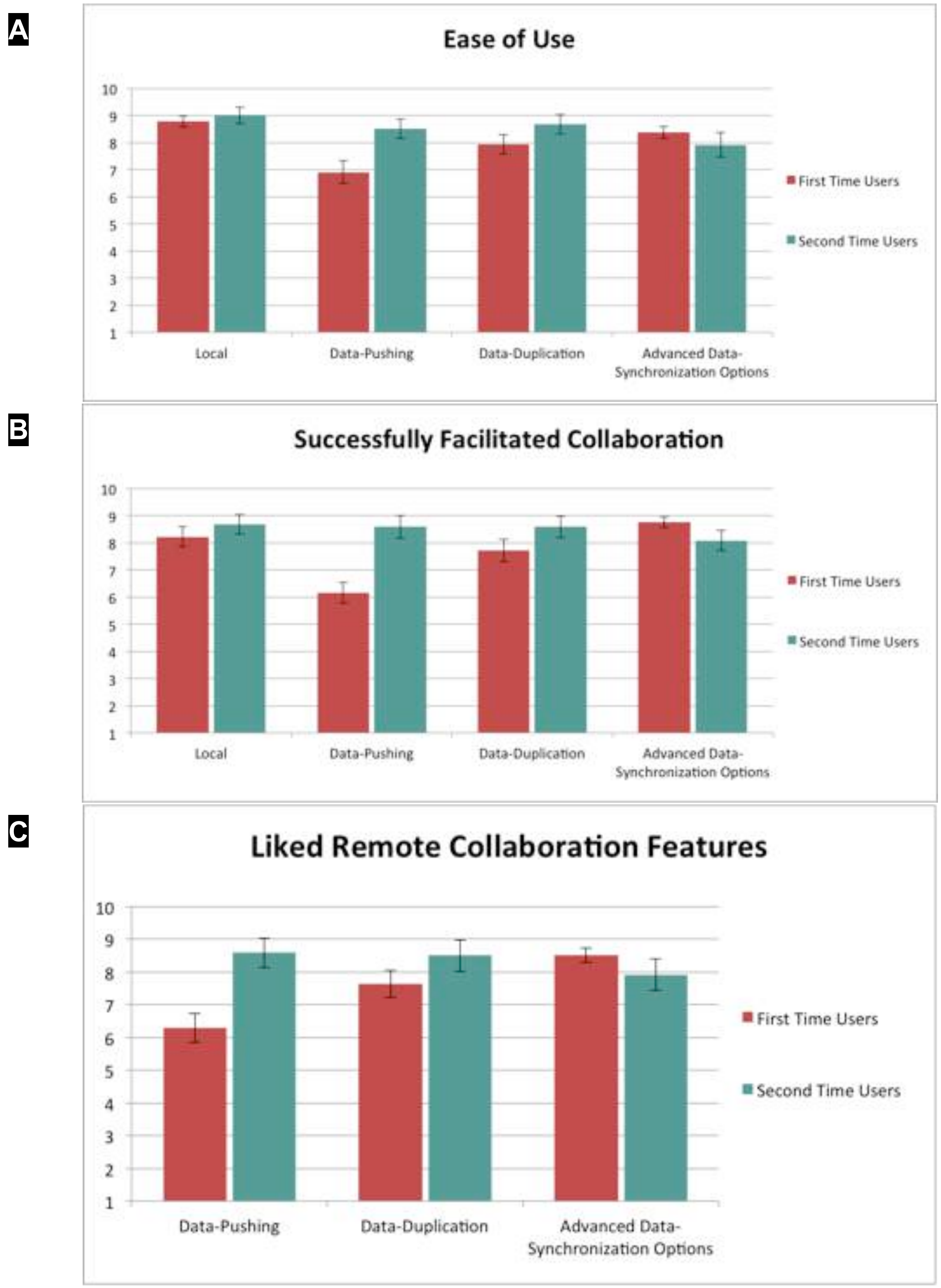


Figure 29. Graphs of task experience dependent survey results depicting answers from both first time users and second time users. Panel A shows reported Ease of Use. Panel B shows how successful users found features at facilitating collaboration. Panel C shows how much users liked the remote collaboration features.

then *advanced data-synchronization options*. I performed a two-tailed t-test to compare the differences, and while trends changed for all three questions, the only significant differences in task experience was that scores for *data-pushing* went up for second time users, with all three questions having p-values < 0.05. Since there were eight groups of first time participants and three groups of second time participants, the first time user group has responses from 32 participants whereas the second time user group has responses from 12 participants. Average scores and their standard errors for the task experience dependent survey results are summarized in Table XVII and visualized in Figure 29.

There were also major differences in how participants ranked the three data sharing techniques based on whether they were completing the task for first or second time. First time participants ranked *advanced data-synchronization options* best, followed closely by *data-duplication*. Second time participants ranked *data-duplication* best, followed by *data-pushing* then *advanced data-synchronization options*. These results are summarized in Table XVIII. A two-tailed t-test was used to examine the differences. Second time users ranked *data-pushing* significantly higher than first time users, and ranked *advanced data-synchronization options* significantly lower than first time users, both with p-values < 0.05.

TABLE XVIII

Average task experience dependent rank for the three data sharing techniques, with 1 being the best and 3 being the worst.

Data Sharing Technique	**Rank [First Time] (1-best, 3-worst)**	**Rank [Second Time] (1-best, 3-worst)**
Data-pushing	2.66 ± 0.10	2.08 ± 0.19
Data-duplication	1.88 ± 0.13	1.58 ± 0.26
Advanced data-synchronization options	1.44 ± 0.12	2.33 ± 0.22

5.2.3.2 Task Experience Dependent Audio / Video Results

When analyzing the coded collaboration modes from the audio / video recordings, one interesting difference stands out between the second time participants and first time participants. Participants who had used the system to complete the task in the past spent more time in the lower modes of collaboration (not collaborating or communicating one-on-one), and therefore less time in the higher modes of collaboration (conferencing or coordinating) than first time participants. However, after comparing the different experience levels with a two-tailed t-test, most differences were not significant. Only the percentage of time spent in one-on-one communication when using *data-pushing* and the percentage of time spent coordinating when using *data-duplication* were significantly different based on task experience, with both p-values < 0.05. Task experience dependent collaboration mode results are summarized in Table XIX and visualized in Figure 30.

TABLE XIX

Average task experience dependent percentage of time spent in each collaboration mode based on data sharing technique.

Data Sharing Technique	Not Collaborating	Communicating	Conferencing	Coordinating
(First Time) Data-pushing	14.6% ± 3.7%	28.1% ± 4.4%	36.2% ± 3.3%	21.0% ± 5.3%
(Second Time) Data-pushing	17.7% ± 4.3%	47.1% ± 4.8%	26.1% ± 4.0%	9.2% ± 3.6%
(First Time) Data-duplication	7.9% ± 3.0%	29.0% ± 5.3%	48.2% ± 4.0%	14.9% ± 3.2%
(Second Time) Data-duplication	17.1% ± 5.6%	38.2% ± 7.1%	42.2% ± 7.7%	2.4% ± 1.5%
(First Time) Advanced data-synchronization options	5.3% ± 1.8%	24.3% ± 5.2%	53.2% ± 4.5%	17.1% ± 4.5%
(Second Time) Advanced data-synchronization options	9.3% ± 6.2%	25.7% ± 5.6%	41.6% ± 4.4%	23.4% ± 8.1%

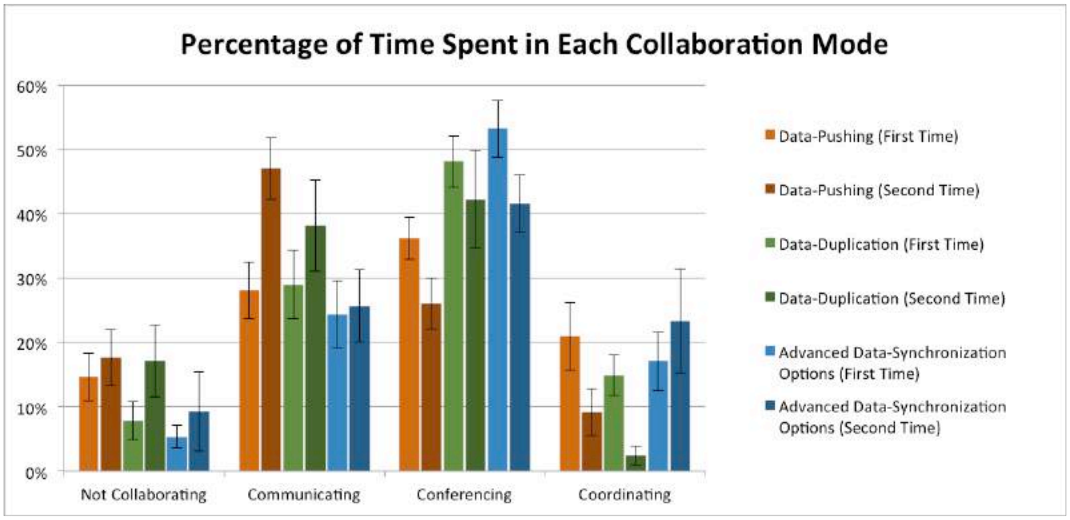


Figure 30. Average task experience dependent percentage of time spent in each collaboration mode based on data sharing technique.

In order to determine if there was any difference with how aware the two teams were of each other based on task experience levels, I compared the amount of time the teams were in the same collaboration mode. This analysis also showed a deterioration of collaboration in second time users. Amount of time that both teams spent in the same collaboration mode dropped for all the data sharing techniques. However, when comparing the experience levels with a two-tailed t-test, there were no significant differences, with all p-values > 0.05. Task experience dependent collaboration mode similarity results are summarized in Table XX and visualized in Figure 31.

TABLE XX

Average task experience dependent percentage of time that the team using the large shared display and the team using the small shared display were in the same collaboration mode as each other based on data sharing technique.

Data Sharing Technique	**Percentage of Time in Same Collaboration Mode (First Time)**	**Percentage of Time in Same Collaboration Mode (Second Time)**
Data-pushing	71.0% ± 6.2%	64.4% ± 6.4%
Data-duplication	81.8% ± 5.0%	65.7% ± 7.0%
Advanced data-synchronization options	90.7% ± 3.3%	79.9% ± 9.4%

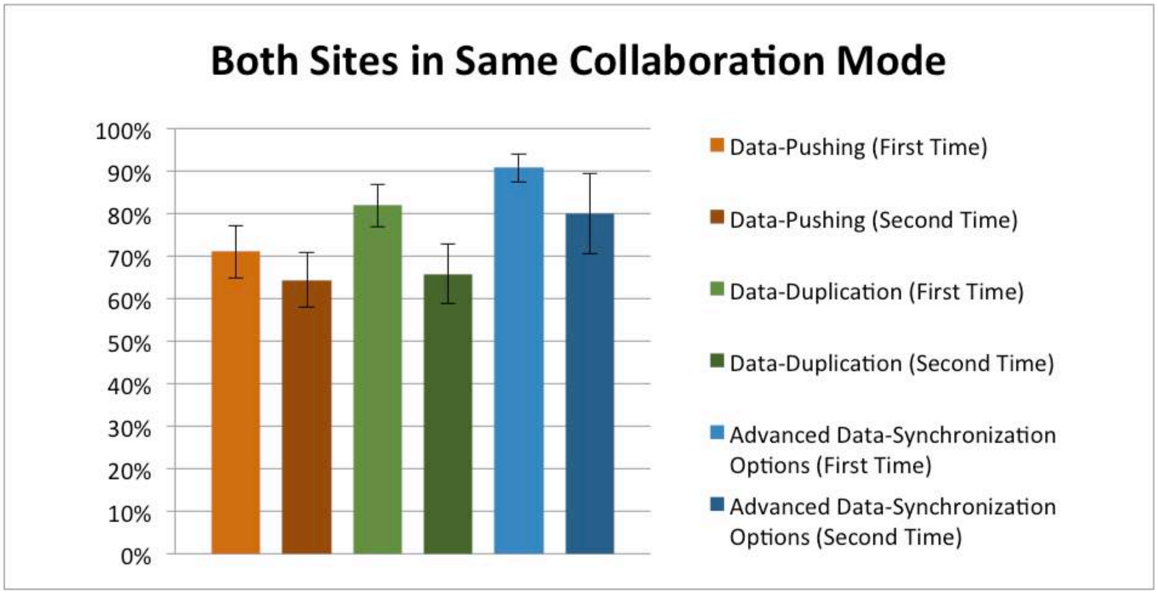


Figure 31. Average task experience dependent percentage of time that the team using the large shared display and the team using the small shared display spent in the same collaboration mode as each other based on data sharing technique.

5.2.3.3 Task Experience Dependent Completion Time Results

To investigate the effects task experience had on completion time, I separated the results from the first time users and the second time users. Not surprisingly, both tasks were completed faster for all three data sharing techniques when second time users were participating. Again all three techniques were completed in nearly the same time on average, whether looking at the first time users or the second time users. Therefore, data sharing technique did not have an effect on completion time regardless of experience level. Table XXI summarizes the completion times for both tasks for first and second time users. These results are visualized in Figure 32.

TABLE XXI

Average task experience dependent completion times (in minutes and seconds) for task 1 and task 2 of the formal user study based on data sharing technique.

Data Sharing Technique	**Completion Time of Task 1 (First Time)**	**Completion Time of Task 1 (Second Time)**	**Completion Time of Task 2 (First Time)**	**Completion Time of Task 2 (Second Time)**
Data-pushing	14:42 ± 1:42	6:35 ± 1:43	3:43 ± 0:41	1:39 ± 0:24
Data-duplication	14:59 ± 2:56	8:46 ± 2:50	3:08 ± 0:41	1:41 ± 0:29
Advanced data-synchronization options	14:31 ± 2:25	8:28 ± 1:09	3:23 ± 0:40	1:39 ± 0:33

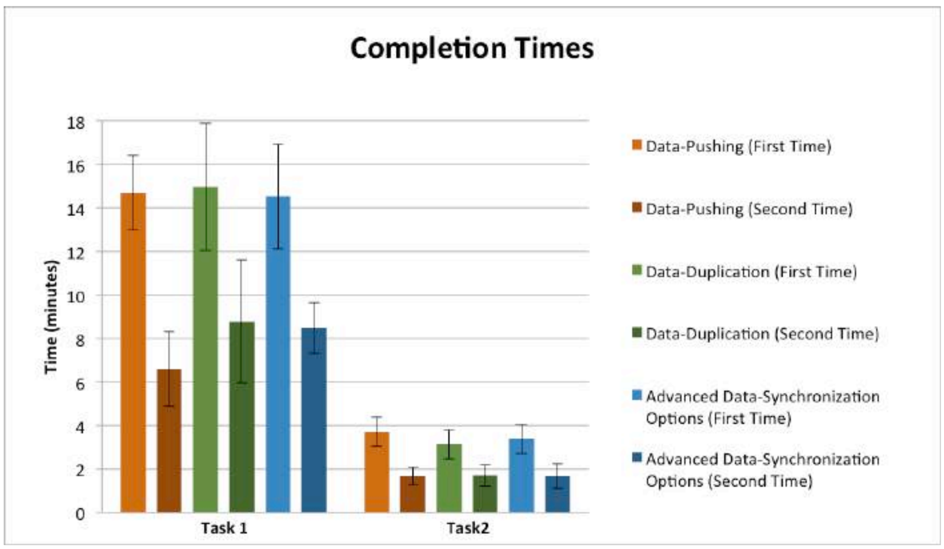


Figure 32. Average task experience dependent completion times for task 1 and task 2 of the formal user study based on data sharing technique.

I also looked at the effects order had on completion time, with respect to task experience. While with first time users there was a clear learning curve that enabled them to complete each run quicker than the previous one, this pattern did not exist with second time users. Furthermore, the completion times for second time users were barley quicker than the final run for first time users. This indicates that after using the tools and performing three runs in the study, that users had learned efficient methods for accomplishing the task collaboratively. Table

XXII summarizes the completion times based on technique order for both tasks for first and second time users. These results are visualized in Figure 33.

TABLE XXII

Average task experience dependent completion times (in minutes and seconds) for task 1 and task 2 of the formal user study based on data sharing technique order.

Technique Order	**Completion Time of Task 1 (First Time)**	**Completion Time of Task 1 (Second Time)**	**Completion Time of Task 2 (First Time)**	**Completion Time of Task 2 (Second Time)**
Technique 1	18:40 ± 2:14	9:19 ± 0:59	4:09 ± 0:38	2:05 ± 0:21
Technique 2	15:12 ± 2:21	6:38 ± 0:18	2:57 ± 0:36	1:14 ± 0:24
Technique 3	10:20 ± 1:21	7:53 ± 3:18	3:07 ± 0:43	1:41 ± 0:30

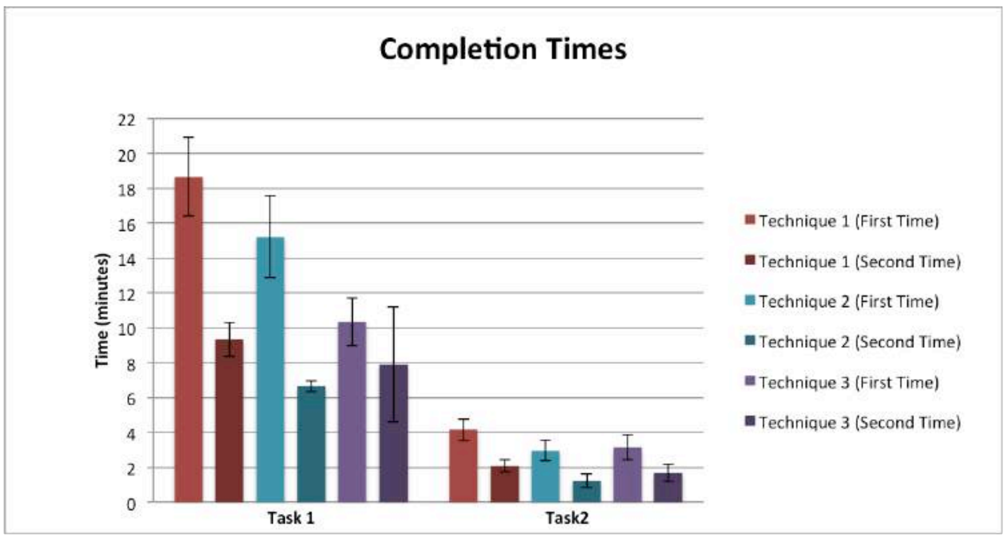


Figure 33. Average task experience dependent completion times for task 1 and task 2 of the formal user study based on data sharing technique order.

5.2.3.4 Differences in Experience Level

There are a few important differences in how participants perceived the data sharing techniques based on whether they had already performed the task before or not. First, *advanced data-synchronization options* scored highest on all three questions for first time users, whereas it scored last on all three questions for second time users. Additionally, second time users scored the *data-pushing* technique significantly higher than the first time users for all three questions, with all p-values < 0.05. Looking a little closed at the scores, it becomes clear that second time users raised the scores of both *data-pushing* and *data-duplication* while keeping the scores of *advanced data-synchronization options* nearly stagnant.

However, by analyzing the audio / video data coded for collaboration modes, the perception change by second time users is not corroborated. The *data-pushing* technique still resulted in teams engaged in lower modes of collaboration significantly more frequently than either of the two continuous data synchronization techniques. Also, the *advanced data-synchronization options* technique still resulted in significantly more time spent with both teams in the same collaboration mode than the *data-pushing* technique.

Second time users also completed the tasks much faster on average than first time users. This is likely due to familiarity with both the task and the tools. The change in perception of ranking all three data sharing techniques nearly identical could be due to familiarity with the data and the fact that the task became repetitive. Also, because the second time users spent less time using with each technique, it may have become more difficult to distinguish between them. However, the actual data about how the group collaborated still supports the notion of continuously synchronized applications leading to enhanced collaboration between PDTs. Greater familiarity could also explain the general dip in collaboration with second time users. Since they were familiar with the task and the tools, second time users may have been able to accomplish the tasks more independently.

CHATPER 6

CONCLUSION

The advent of big-data has introduced the need for tools that can allow researchers to access, visualize, and explore their data. As we enter a world that is more globally connected than ever before, these researchers have the ability to collaborate with experts around the world. However, the physical distance between collaborators has created a barrier that has fueled the research and development of technology that can enable groups of people connect and share data. While existing commercial software has begun to address the issues surrounding remote collaboration, many of the solutions are limited to a particular task or a particular set of hardware.

This dissertation has sought to create a more flexible solution that provides a platform for researchers to share arbitrary data and collaborate in real-time with other remotely located teams. Beyond developing a technical solution, this dissertation has sought to provide an understanding on how PDTs work together, and how the synchronization of data being manipulated in shared applications affects the quality of the collaboration. I have demonstrated, through experimental studies, that providing multi-user applications that can be synchronized with a remote site can improve the quality of collaboration for both teams and improve awareness of the remote collaborators. These results provide insight and design implications for creating flexible data-conferencing software.

This chapter concludes the dissertation by outlining the main contributions as well as providing potential areas of future research that would continue to enhance collaboration between distributed teams of experts.

6.1 Contributions

This dissertation aimed to explore three main questions regarding remote collaboration between PDTs:

1) Does providing continuously synchronized applications improve the quality of collaboration and awareness of the remote team?
2) Can synchronizing applications provide remote teams the same quality of collaboration as local teams?
3) Does the size of a shared display have an effect on the quality of collaboration and awareness of the remote team?

In order to answer these questions I performed both a longitudinal user study and a formal user study between PDTs. In both studies, I had a group of participants use data-conferencing groupware with and without continuous synchronization features enabled. Additionally, in both user studies, one team used a large shared display (approximately 25 Mpixels) while the other team used a small shared display (approximately 8 Mpixels). These studies allowed me to investigate the effects of data synchronization, the differences between the local team and the remotely collaborating group as a whole, and the effects that display size had on each team.

6.1.1 Effects of Data Synchronization on Remote Collaboration

Data synchronization significantly improved collaboration between remote teams in multiple ways. First, users perceived that systems with synchronized data were significantly easier to use, better at facilitating collaboration, and more liked. This supported my first hypothesis – that continuous data synchronization would lead to enhanced collaboration. Users did not perceive any significant differences between the two techniques that provided continuous synchronization. This was likely due to the fact that each technique had its own unique pros and cons. When asked for any further comments, users provided some insights on this topic. For the *data-duplication* technique, users felt that being able to see the cursors of their remote collaborators was extremely helpful, writing comments such as:

> *"Seeing the cursors of the collaborators granted a feeling of locality…"*
> *"I liked being able to see the remote pointers – very helpful."*

However, users did feel constrained in that they couldn't work in parallel as well, in particular having to take turns on who was panning and zooming the shared map:

> *"… people were more restrained in the pan and zoom business."*

For the *advanced data-synchronization options* technique, users felt that being able to have independent control over certain items, such as map pan and zoom, while keeping other items synchronized, such as marker locations, was extremely helpful:

> *"Advanced Data-Sync works well on these displays because you could have one window open to make [both] local and global changes."*

However, users felt confused at times about which properties were and were not currently synchronized:

> *"… you could go a while w/o syncing & [get] startled when another member moved [the application] in a way you weren't expecting"*

When investigating the data from the audio / video analysis, it supports participants' perception for the most part. Using the continuous synchronization techniques, users spent less time in the lower modes of collaboration (not collaborating or only communicating one-one-one) and more time in the higher mode of collaboration (conferencing as a group or coordinating the efforts of the group between the distributed teams). In this instance though, the *advanced data-*

synchronization options technique was significantly better overall than the *data-duplication* technique as well as the *data-pushing* technique. Additionally, using the continuous synchronization techniques led to the two teams being more inline, resulting in more time spent in the same collaboration mode as each other. Again, the *advanced data-synchronization options* technique outperformed the *data-duplication* technique as well as the *data-pushing* technique.

6.1.2 Comparing Remote Collaboration to Local Collaboration

The collaboration between each local team was reported as superior to the remote collaboration in terms of ease of use regardless of which data sharing technique was being used. However, this was not a significant difference for either *data-duplication* or *advanced data-synchronization options*. When looking at how participants scored how successful the collaboration was however, local collaboration did not rank the best. Participants actually found the *advanced data-synchronization options* technique most successful, though nearly identical to the local collaboration. Again there were not significant differences between local collaboration and the remote collaboration when using *data-duplication* or *advanced data-synchronization options*.

These results support my second hypothesis by indicating that for the most part local collaboration was still superior to remote collaboration, but that enabling continuous data synchronization techniques makes the distributed workflow easier. Also, the fact that the *advanced data-synchronization options* technique actually scored slightly higher on its ability to facilitate collaboration than the local tools can perhaps be attributed to the fact that its partial synchronization features can't even be utilized between local participants – there is no way for an SRSD to share an application with itself. This feature was shown to have helped coordinated workflows between distributed teams, but was unavailable for the local teams to uses amongst themselves.

6.1.3 Effects of Display Size on Remote Collaboration

The formal user study elucidated some interesting differences between teams using a large shared display and teams using a small shared display. When glancing at the results of the participant survey or the coded collaboration modes from the audio / video recordings, it may appear that large shared displays actually hinder collaboration. However, the increased lack of collaboration can actually be attributed to the large shared display making it easier for teams to analyze complex data and complete their task. Since they were working with a team that did not have the same technological affordances, they were forced to wait for the other team to catch up more frequently. It is therefore expected that PDTs with both sites having a large shared display would be better able to analyze the data and complete the tasks faster than PDTs with both sites having a small shared display.

Also, large shared display participants' perception of the collaboration suffered in comparison to the small display participants. The fact that the teams using the large shared display ended up waiting more frequently could also explain this drop in scores on the participant survey. Therefore the results of this study line up with the work outlined in Chapter 2 that provide evidence that large display environments enable collaboration and significantly amplify the way users make sense of large-scale, complex data.

These results from the survey showed that participants using the small shared display preferred the *data-duplication* technique. However, the collaboration modes coded from the audio/video recordings showed an improved collaboration when using the *advanced data-synchronization options* technique. Therefore my third hypothesis, that *advanced data-*

synchronization options would be preferable for participants using a small shared display, remains inconclusive.

6.2 Design Implications

Synchronization of data is a key factor for real-time collaboration between PDTs. Data-conferencing software should be designed to allow for continuous synchronization of shared content that can be manipulated by any participant. Awareness of the remote collaborators and the state of a shared application are both important factors for successful data conferencing. Visually representing the remote collaborators, such as showing their cursor, will lead to an improved presence. Having a visual representation of the state of an application that is only partially synchronized could decrease confusion.

Since the *advanced data-synchronization options* technique was the top all around performer, user controlled partial synchronization techniques have been shown to help successfully facilitate collaboration. Groupware should incorporate partial synchronization features that can be utilized by either remotely collaborating teams or co-located individuals. These partial synchronization techniques should be dynamic and user controlled so that any participant of a collaborative session can adjust the synchronization settings to best fit the needs of the group throughout a collaborative session.

Overall data-conferencing software should be flexible. Distributed teams collaborate in many ways. Status report meetings, coordinated search tasks, and cooperative exploration may require a different set of tools. Each type of collaboration requires its participants to communicate in different ways. Providing users with groupware that contains a set of instruments not specifically designed for just one task enables them to leverage necessary features and improve collaboration across distance.

6.3 Areas of Future Research

The work done in this dissertation does have certain limitations. All data sharing techniques are one-to-one, enabling two SRSDs to share data with each other. In order to communicate between three or more sites, multiple one-to-one sharing connections would be required. Moving forward, the *data-duplication* and *advanced data-synchronization options* techniques could be expanded to allow for any application to be shared with any number of other SRSDs. This way PDTs located in three or more locations could all see and work on synchronized copies of the same applications. Also, the shared portal created in the *data-duplication* technique has a fixed physical size, meaning that when either side scales the portal larger or smaller it must maintain its aspect ratio. Since shared display spaces have a wide variety of configurations, it would be beneficial to have a more flexible portal window that can freely resize, updating both is visible size as well as its shared physical size. Finally, the *advanced data-synchronization options* technique could benefit from two main improvements, which were elucidated by user comments. In order to provide remote collaborator awareness, it would be beneficial to add an abstract notion of cursor location. Cursor location is not as simple as the *data-duplication* scenario since the shared application may not be fully synchronized, and therefore be showing a different visual representation of the data. Also, to avoid confusion, it would be beneficial to have a persistent visual indicator of the current synchronization state, so that teams are aware of what their remote partners are viewing. This visual indicator could also serve to identify when unsynchronized aspects of an application are going to be resynchronized.

The aspects of collaboration that I focused my research on were asymmetric synchronous collaboration scenarios for interdisciplinary coordinated work across distance. More research would be needed to test the principles learned in this dissertation to see if the lessons could be applied to groups working on the same task, or teams who are not working in real-time with each other. Additionally, a study focused on type of task could help elucidate whether certain forms of data synchronization are best suited for certain tasks (i.e. status reports, search and analysis tasks, exploration of unknown data, etc.). Finally, it would be interesting to investigate the long-term effects of using a flexible data-conferencing system between PDTs of domain experts in authentic work scenarios. While this was attempted by studying the SAGE2 Development meetings, this setting consisted of experts in the same field at both locations and for the most part ideas were simply being shared during the collaborative sessions.

One aspect of collaboration that was not measured in these studies was how productive participants were during each mode of collaboration. It would be interesting to reanalyze and code the videos based on slightly modified criteria. Instead of just breaking collaboration mode into communication, conferencing, or coordinating, it could bring forth interesting data to investigate the quality of collaboration that occurred during each of those modes. For example, in a group conference how often were the teams in consensus and how often were they in conflict with each other. This could help determine a more exact type of communication that leads to enhanced collaboration, and in turn could lead to developing tools and techniques that would assist collaborators achieve those types of communication more often.

This dissertation contributes to the field of computer supported cooperative work. However, there still remains a wide array of challenges and research topics that remain open for investigation. While the above future areas of research are not an exhaustive list, they present a few next steps to help continue improving how people coordinate work with each other.

6.4 Final Remarks

The results of this research are a demonstration of the fluidity of people, who can work together even across great distance. I have argued that computer supported data synchronization techniques fundamentally impact how people collaborate across distance. The techniques described in this thesis have shown that distance is not a barrier that can't be overcome. However, this research only contributes a portion of knowledge necessary to enable truly seamless data-conferencing with distributed teams using heterogeneous technologies.

CITED LITERATURE

[1] Google+ Hangout. Retrieved July 7, 2014 from http://www.google.com/+/learnmore/hangouts/.

[2] Skype. Retrieved July 7, 2014 from http://www.skype.com/en/what-is-skype/.

[3] WebEx. Retrieved July 7, 2014 from http://www.webex.com/.

[4] Oblong Industries, Inc., “Introducing Mezzanine: the Future of Conference Room Collaboration,” *White Paper*. 2013.

[5] T. Marrinan, J. Aurisano, A. Nishimoto, K. Bharadwaj, V. Mateevitsi, L. Renambot, L. Long, A. Johnson, and J. Leigh, "SAGE2: A New Approach for Data Intensive Collaboration Using Scalable Resolution Shared Displays," in *Proceedings of the IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing* (CollaborateCom ‘14), 2014, pp 177-186.

[6] L. Renambot, A. Rao, R. Singh, J. Byungil, N. Krishnaprasad, V. Vishwanath, V. Chandrasekhar, N. Schwarz, A. Spale, C. Zhang, G. Goldman, J. Leigh, and A. Johnson, “SAGE: the Scalable Adaptive Graphics Environment,” in *Workshop on Advanced Collaborative Environments* (WACE ‘04). 2004.

[7] B. Jeong, L. Renambot, R. Jagodic, R. Singh, J. Aguilera, A. Johnson, and J. Leigh, “High-performance dynamic graphics streaming for scalable adaptive graphics environment,” in *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing* (SC ‘06). 2006.

[8] J. Leigh, L. Renambot, A. Johnson, R. Jagodic, H. Hur, E. Hofer, and D. Lee, “Scalable Adaptive Graphics Middleware for Visualization Streaming and Collaboration in Ultra Resolution Display Environments,” in *Proceedings of Workshop on Ultrascale Visualization* (UltraVis ‘08). 2008, pp 47-54.

[9] C. Johnson, R. Ross, S. Ahern, J. Ahrens, W. Bethel, K. L. Ma, M. Papka, J.V. Rosendale, H. W. Shen, and J. Thomas, “Visualization and knowledge discovery: Report from the DOE/ASCR workshop on visual analysis and data exploration at extreme scale”, Salt Lake City. 2007.

[10] J. Cummings, T. Finholt, I. Foster, C. Kesselman, and K.A. Lawrence, “Beyond being there: A blueprint for advancing the design, development, and evaluation of virtual organizations.” 2008.

[11] C. Andrews, A. Endert, and C. North, “Space to think: large high-resolution displays for sensemaking,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI ‘10). 2010, pp 55-64.

[12] R. Ball, and C. North, “Analysis of user behavior on high-resolution tiled displays,” in *Proceedings of the 2005 IFIP TC13 international conference on Human-Computer Interaction* (INTERACT ‘05), 2005, pp 350-363.

[13] R. Ball, and C. North, “Effects of tiled high-resolution display on basic visualization and navigation tasks,” in *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (CHI EA ‘05). 2005, pp 1196-1199.

[14] R. Ball, C. North, and D. A. Bowman, “Move to improve: promoting physical navigation to increase user performance with large displays,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI ‘07). 2007, pp 191-200.

[15] X. Bi, and R. Balakrishnan, “Comparing usage of a large high-resolution display to single or dual desktop displays for daily work,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI ‘09). 2009, pp 1005-1014.

[16] 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*. 2003, vol. 3, pp 9-16.

[17] D. S. Tan, J. K. Stefanucci, D. R. Proffitt, and R. Pausch, “The Infocockpit: providing location and place to aid human memory,” in *Proceedings of the 2001 workshop on Perceptive user interfaces* (PUI ‘01). 2001, pp 1-4.

[18] D. S. Tan, D. Gergle, P. Scupelli, and R. Pausch, “With similar visual angles, larger displays improve spatial performance,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI ‘03). 2003, pp 217-224.

[19] B. Yost, Y. Haciahmetoglu, and C. North, “Beyond visual acuity: the perceptual scalability of information visualizations for large displays,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI ‘07). 2007, pp 101-110.

[20] C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon, J. C. Hart, “The CAVE: audio visual experience automatic virtual environment,” in *Communications of the ACM*. 1992, vol. 35 no. 6, pp 64-72.

[21] C. Cruz-Neira, D. J. Sandin, Thomas A. DeFanti, “Surround-screen projection-based virtual reality: the design and implementation of the CAVE,” in *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques* (SIGGRAPH '93). 1993, pp 135-142.

[22] PowerWall. July 14, 2014 from http://www.lcse.umn.edu/research/powerwall/powerwall.html.

[23] G. Humphreys, I. Buck, M. Eldridge, and P. Hanrahan, “Distributed rendering for scalable displays,” in *Proceedings of the 2000 ACM/IEEE conference on Supercomputing* (SC ‘00). 2000, no. 30.

[24] T. A. DeFanti, J. Leigh, M. D. Brown, D. J. Sandin, O. Yu, C. Zhang, R. Singh, E. He, J. Alimohideen, N. K. Krishnaprasad, R. Grossman, and Marco Mazzucco, “Teleimmersion and Visualization with the OptIPuter,” in *Proceedings of the 12th International Conference on Artificial Reality and Telexistence* (ICAT 2002). 2002.

[25] G. Humphreys, M. Eldridge, I. Buck, G. Stoll, M. Everett, and P. Hanrahan, “WireGL: a scalable graphics system for clusters,” in *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques* (SIGGRAPH ‘01). 2001, pp 129-140.

[26] G. Humphreys, M. Houston, R. Ng, R. Frank, S. Ahern, P. D. Kirchner, and J. T. Klosowski, “Chromium: a stream-processing framework for interactive rendering on clusters,” in *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques* (SIGGRAPH ‘02). 2002, pp 693-702.

[27] S. Eilemann, M. Makhinya, and R. Pajarola, “Equalizer: A Scalable Parallel Rendering Framework,” *IEEE Transactions on Visualization and Computer Graphics* (TVCG), 2009, vol. 15 no. 3, pp 436-452.

[28] K. Doerr and F. Kuester, “CGLX: A Scalable, High-Performance Visualization Framework for Networked Display Environments,” *IEEE Transactions on Visualization and Computer Graphics* (TVCG), 2011, vol. 17 no. 3, pp 320-332.

[29] A. Febretti, A. Nishimoto, V. Mateevitsi, L. Renambot, A. Johnson, and J. Leigh, “Omegalib: A multi-view application framework for hybrid reality display environments,” in *Proceedings of IEEE Virtual Reality* (VR ‘14). 2014, pp 9-14.

[30] S. Yokoyama and H. Ishikawa, “Parallel distributed rendering of HTML5 canvas elements,” in *Proceedings of the 11th International Conference on Web Engineering* (ICWE ‘11). 2011, pp 331-345.

[31] N. Schwarz, S. Venkataraman, L. Renambot, N. Krishnaprasad, V. Vishwanath, J. Leigh, A. Johnson, G. Kent, and A. Nayak, “Vol-a-Tile - A Tool for Interactive Exploration of Large Volumetric Data on Scalable Tiled Displays,” in *Proceedings of the Conference on Visualization* (VIS ‘04). 2004.

[32] D. Lee, S. A. Munson, B. Congleton, M. W. Newman, M. S. Ackerman, E. C. Hofer, and T. A. Finholt, “Montage: a platform for physically navigating multiple pages of web content,” in *CHI '09 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '09). 2009, pp 4477-4482.

[33] G. P. Johnson, G. D. Abram, B. Westing, P. Navrátil, and K. Gaither, “DisplayCluster: An Interactive Visualization Environment for Tiled Displays,” in *Proceedings of IEEE International Conference on Cluster Computing* (CLUSTER ‘12). 2012, pp 239-247.

[34] LambdaVision. Retrieved July 18, 2014 from http://www.evl.uic.edu/cavern/lambdavision/.

[35] HIPerWall. Retrieved July 18, 2014 from http://hiperwall.calit2.uci.edu/.

[36] HyperWall-2. Retrieved July 18, 2014 from http://www.nas.nasa.gov/hecc/resources/viz_systems.html.

[37] HIPerSpace. Retrieved July 18, 2014 from http://vis.ucsd.edu/mediawiki/index.php/Research_Projects:_HIPerSpace.

[38] TACC Visualization Resources. Retrieved July 18, 2014 from https://www.tacc.utexas.edu/resources/visualization.

[39] OptIPresence. Retrieved July 18, 2014 from http://vis.ucsd.edu/mediawiki/index.php/Research_Projects:_OptIPresence_Tele-Immersion_Testbed.

[40] Reality Deck. Retrieved July 18, 2014 from http://labs.cs.sunysb.edu/labs/vislab/reality-deck-home/.

[41] JavaScript. Retrieved August 7, 2015 from http://www.w3.org/standards/webdesign/script.

[42] WebGL. Retrieved August 7, 2015 from https://www.khronos.org/webgl/.

[43] WebRTC. Retrieved August 7, 2015 from http://www.webrtc.org/.

[44] Node.js. Retrieved July 7, 2014 from http://nodejs.org/.

[45] ImageMagick. Retrieved August 12, 2015 from http://www.imagemagick.org/script/index.php.

[46] FFmpeg. Retrieved August 12, 2015 from https://www.ffmpeg.org/.

[47] ExifTool. Retrieved August 12, 2015 from http://www.sno.phy.queensu.ca/~phil/exiftool/.

[48] J. Leigh, A. Johnson, L. Renambot, T. Peterka, B. Jeong, D. Sandin, J. Talandis, R. Jagodic, S. Nam, H. Hur, and Y. Sun, “Scalable resolution display walls,” in *Proceedings of the IEEE*. 2013, vol. 101 no. 1, pp 115-129.

[49] Omicron. Retrieved February 13, 2014 from http://github.com/uic-evl/omicron/.

[50] D. C. Engelbart, “Augmenting Human Intellect: A Conceptual Framework,” *Summary Report for Director of Information Sciences, Air Force Office of Scientific Research*. 1962.

[51] D. C. Engelbart and W. K. English, “A Research Center for Augmenting Human Intellect,” in *AFIPS Conference Proceedings of the 1968 Fall Joint Computer Conference*. 1968, vol. 33, pp 395-410.

[52] Lotus Development Corporation, “Groupware - Communication, Collaboration and Coordination.” 1995.

[53] S. Elrod, R. Bruce, R. Gold, D. Goldberg, F. Halasz, W. Janssen, D. Lee, K. McCall, E. Pedersen, K. Pier, J. Tang, and B. Welch, “Liveboard: a large interactive display supporting group meetings, presentations, and remote collaboration,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI ‘92). 1992, pp 599-607.

[54] R. MacKenzie, K. Hawkey, K. S. Booth, Z. Liu, P. Perswain, and S. S. Dhillon, “LACOME: a multi-user collaboration system for shared large displays,” in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work Companion* (CSCW ‘12). 2012, pp 267-268.

[55] M. Rittenbruch, “CubIT: large-scale multi-user presentation and collaboration,” in *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces* (ITS ‘13). 2013, pp 441-444.

[56] J. T. Biehl, W. T. Baker, B. P. Bailey, D. S. Tan, K. M. Inkpen, and M. Czerwinski, “Impromptu: a new interaction framework for supporting collaboration in multiple display environments and its field evaluation for co-located software development,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI ‘08). 2008, pp 939-948.

[57] K. Cheng, J. Li, and C. Müller-Tomfelde, “Supporting interaction and collaboration on large displays using tablet devices,” in *Proceedings of the International Working Conference on Advanced Visual Interfaces* (AVI ‘12). 2012, pp 774-775.

[58] C. Müller-Tomfelde, K. Cheng, and J. Li, “Pseudo-direct touch: interaction for collaboration in large and high-resolution displays environments,” in *Proceedings of the 23rd Australian Computer-Human Interaction Conference* (OzCHI ‘11). 2011, pp 225-228.

[59] K. Ponto, K. Doerr, T. Wypych, J. Kooker, and F. Kuester, “CGLXTouch: A multi-user multi-touch approach for ultra-high-resolution collaborative workspaces,” in *Future Generation Computer Systems* (FGCS). 2001, vol. 27 no. 6, pp 649-656.

[60] D. Lowet and D. Goergen, “Co-browsing dynamic web pages,” in *Proceedings of the 18th International Conference on World Wide Web* (WWW ‘09). 2009, pp 941-950.

[61] E. Arroyo, V. Righi, R. Tarrago, and J. Blat, “A remote multi-touch experience to support collaboration between remote museum visitors,” in *Proceedings of the 13th IFIP TC 13 International Conference on Human-Computer Interaction - Volume Part IV* (INTERACT ‘11). 2011, pp 462-465.

[62] M. Kuechler and A. M. Kunz, “Collaboard: a remote collaboration groupware device featuring an embodiment-enriched shared workspace,” in *Proceedings of the 16th ACM International Conference on Supporting Group Work* (GROUP ‘10). 2010, pp 211-214.

[63] S. Yarosh, A. Tang, S. Mokashi, and G. D. Abowd, “"almost touching": parent-child remote communication using the sharetable system,” in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (CSCW ‘13). 2013, pp 181-192.

[64] R. Ocker, M. Rosson, D. Kracaw, and S. Hiltz, “Training Students to Work Effectively in Partially Distributed Teams,” in *ACM Transactions on Computing Education* (TOCE). 2009, vol. 9 no. 1, article 6.

[65] H. Kim and S. Snow, “Collaboration on a large-scale, multi-touch display: asynchronous interaction and multiple-input use,” in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion* (CSCW ‘13). 2013, pp 165-168.

[66] N. Yamashita, H. Kuzuoka, K. Hirata, S. Aoyagi, and Y. Shirai, “Supporting fluid tabletop collaboration across distances,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI ‘11). 2011, pp 2827-2836.

[67] A. D. Balakrishnan, S. R. Fussell, and S. Kiesler, “Do visualizations improve synchronous remote collaboration?,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI ‘08). 2008, pp 1227-1236.

[68] J. R. Epps and B. S. Close, “A study of co-worker awareness in remote collaboration over a shared application,” in *CHI ‘07 Extended Abstracts on Human Factors in Computing Systems* (CHI EA ‘07). 2007, pp 2363-2368.

[69] M. R. Morris and J. Teevan, “Collaborative Web Search: Who, What, Where, When, and Why,” A Publication in the Morgan & Claypool Publishers series *Synthesis Lectures on Information Concepts, Retrieval, and Services*. 2010.

[70] B. Evans and E. Chi, “Towards a Model of Understanding Social Search,” in *Proceedings of Computer Supported Cooperative Work* (CSCW '08). 2008, pp 485-494.

[71] J. Rama and J. Bishop, “A survey and comparison of CSCW groupware applications,” in *Proceedings of the 2006 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries* (SAICSIT '06). 2006, pp 198-205.

[72] J. Dyck, C. Gutwin, N. Graham, and D. Pinelle, “Beyond the LAN: Techniques from network games for improving groupware performance,” in *Proceedings of the 2007 International ACM Conference on Supporting Group Work* (GROUP ‘07). 2007, pp 291-300.

[73] K. F. White and W. G. Lutters, “Structuring Cross-Organizational Knowledge Sharing,” in *Proceedings of the International ACM Conference on Supporting Group Work* (GROUP ‘07). 2007, pp 187-196.

[74] E. Hornecker, P. Marshall, and Y. Rogers, “From entry to access: how shareability comes about,” in *Proceedings of the Conference on Designing Pleasurable Products and Interfaces* (DPPI ‘07). 2007, pp 328-342.

[75] Dimitracopoulou, “Designing Collaborative Learning Systems: Current Trends & Future Research Agenda,” in *Proceedings of the Conference on Computer Support for Collaborative Learning: Learning 2005: the Next 10 Years!* (CSCL '05). 2005, pp 115-124.

[76] A. Hayes and K. Krippendorff, “Answering the Call for a Standard Reliability Measure for Coding Data,” in *Communication Methods and Models*. 2007, pp 77-89.

APPENDICES

APPENDIX A

Permission to use previously published materials by IEEE (https://www.ieee.org/publications_standards/publications/rights/permissions_faq.pdf).

- **Does IEEE require individuals working on a thesis or dissertation to obtain formal permission for reuse?**

  The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you must follow the requirements listed below:

  Textual Material

  Using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.

  In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.

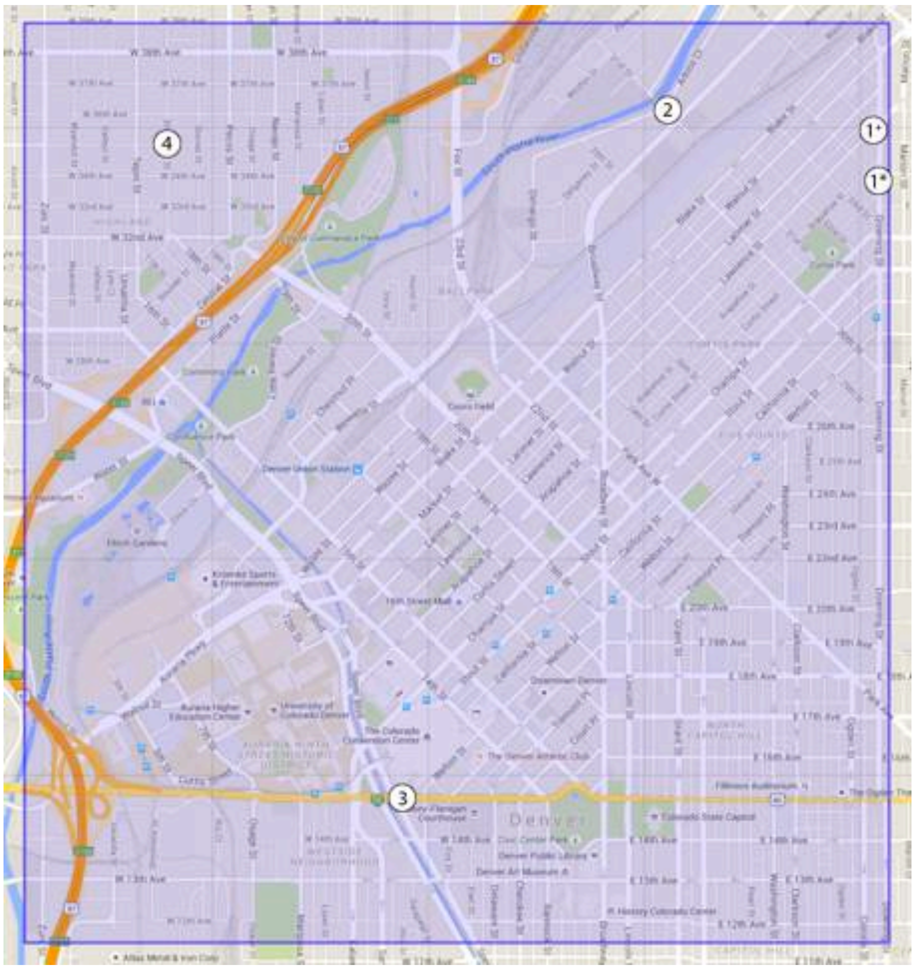
  If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

T. Marrinan, J. Aurisano, A. Nishimoto, K. Bharadwaj, V. Mateevitsi, L. Renambot, L. Long, A. Johnson, and J. Leigh, "SAGE2: A New Approach for Data Intensive Collaboration Using Scalable Resolution Shared Displays," in *Proceedings of the IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing* (CollaborateCom '14), 2014, pp 177-186. © 2014 IEEE.

APPENDIX B

Data given to participants of the formal user study on a sheet of paper. This data served as "prior knowledge" for the tasks. Numbers correspond to approximate location on the map.

Team A - Denver



1. Coffee Shop on 34th St. and Downing St. (*) moved a few blocks to 35th St. and Lawrence St. (+).

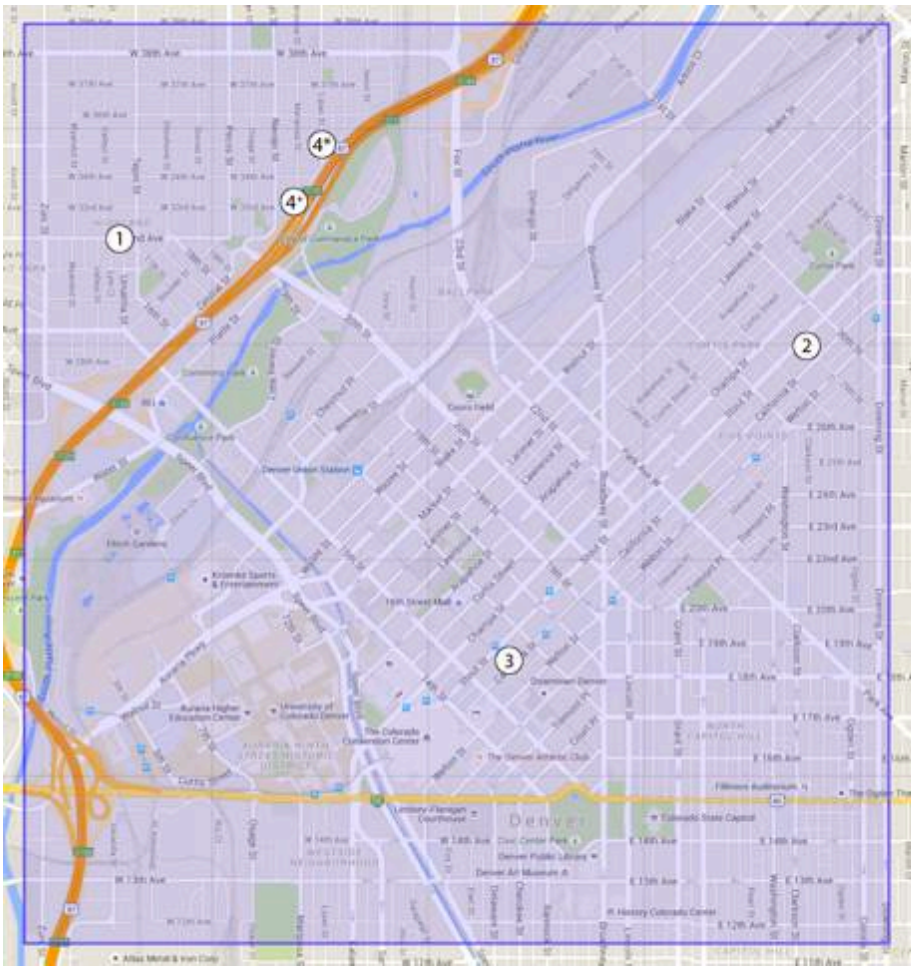
2. New Coffee Shop opened 2 weeks ago on the S corner of 31st St. and Arkins Ct.

3. Construction on Highway 70, West of Broadway St., finished 6 months ago adding another lane in each direction.

4. 35th Ave. between Zuni St. and Navajo St. is currently closed for construction

APPENDIX B (Continued)

Team B – Denver



1. Parking lot built last week on W 32nd Ave. between Vallejo St. and Tejon St.

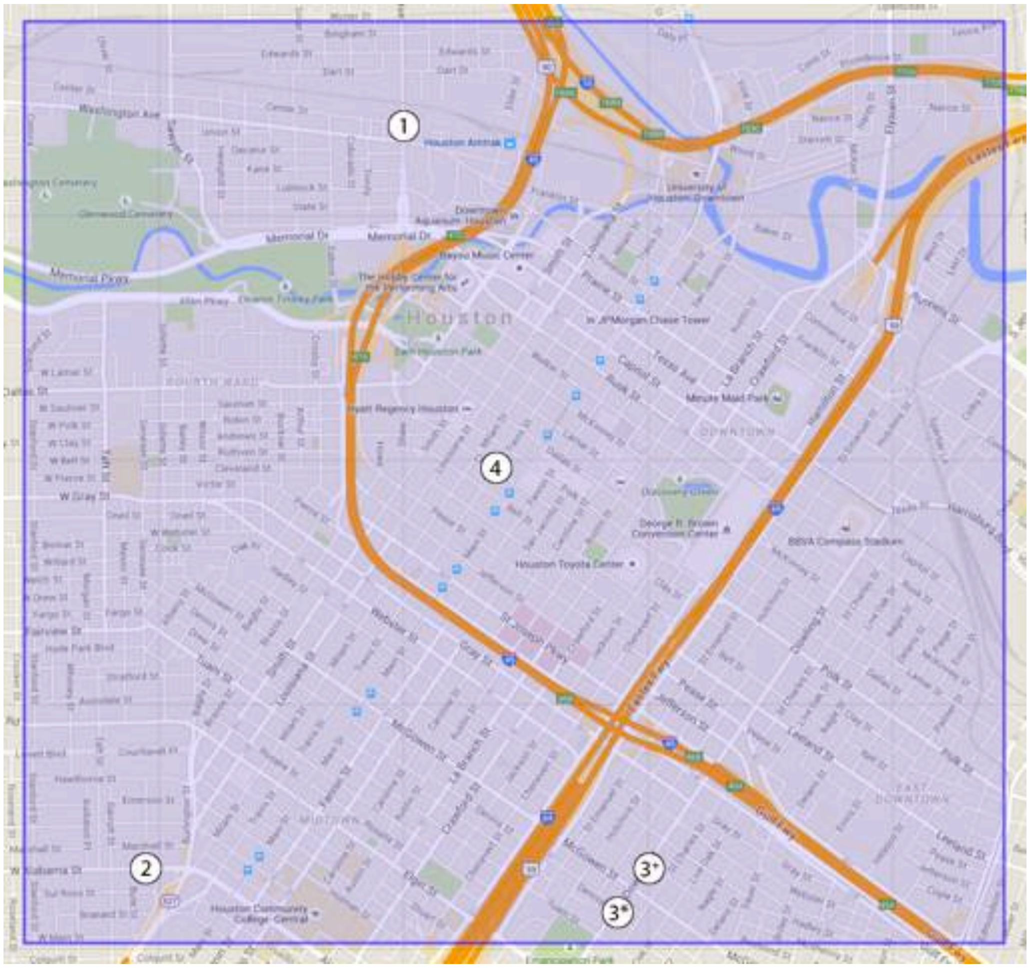
2. Parking lot on Stout St. demolished 6 weeks ago to expand the park between 28th St. and 30th St.

3. Building next to the Donut Shop on 16th St. and California St. recently updated its roof to white.

4. Donut Shop on 35th St. and Lipan St. (*) moved a few blocks down to the corner of 33rd Ave. and Kalamath St. (+).

APPENDIX B (Continued)

Team A – Houston



1. Houston Ave. between Edwards St. and Lubbock St. is currently closed for construction.

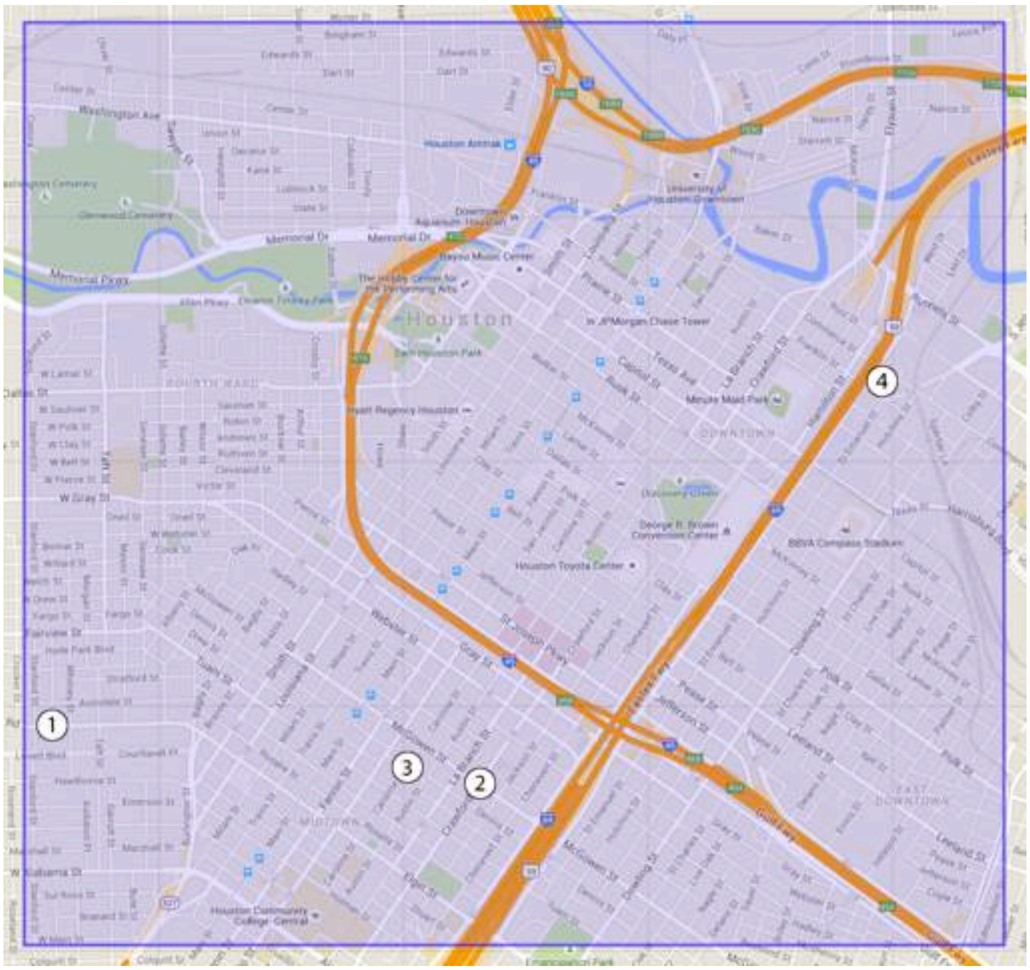
2. Construction on W. Alabama St. finished 2 months ago adding another lane in each direction.

3. Coffee Shop on Dowling St. and Dennis St. (*) moved a couple blocks to Dowling St. and McIlhenny St. (+).

4. Travis St. became a two-way street between Pease St. and McKinney.

APPENDIX B (Continued)

Team B – Houston



1. Buildings on Westheimer Rd. between Stanford St. and Whitney St. updated their roofs one month ago – now are red.

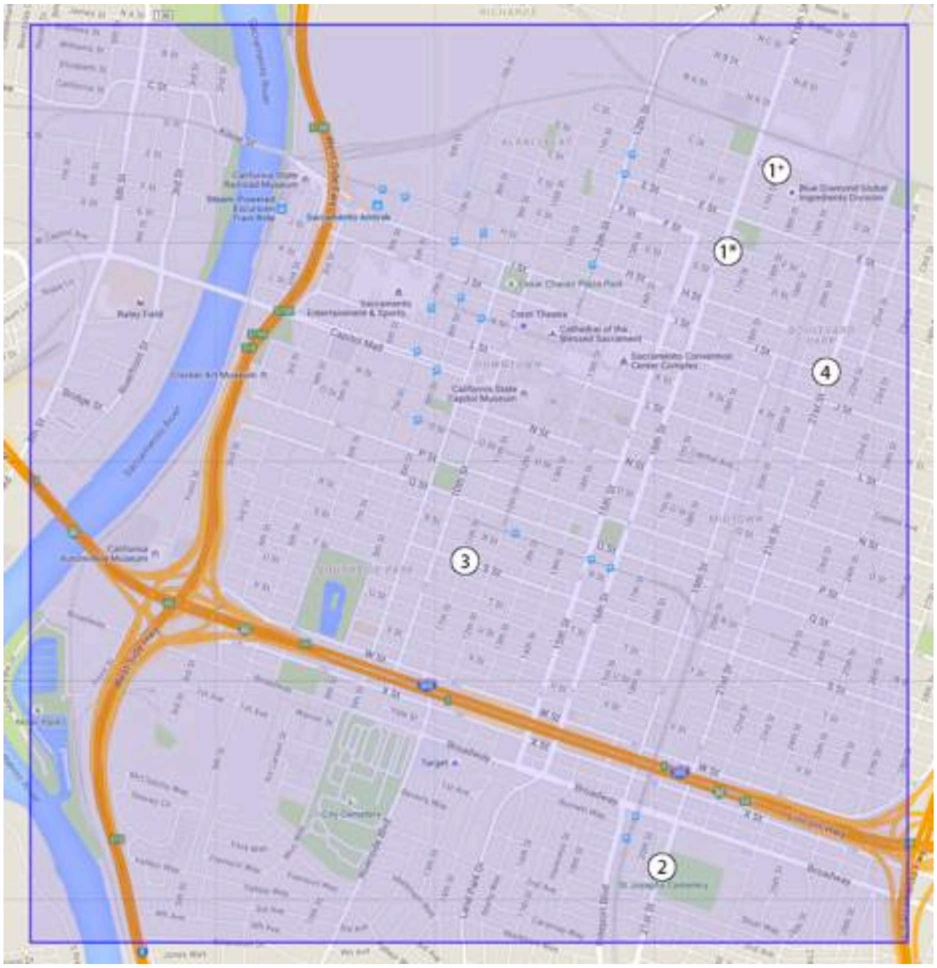
2. Parking lot off of McGowen St. and Crawford St. demolished 2 weeks ago.

3. Building on E corner Dennis St. and Caroline St. updated its roof – now is black.

4. Donut Shop on Chartres St. and Commerce St. moved one block to the NE corner of Franklin St. and St. Emanuel St.

APPENDIX B (Continued)

Team A – Sacramento



1. Coffee Shop off of 16th St. between F St. and Eggplant Alley (*) recently relocated to the SW corner of 17th St. and C St. (+).

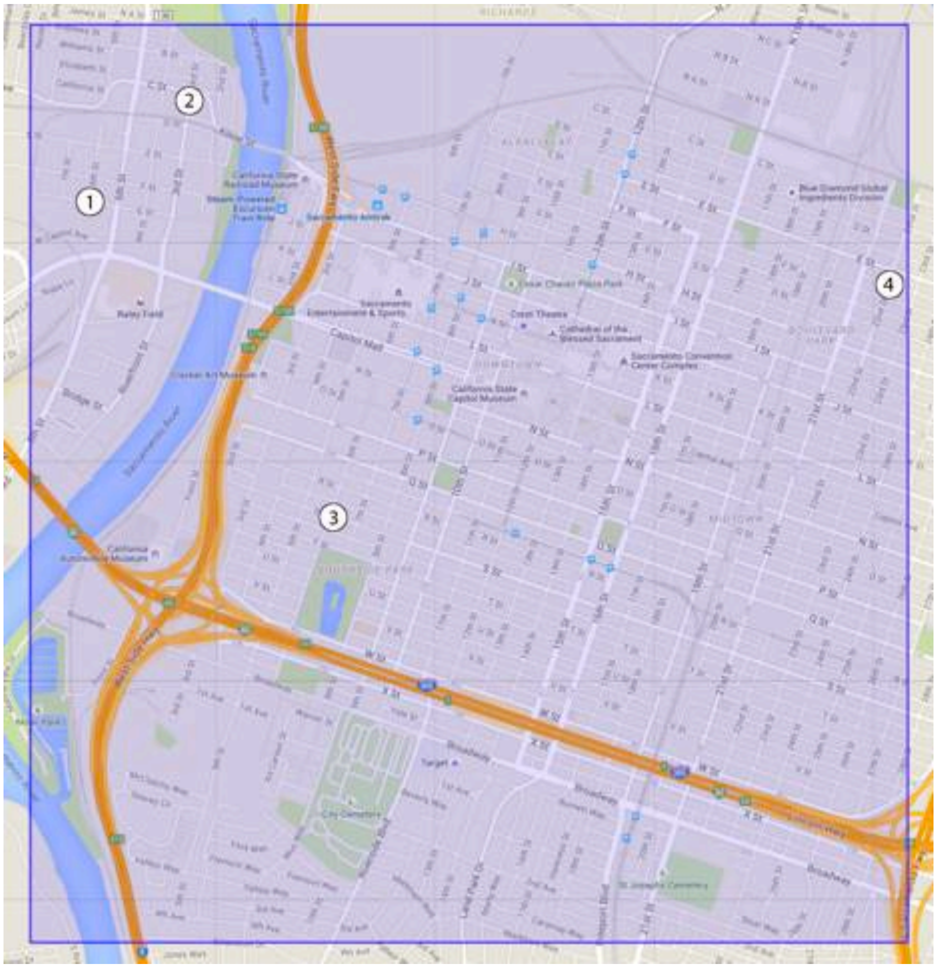
2. 21st St. added another lane in each direction South of Broadway.

3. S St. between 9th St. and 13th St. is currently closed for construction.

4. New Coffee shop just opened yesterday on the SW corner of I St. and 21st St.

APPENDIX B (Continued)

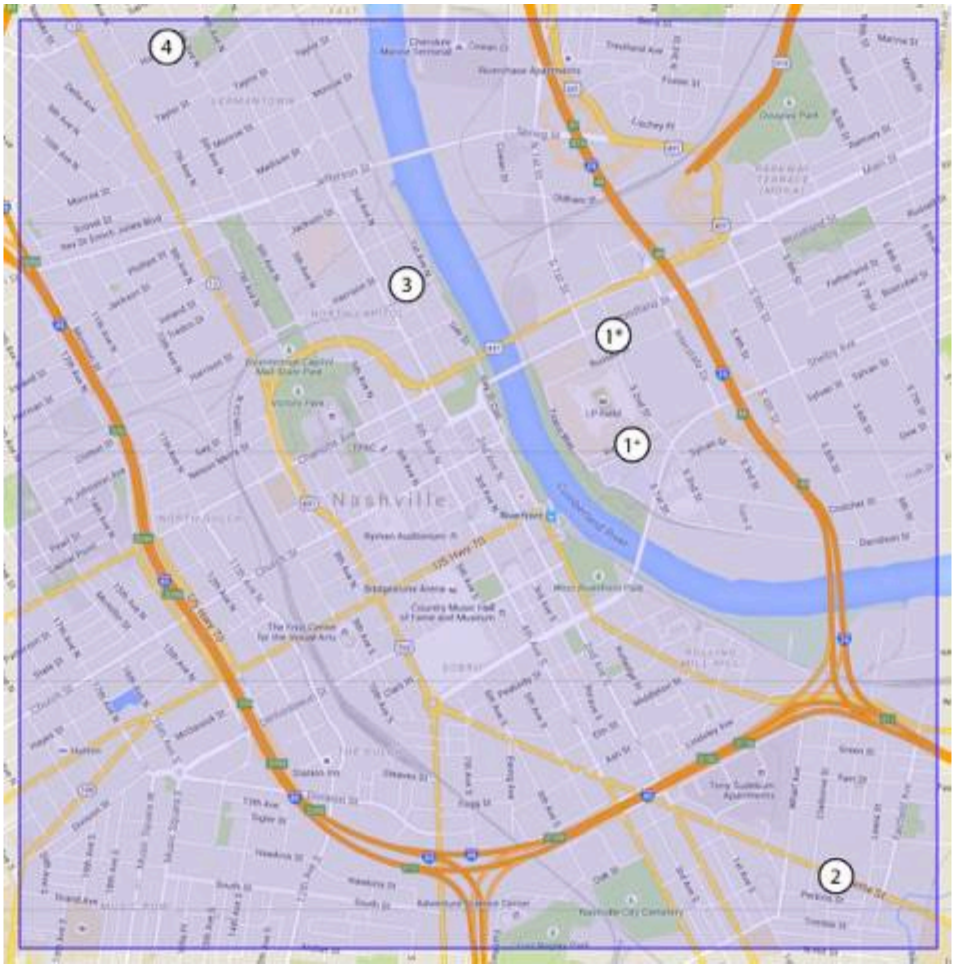
Team B – Sacramento



1. Buildings on G St. between 5th St. and 7th St. updated their roofs three weeks ago – now are white.
2. Parking lot off of 3rd St and C St. demolished 1 month ago.
3. Donut Shop on 6th St. and S St. moved down the block to the corner of 6th St. and Salons Alley.
4. New Donut Shop opened 3 weeks ago at the NE corner of 22nd St. and Eggplant Alley.

APPENDIX B (Continued)

Team A – Nashville



1. Coffee Shop on S. 2nd St. between Woodland St. and Russell St. (*) moved to Victory Ave. and S. 1st St. 2 months ago. (+).

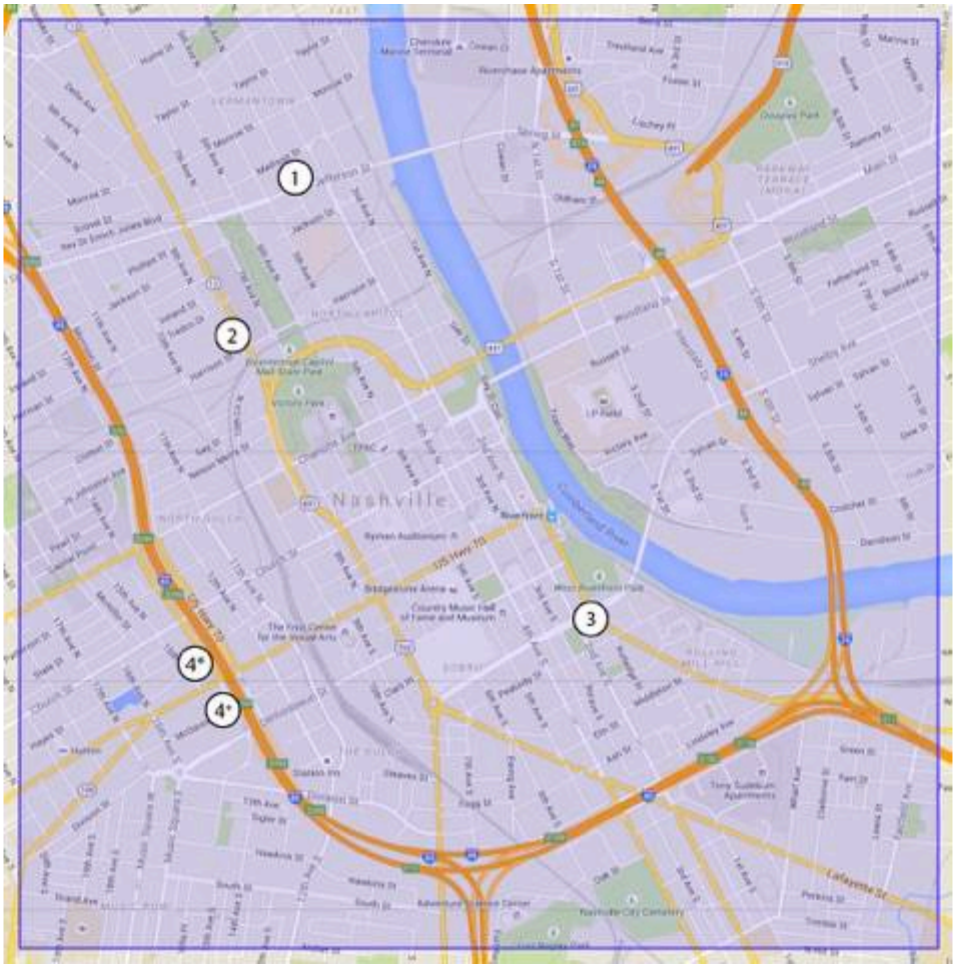
2. Lafayette St. between Wharf Ave. and Fairfield Ave. is closed for construction for the next 6 months.

3. 2nd Ave. N. between Jackson St. and 431 has added a lane in each direction.

4. Coffee Shop on Hume St. between 6th Ave. N. and 7th Ave. N. moved a few blocks East to SE corner of Hume St. and 5th Ave. N.

APPENDIX B (Continued)

Team B – Nashville



1. Buildings on the corner of 4th Ave N and Jefferson St. updated their roofs to green three weeks ago.

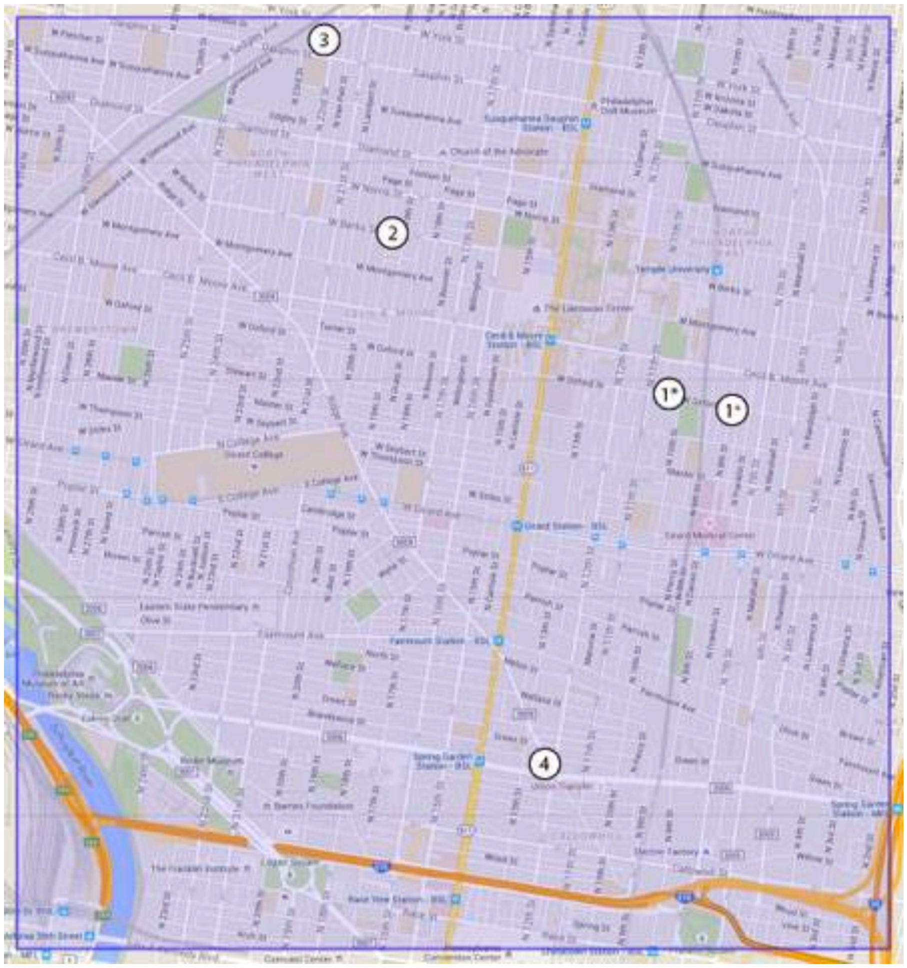
2. The large building on Harrison St. and Rosa L Parks Blvd. / Road 12 and the Nashville Farmers Market updated their roofs to black two months ago.

3. Parking lot on the SW corner of Korean Veterans Blvd. and 24 was demolished 3 weeks ago.

4. Donut Shop on Grundy St. between 14th Ave. S. and 15th Ave. N. (*) moved a few blocks South to the SW corner of McGavock St. and 14th Ave. S. (+).

APPENDIX B (Continued)

Team A – Philadelphia



1. Coffee Shop on W. Oxford St. and Chesapeake Pl. (*) moved to W. Oxford St. and N. 8th St. (+).

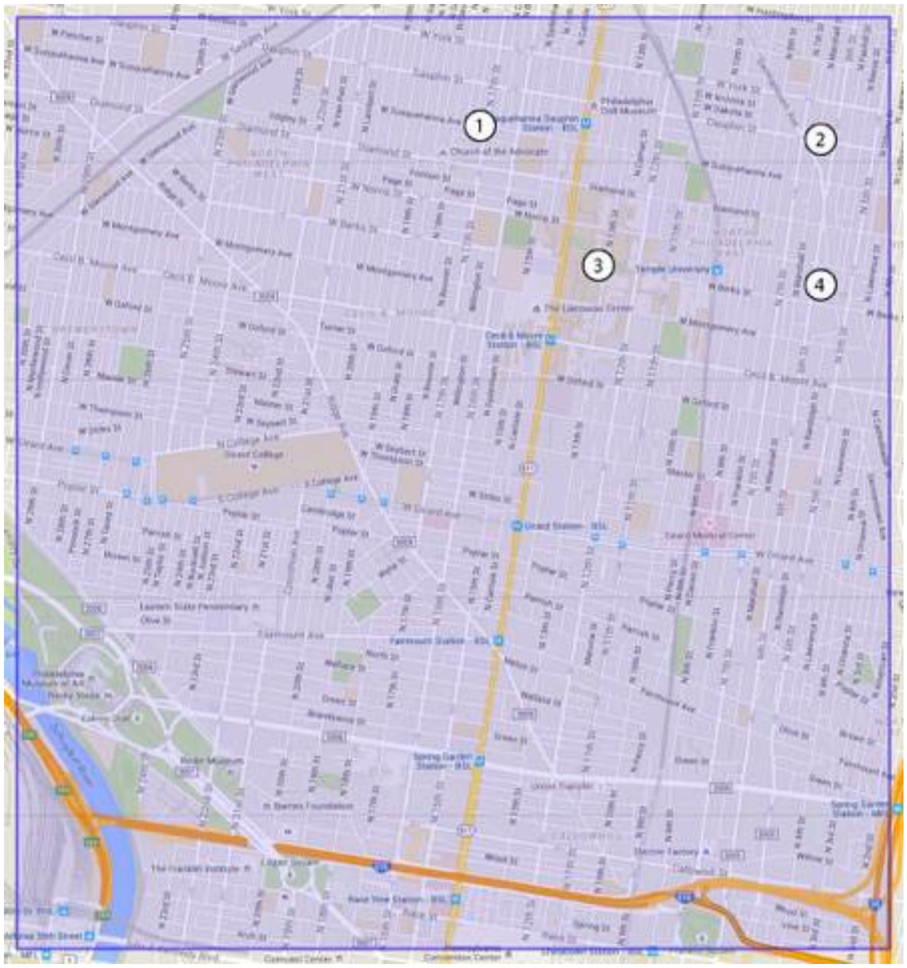
2. W. Berks St. between N. 17th St. and N. 22nd St. became a 2-way street 4 months ago.

3. W. Glenwood Ave. between Dauphin St. and W. York St. is closed for construction for the next year.

4. Coffee Shop on the corner of Spring Garden St. and Ridge Ave. moved a block North to the SW corner of Green St. and N. 12th St 3 weeks ago.

APPENDIX B (Continued)

Team B – Philadelphia



1. Buildings on W. Susquehanna Ave. and French St. between N. 17th St. and N. 18th St. updated their roofs to white 6 months ago.

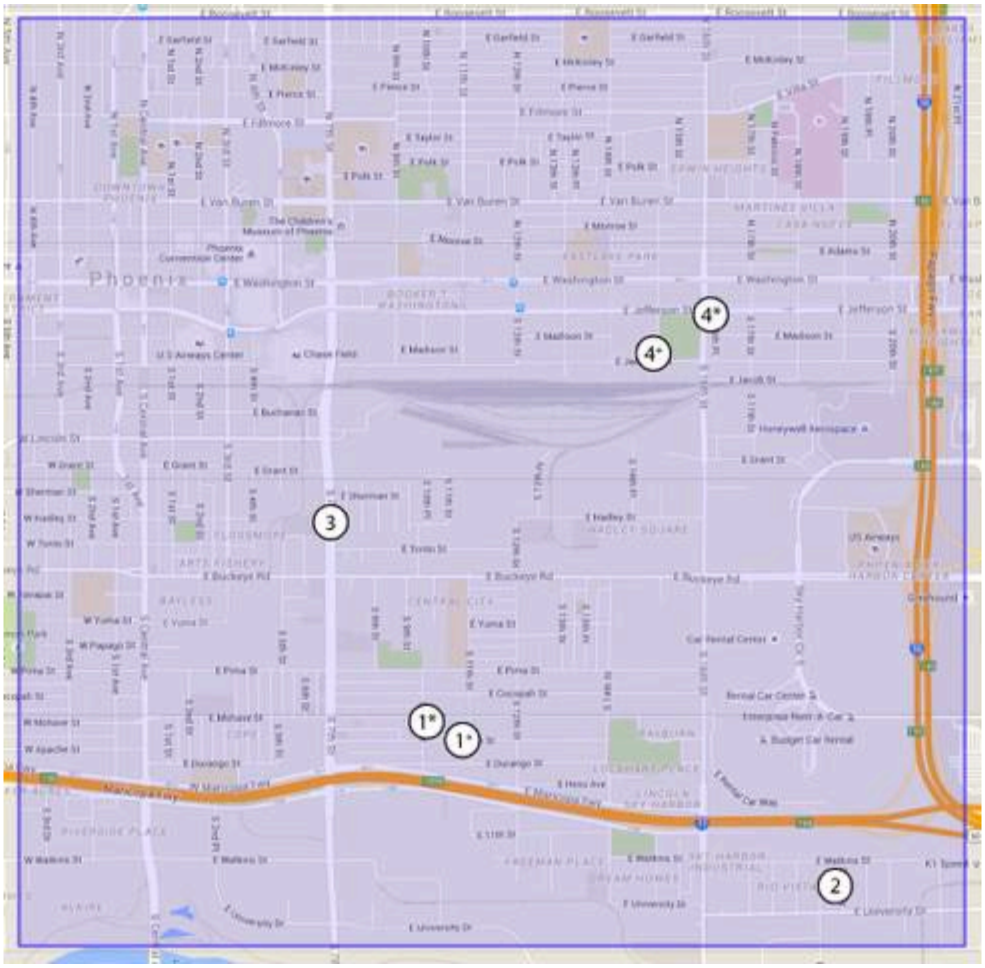
2. Parking lot finished construction on NE corner of Dauphin St. and N. Marshall St.

3. Donut Shop just opened on the SE corner of N. 13th St. and Polett Walk two weeks ago.

4. The large building on 6th St. between W. Norris St. and W. Berks St. updated its roof to red one month ago.

APPENDIX B (Continued)

Team A – Phoenix



1. Coffee Shop on E. Mohave St. between S. 7th St. and S. 10th St. move to the NW corner of E. Apache St. and S. 11th St one month ago.

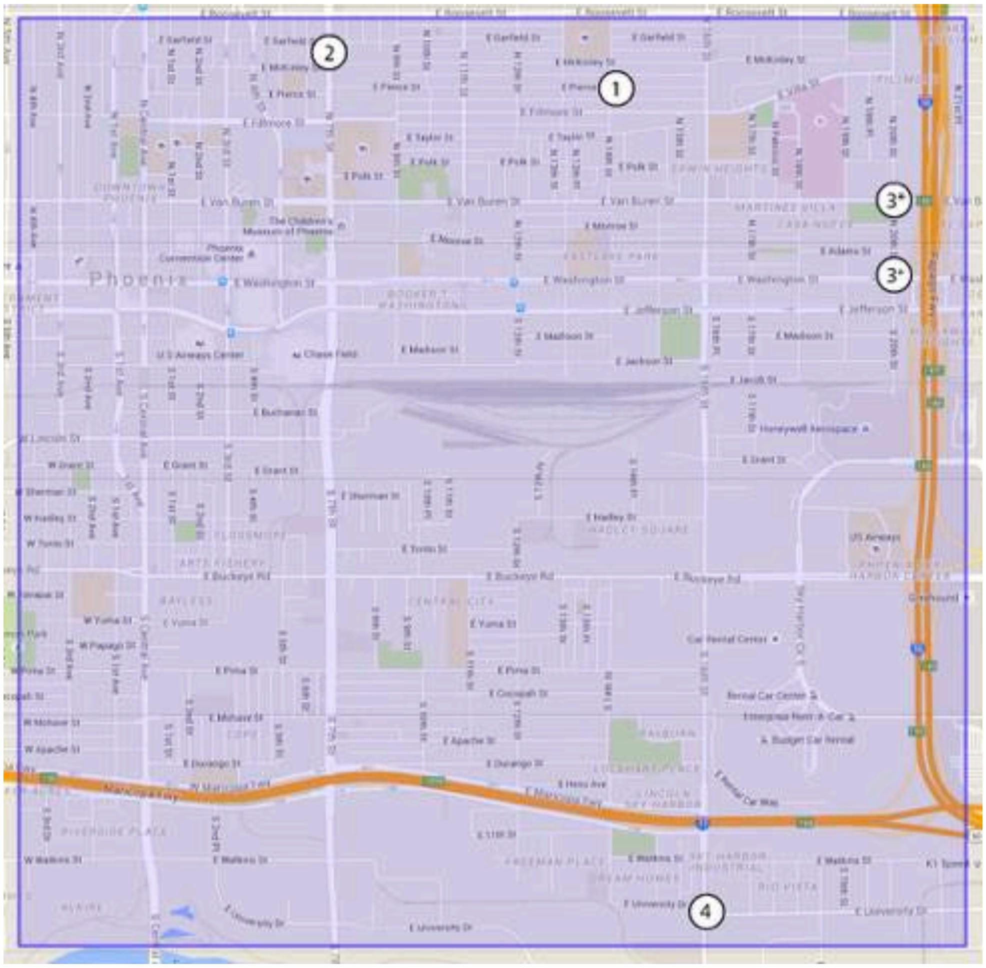
2. S. 17th St. through S. 21st St. added a lane in each direction between E. Watkins St. and E. University Dr.

3. S. 7th St. is closed for construction between E. Lincoln St. and E. Buckeye Rd. for the next 6 months.

4. Coffee Shop on E. Jefferson and S. 16th St. moved 1 block South West to the NW corner of E. Jackson St. and S. 15th St.

APPENDIX B (Continued)

Team B – Phoenix



1. A new parking lot just finished construction and is now open on the SW corner of N. 14th St. and E. Pierce St.

2. The buildings on both sides of 7th St. between E. Roosevelt St. and E. Pierce St. updated their roofs to black 2 weeks ago.

3. Donut Shop on E. Van Buren St. and N. 20th St. move a few blocks South to E. Washington St. and N. 20th St. 4 months ago.

4. The parking lot on the corner of E. University Dr. and S. 16th St. was demolished 2 months ago.

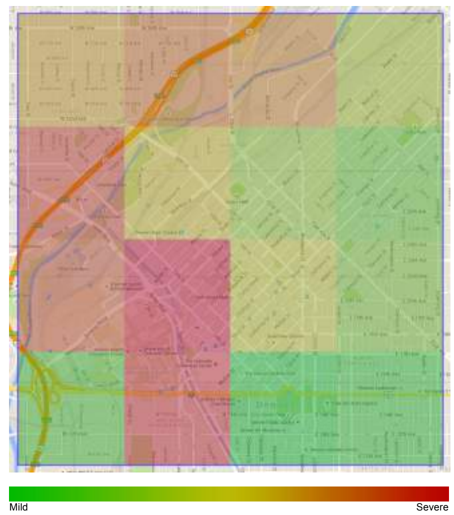
APPENDIX C

Extra data given to participants for task 2 of the formal user study.

Denver – Team A

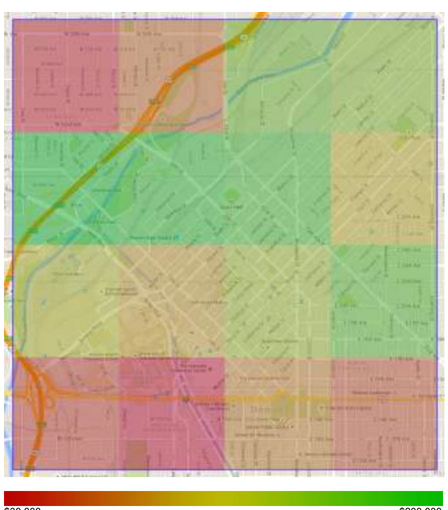
Crime Rates:

Storm Damage:

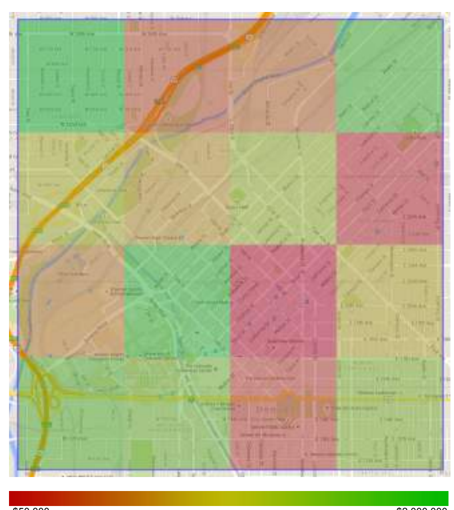


Denver – Team B

Average Family Income:



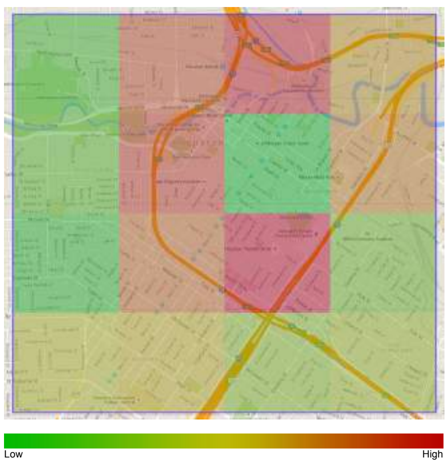
Average Business Profits:



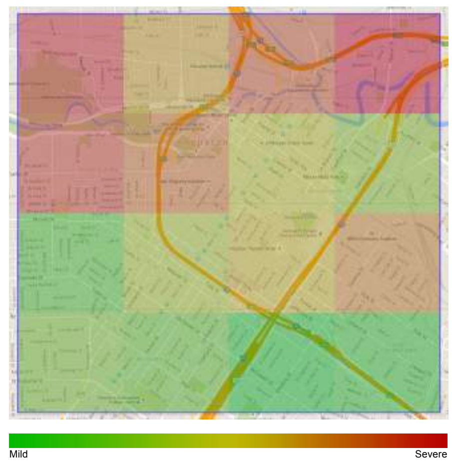
APPENDIX C (Continued)

Houston – Team A

Crime Rates:

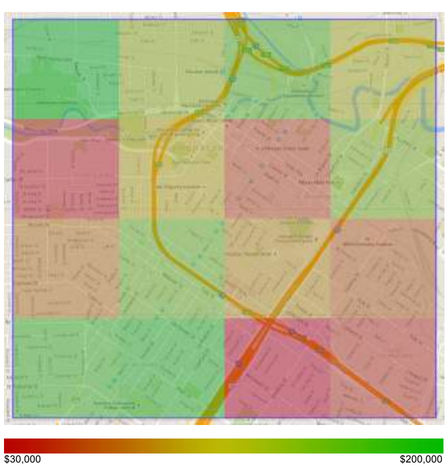


Storm Damage:

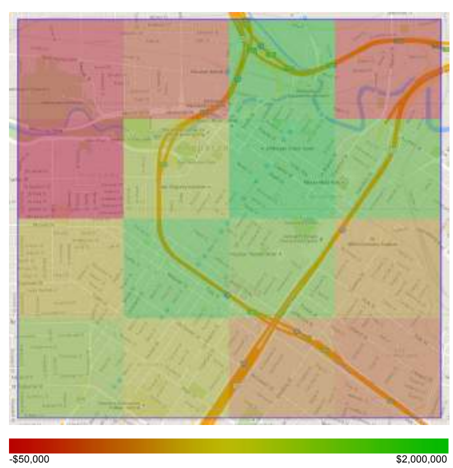


Houston – Team B

Average Family Income:



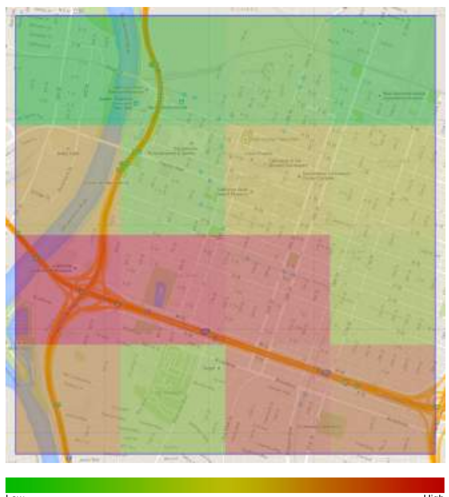
Average Business Profits:



APPENDIX C (Continued)

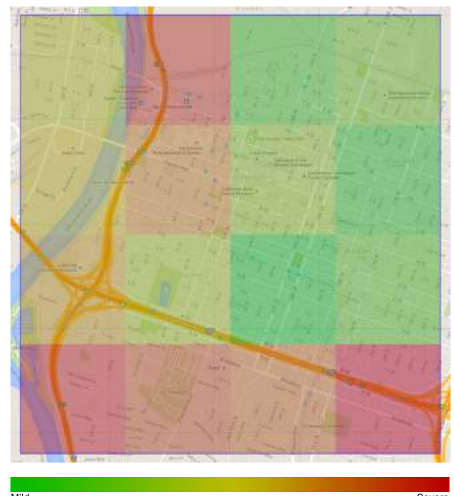
Sacramento – Team A

Crime Rates:



Low High

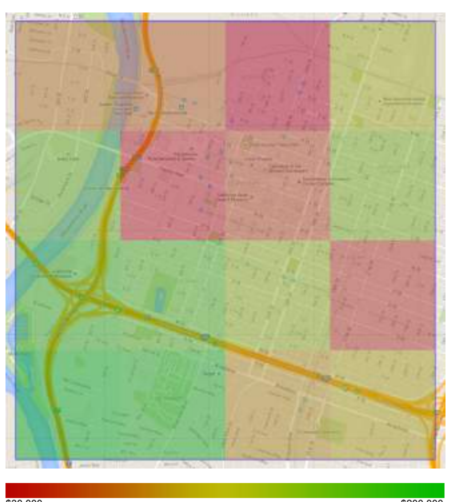
Storm Damage:



Mild Severe

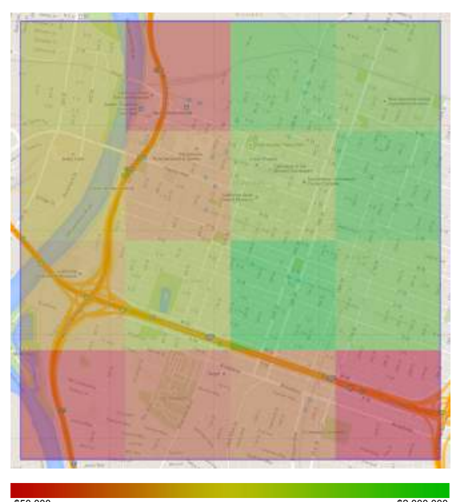
Sacramento – Team B

Average Family Income:



$30,000 $200,000

Average Business Profits:



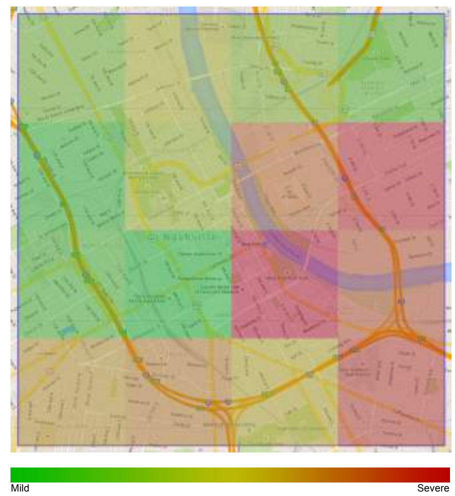
-$50,000 $2,000,000

APPENDIX C (Continued)

Nashville – Team A

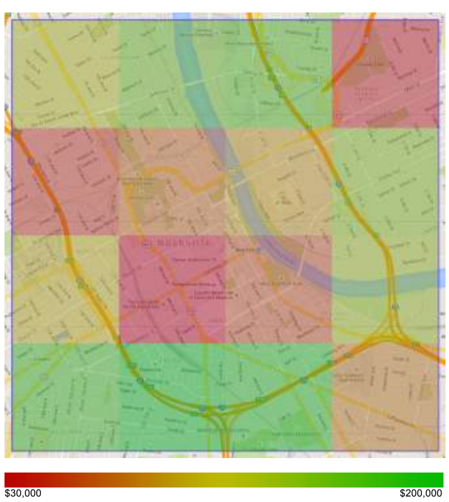
Crime Rates:

Storm Damage:



Nashville – Team B

Average Family Income:

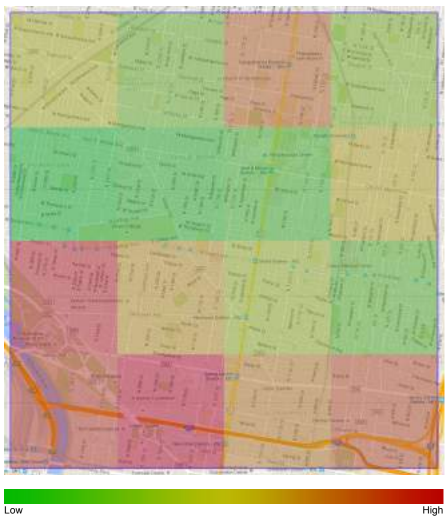


Average Business Profits:

APPENDIX C (Continued)

Philadelphia – Team A

Crime Rates:



Storm Damage:

Philadelphia – Team B

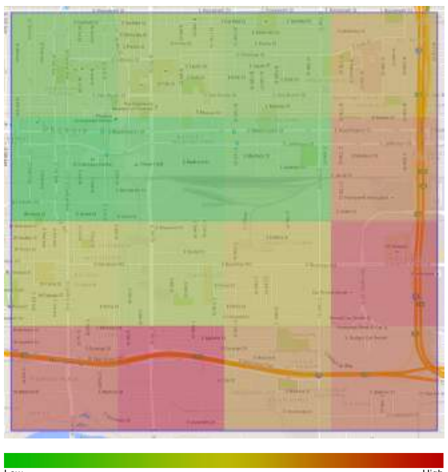
Average Family Income:

Average Business Profits:

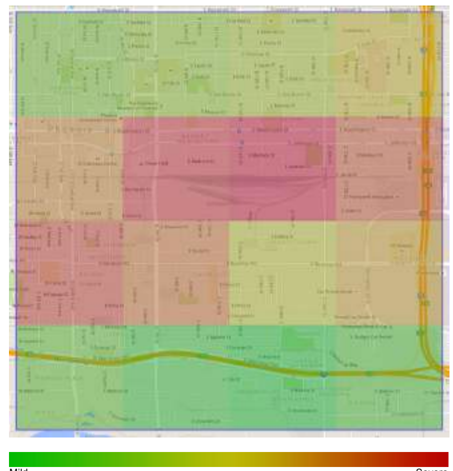
APPENDIX C (Continued)

Phoenix – Team A

Crime Rates:

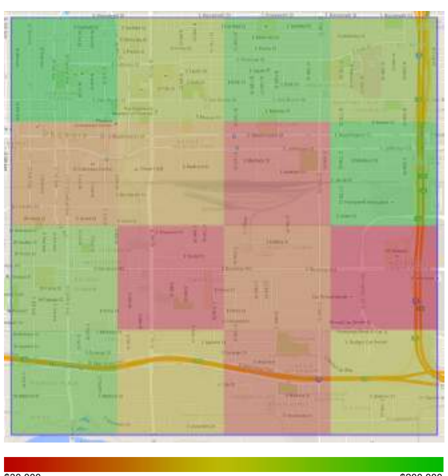


Storm Damage:

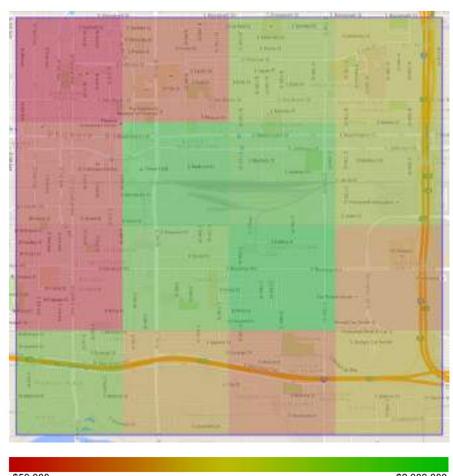


Phoenix – Team B

Average Family Income:



Average Business Profits:



APPENDIX D

Survey used for the longitudinal user study.

University of Illinois at Chicago
Department of Computer Science

Survey questions
Data-Intensive Remote Collaboration using
Scalable Visualizations in Heterogeneous Display Spaces

1. Please rate the **level of participation** on a scale of 1 to 10, with 1 being low and 10 being high (5 being average).

	1 low	2	3	4	5	6	7	8	9	10 high
You as an individual (compared to rest of group)										
Local Site (compared to remote site)										
Remote Site (compared to local site)										

2. Please rate the **ease of use** of collaborative interaction amongst your **local group** on a scale of 1 to 10, with 1 being difficult and 10 being easy.

1 difficult	2	3	4	5	6	7	8	9	10 easy

3. Please rate the **ease of use** of collaborative interaction with the **remote group** on a scale of 1 to 10, with 1 being difficult and 10 being easy.

1 difficult	2	3	4	5	6	7	8	9	10 easy

4. Please rate how successful you felt the tool was at **facilitating local collaboration** on a scale from 1 to 10, with 1 being not successful and 10 being very successful.

1 not	2	3	4	5	6	7	8	9	10 very

5. Please rate how successful you felt the tool was at **facilitating remote collaboration** on a scale from 1 to 10, with 1 being not successful and 10 being very successful.

1 not	2	3	4	5	6	7	8	9	10 very

6. Please rate how much you **liked** using the **remote collaboration features** of the tool on a scale from 1 to 10, with 1 being disliked and 10 being liked.

1 dislike	2	3	4	5	6	7	8	9	10 like

APPENDIX D (Continued)

After Final Session ONLY

7. Please **rank** the three **data-synchronization techniques** used throughout this study, with 1 being the best and 3 being the worst.

	1	2	3
Data-Pushing			
Data-Duplication			
Advanced Data-Synchronization Options			

8. Any further comments about the collaborative tools used during this study

APPENDIX E

Survey used for the formal user study.

University of Illinois at Chicago
Department of Computer Science

Survey questions
Data-Intensive Remote Collaboration using
Scalable Visualizations in Heterogeneous Display Spaces

1. Please rate the **ease of use** of collaborative interaction amongst your **local group** on a scale of 1 to 10, with 1 being difficult and 10 being easy.

1 difficult	2	3	4	5	6	7	8	9	10 easy

2. Please rate how successful you felt the tool was at **facilitating local collaboration** on a scale from 1 to 10, with 1 being not successful and 10 being very successful.

1 not	2	3	4	5	6	7	8	9	10 very

3. Please rate the **ease of use** of collaborative interaction with the **remote group** for **data-pushing** on a scale of 1 to 10, with 1 being difficult and 10 being easy.

1 difficult	2	3	4	5	6	7	8	9	10 easy

4. Please rate how successful you felt **data-pushing** was at **facilitating remote collaboration** on a scale from 1 to 10, with 1 being not successful and 10 being very successful.

1 not	2	3	4	5	6	7	8	9	10 very

5. Please rate how much you **liked** using the **remote collaboration features** of **data-pushing** on a scale from 1 to 10, with 1 being disliked and 10 being liked.

1 dislike	2	3	4	5	6	7	8	9	10 like

6. Please rate the **ease of use** of collaborative interaction with the **remote group** for **data-duplication** on a scale of 1 to 10, with 1 being difficult and 10 being easy.

1 difficult	2	3	4	5	6	7	8	9	10 easy

7. Please rate how successful you felt **data-duplication** was at **facilitating remote collaboration** on a scale from 1 to 10, with 1 being not successful and 10 being very successful.

1 not	2	3	4	5	6	7	8	9	10 very

Data-intensive Remote Collaboration using Scalable Visualizations in Heterogeneous Display Spaces.
Survey questions version: 1, 05/21/2015, Page 1 of 2

APPENDIX E (Continued)

8. Please rate how much you **liked** using the **remote collaboration features** of **data-duplication** on a scale from 1 to 10, with 1 being disliked and 10 being liked.

1 dislike	2	3	4	5	6	7	8	9	10 like

9. Please rate the **ease of use** of collaborative interaction with the **remote group** for **advanced data-synchronization options** on a scale of 1 to 10, with 1 being difficult and 10 being easy.

1 difficult	2	3	4	5	6	7	8	9	10 easy

10. Please rate how successful you felt **advanced data-synchronization options** was at **facilitating remote collaboration** on a scale from 1 to 10, with 1 being not successful and 10 being very successful.

1 not	2	3	4	5	6	7	8	9	10 very

11. Please rate how much you **liked** using the **remote collaboration features** of **advanced data-synchronization options** on a scale from 1 to 10, with 1 being disliked and 10 being liked.

1 dislike	2	3	4	5	6	7	8	9	10 like

12. Please **rank** the three **data-synchronization techniques** used throughout this study, with 1 being the best and 3 being the worst.

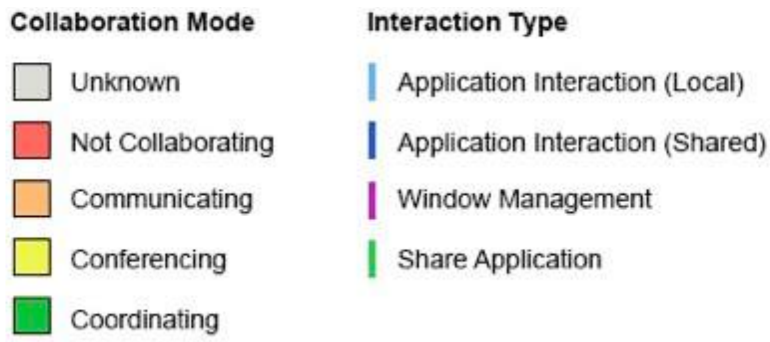
	Rank
Data-Pushing	
Data-Duplication	
Advanced Data-Synchronization Options	

13. Any further comments about the collaborative tools used during this study.

APPENDIX F

Timeline visualizations of the collaboration modes and user interaction for each trial in the formal user study.

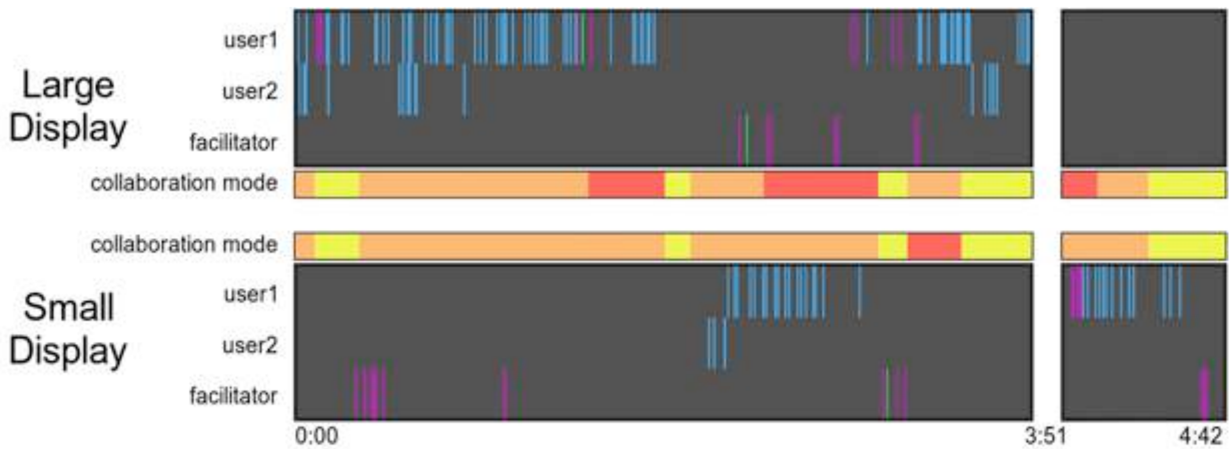
The legend for all timeline visualizations is below:

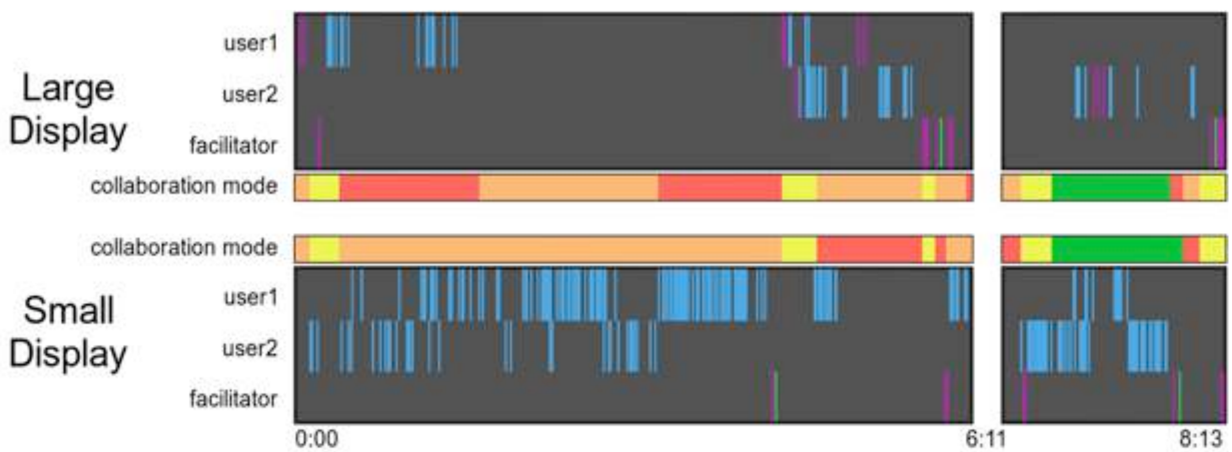


APPENDIX F (Continued)

The first set shows normalized timeline visualizations that are the same width regardless of task completion time. The timelines are ordered from quickest to slowest for each data sharing technique.

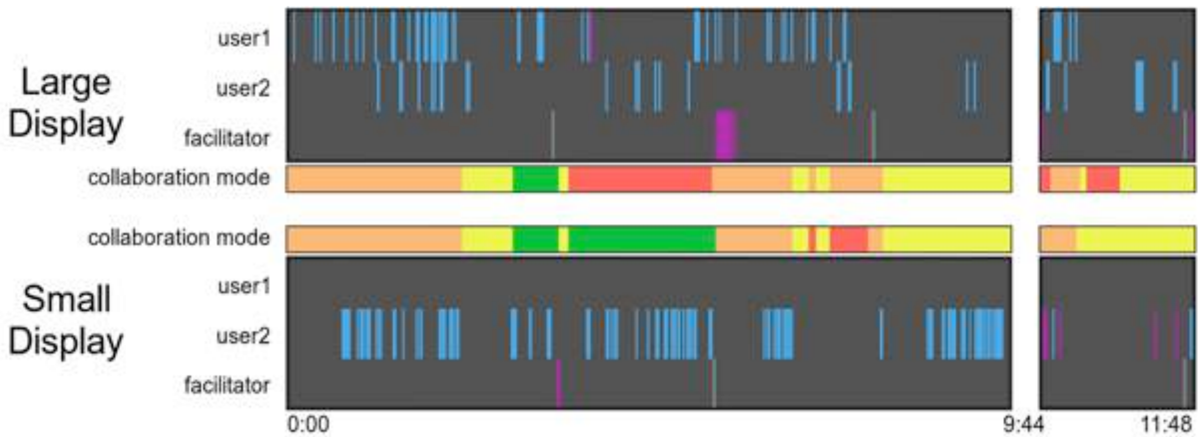
Normalized *data-pushing* timelines:

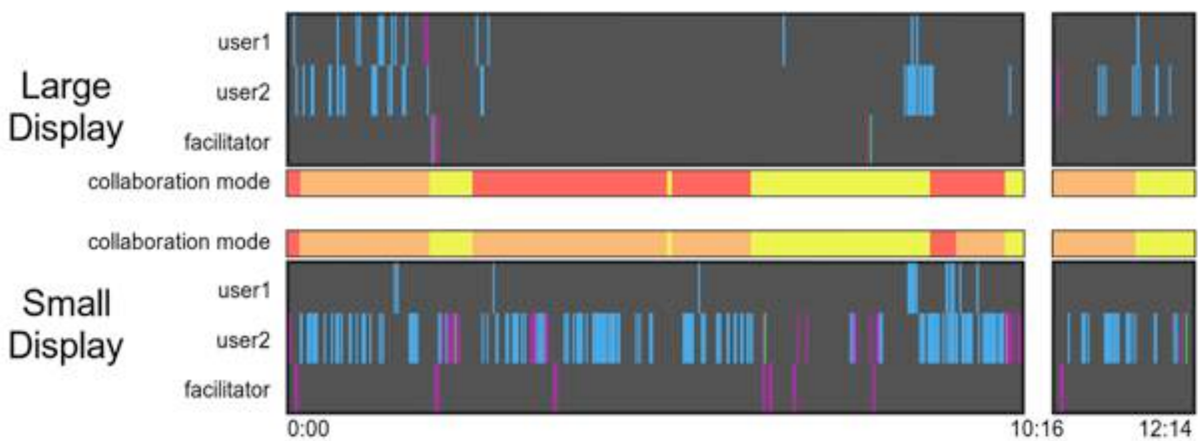


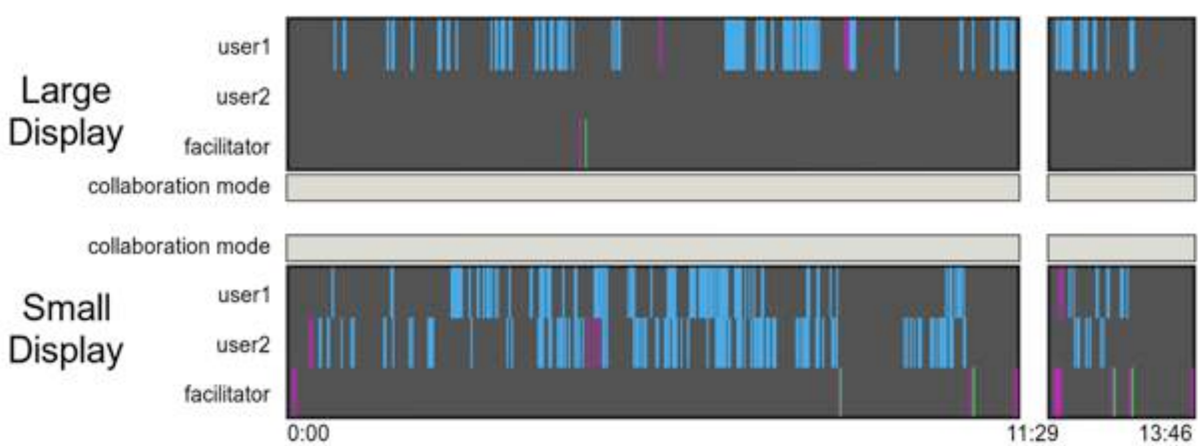


APPENDIX F (Continued)

Normalized *data-pushing* timelines (continued):

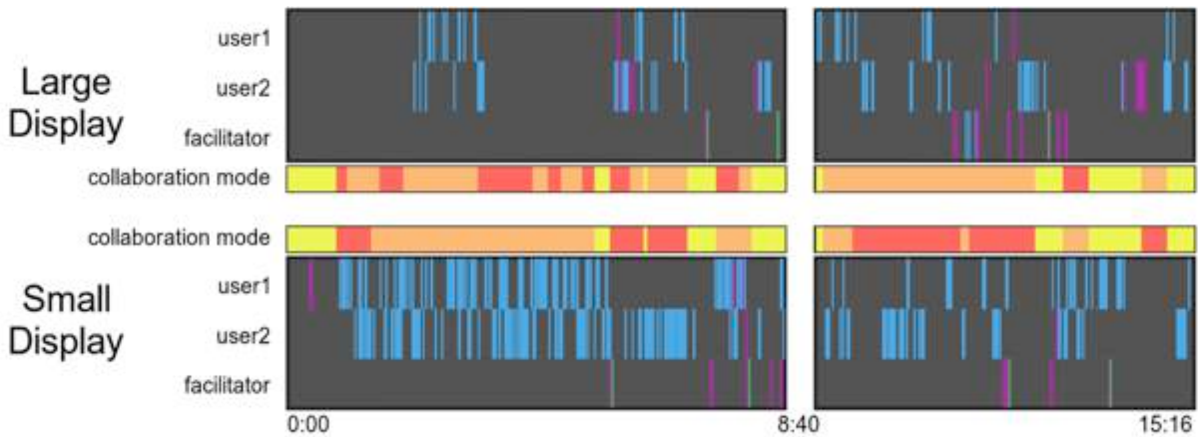


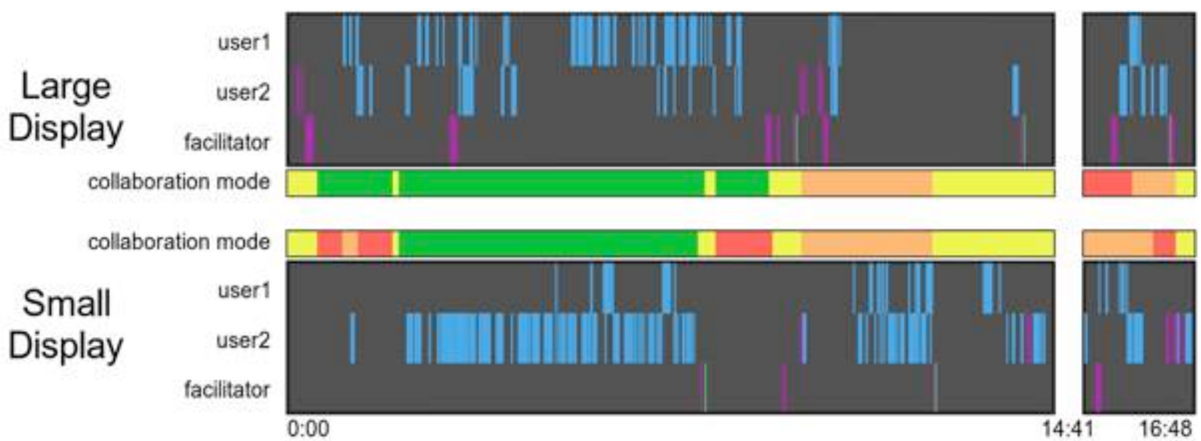


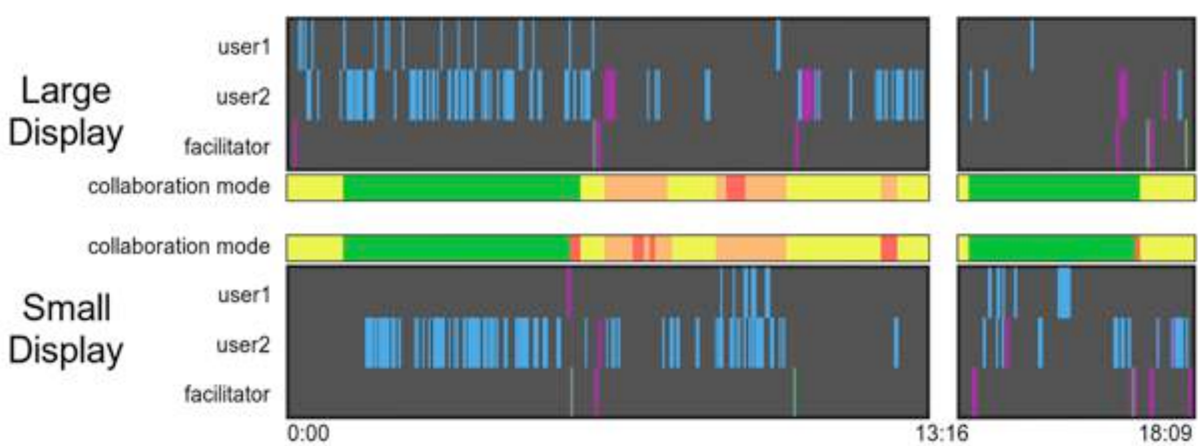


APPENDIX F (Continued)

Normalized *data-pushing* timelines (continued):

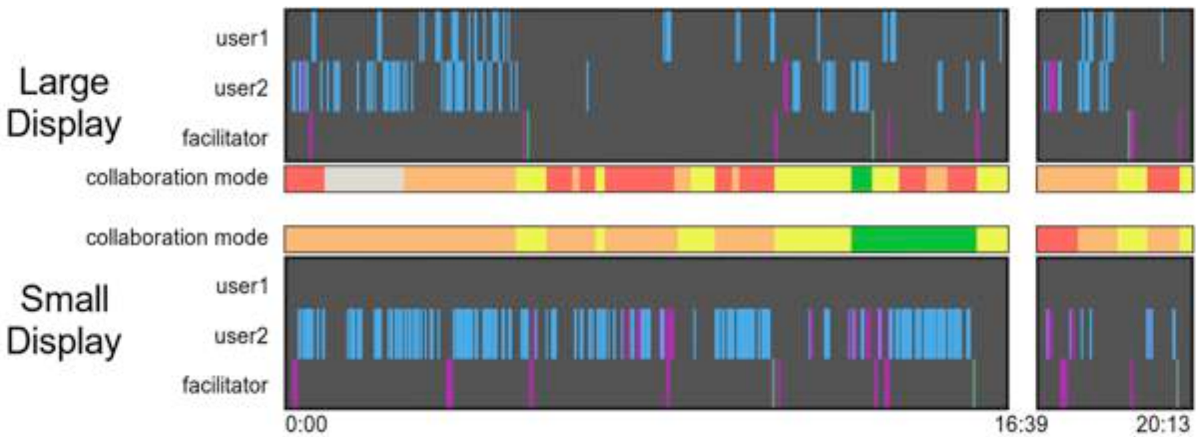


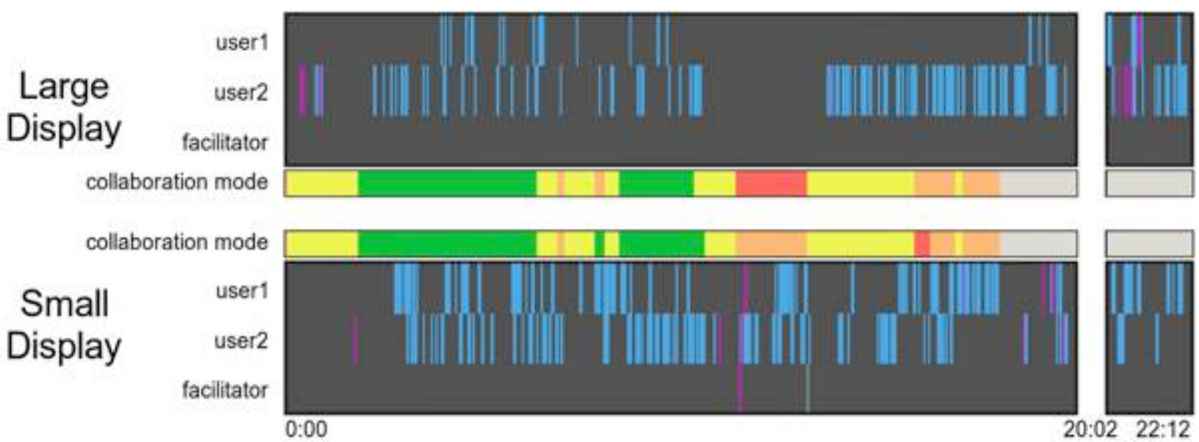


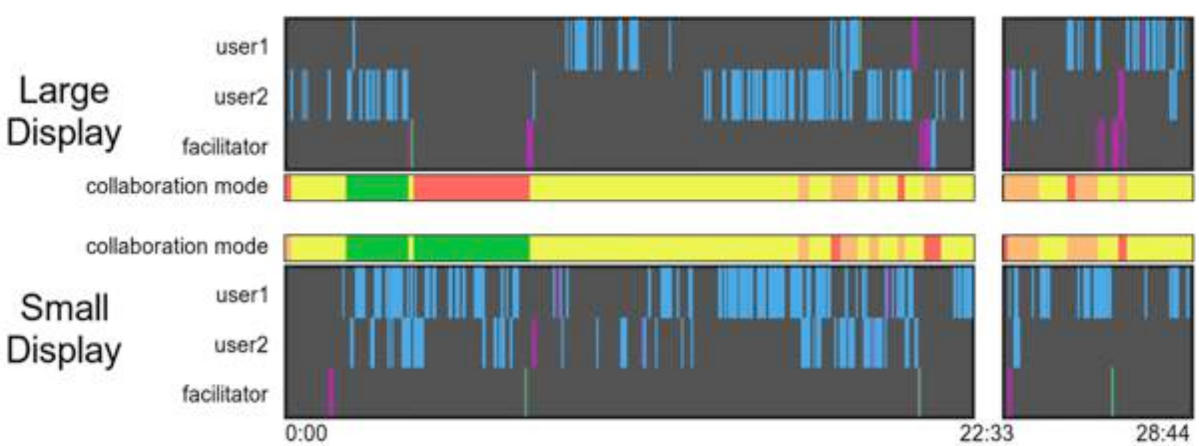


APPENDIX F (Continued)

Normalized *data-pushing* timelines (continued):

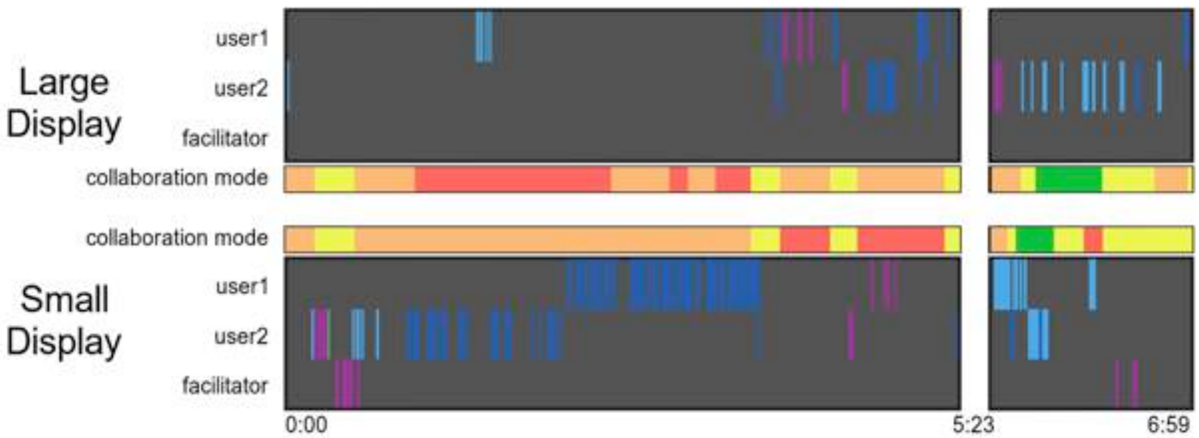


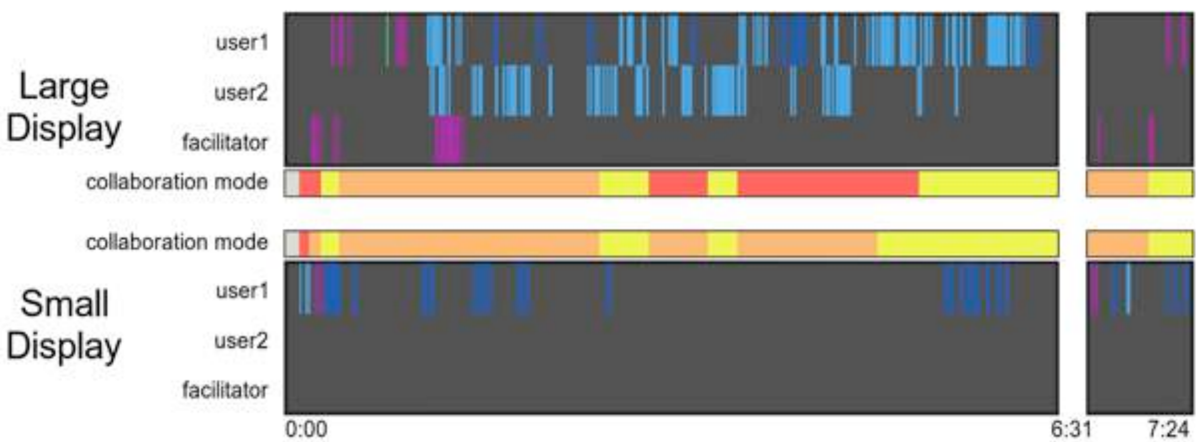


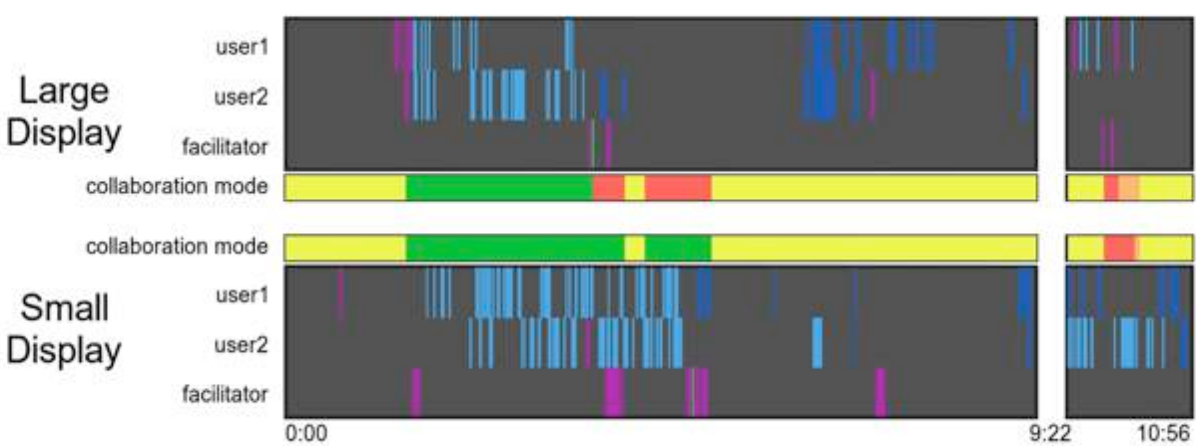


APPENDIX F (Continued)

Normalized *data-duplication* timelines:

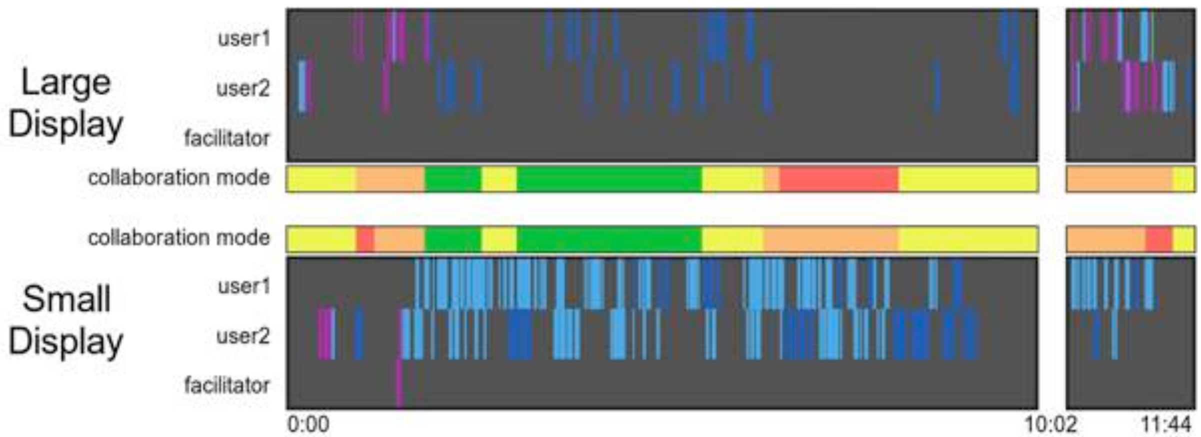


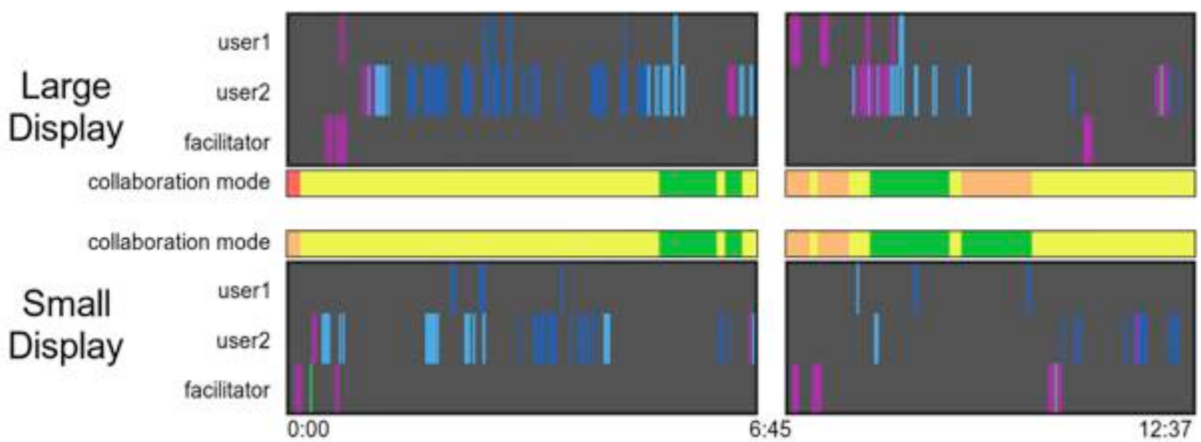


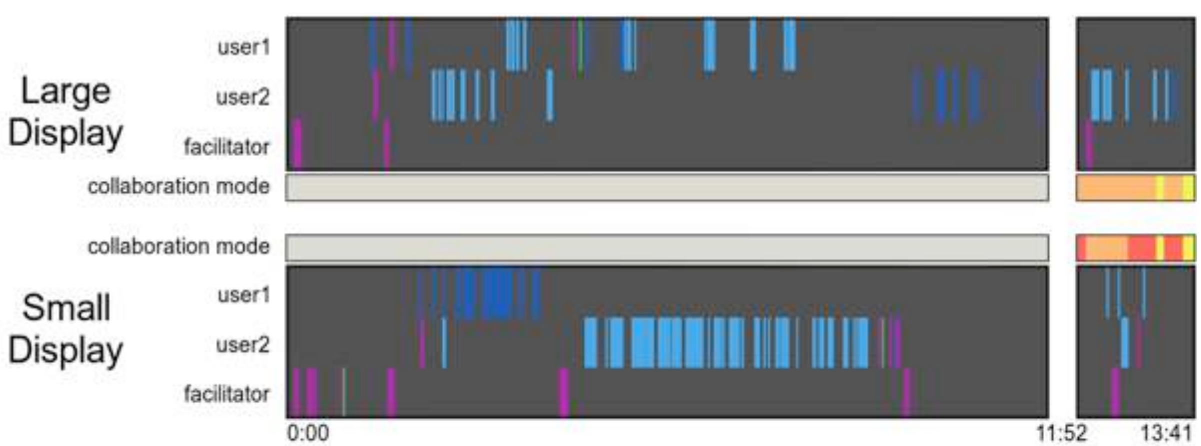


APPENDIX F (Continued)

Normalized *data-duplication* timelines (continued):

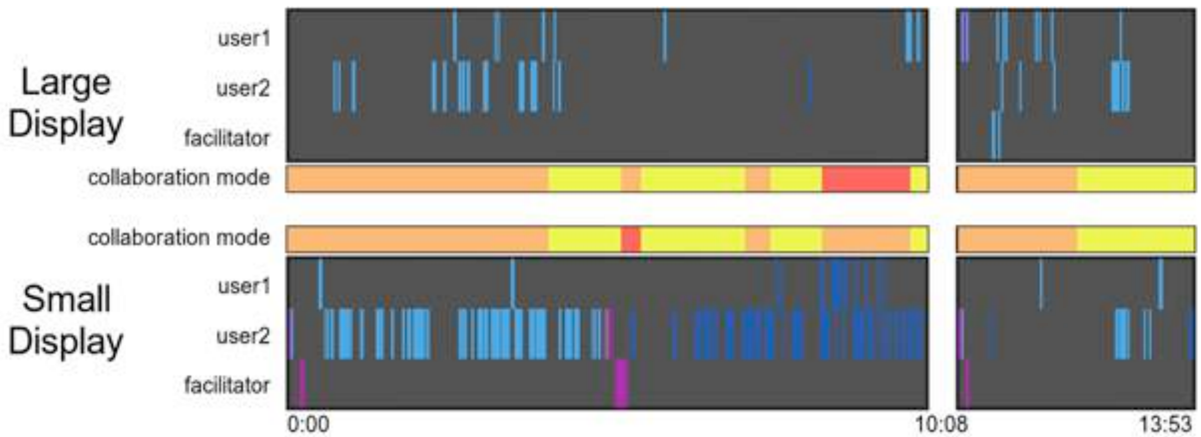


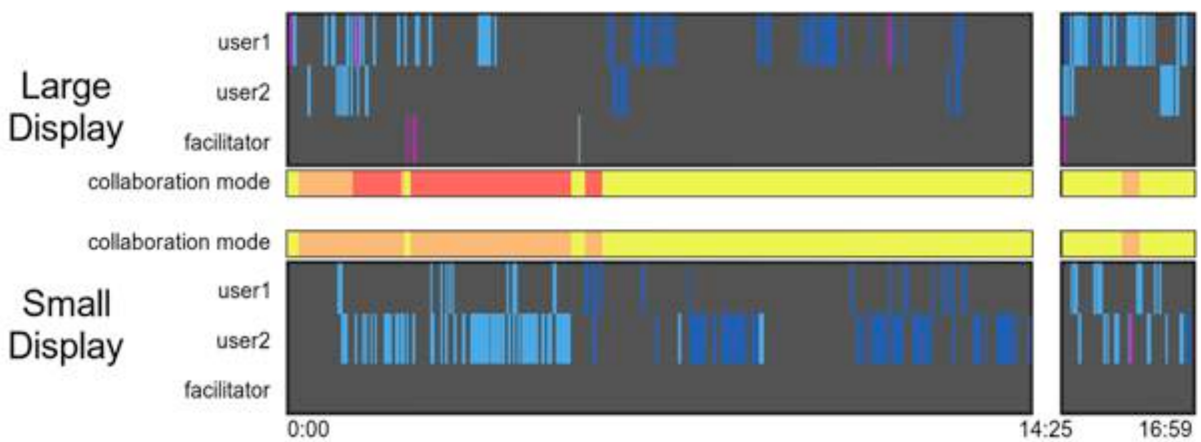


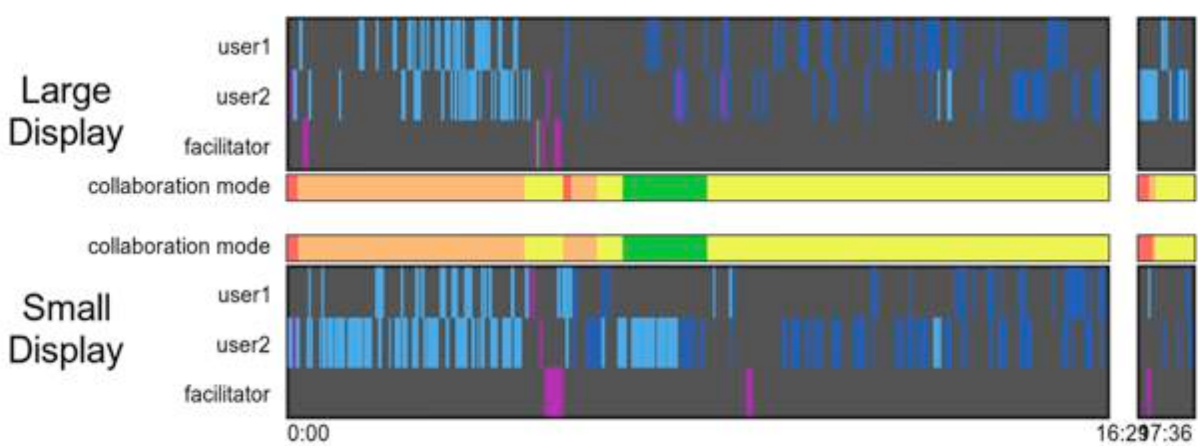


APPENDIX F (Continued)

Normalized *data-duplication* timelines (continued):

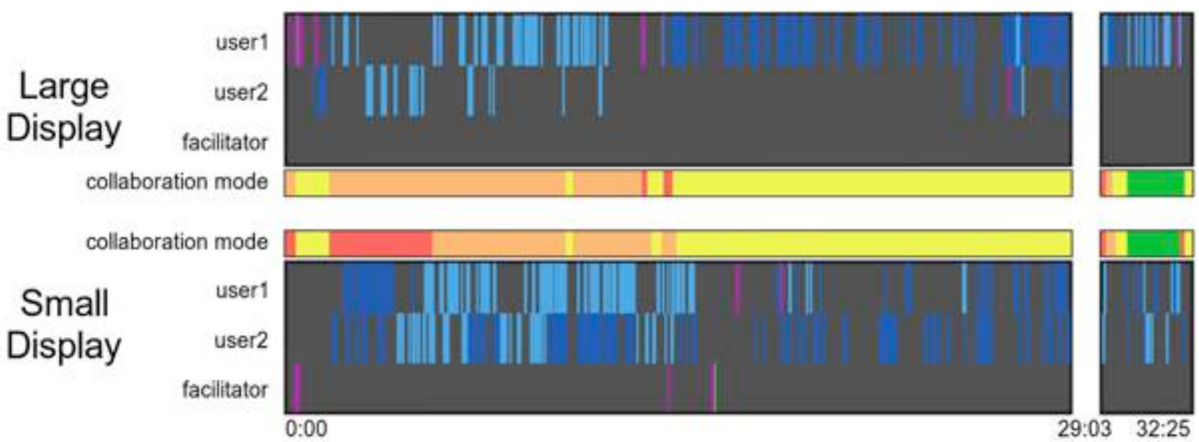






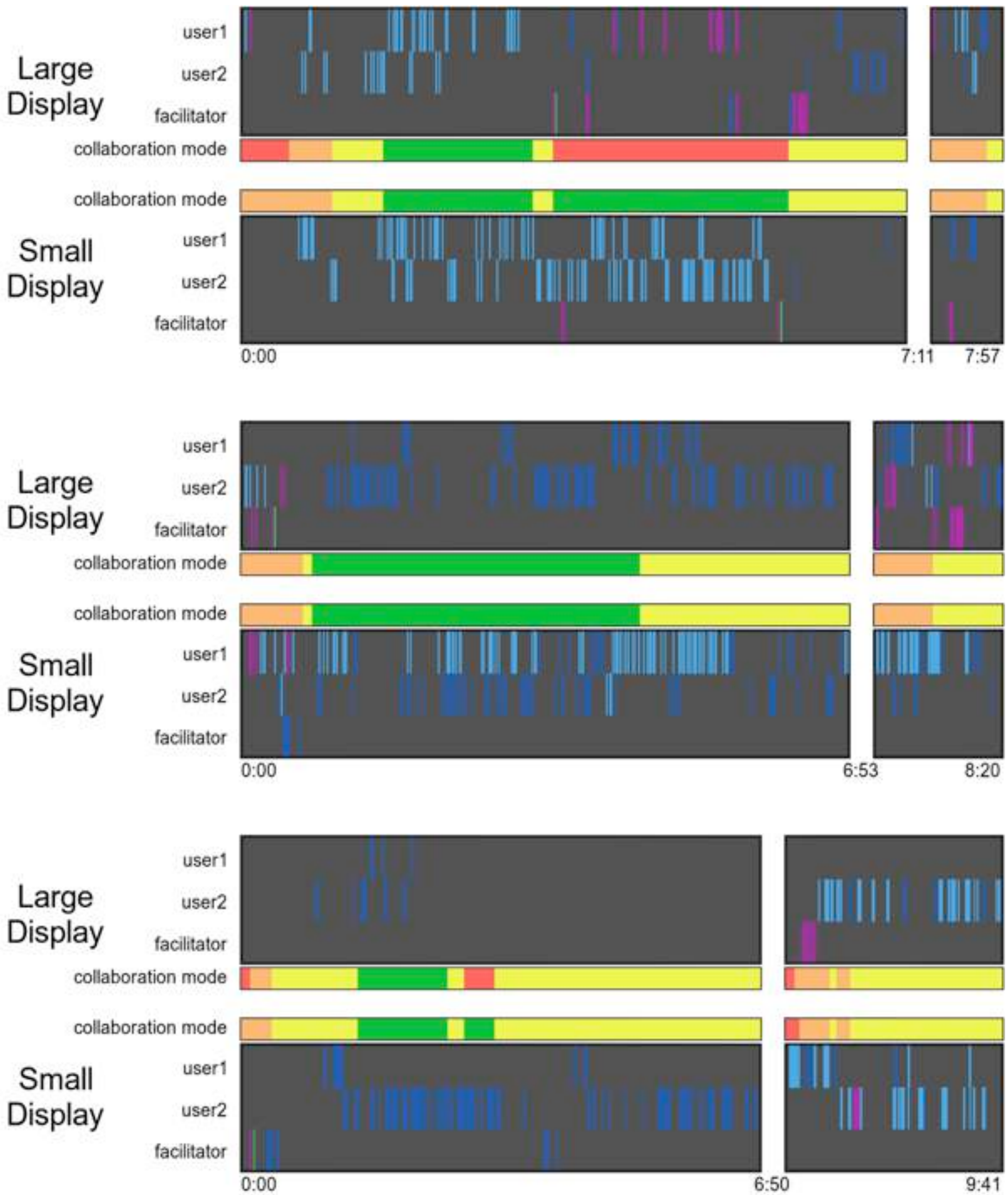
APPENDIX F (Continued)

Normalized *data-duplication* timelines (continued):



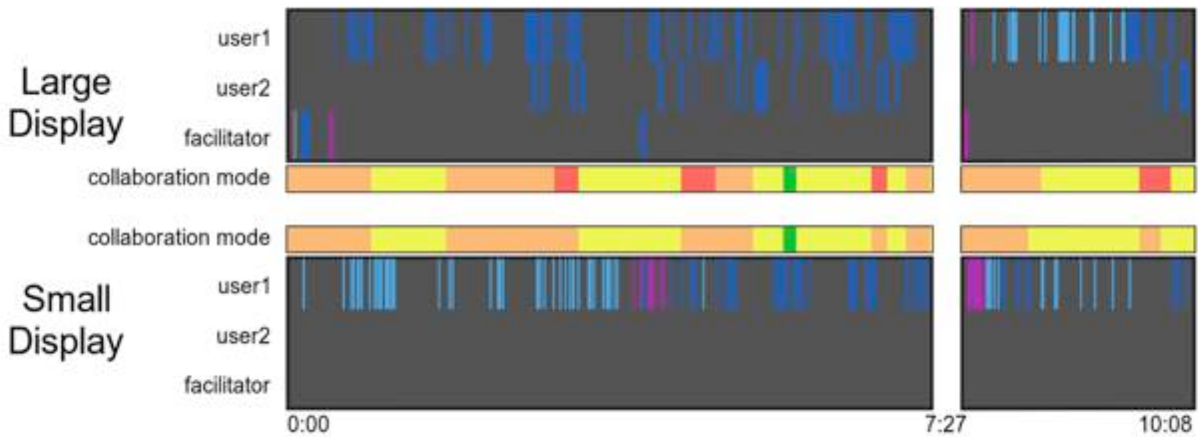
APPENDIX F (Continued)

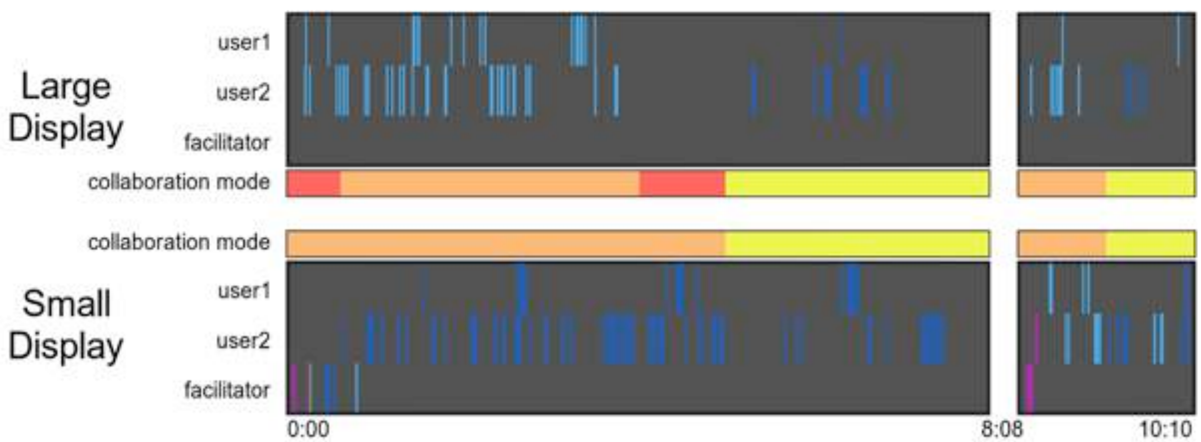
Normalized *advanced data-synchronization options* timelines:

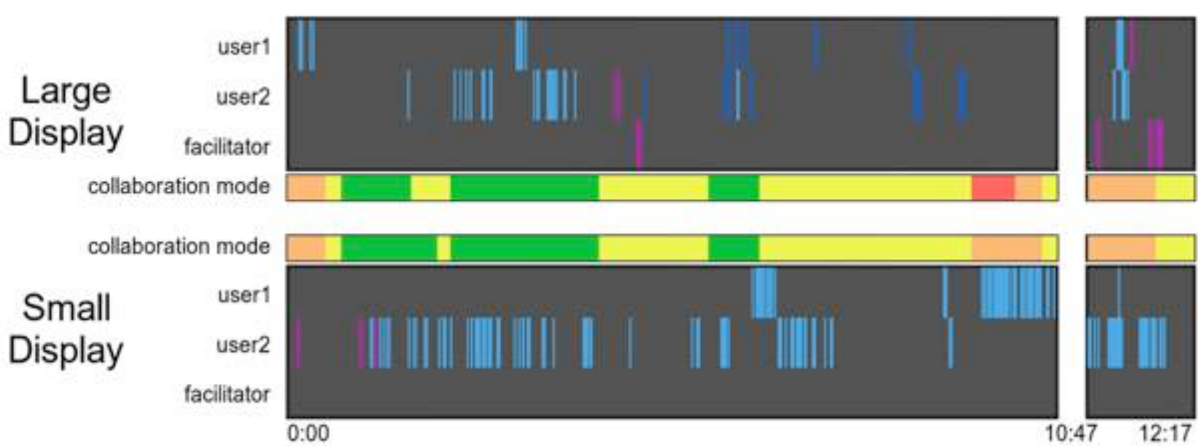


APPENDIX F (Continued)

Normalized *advanced data-synchronization options* timelines (continued):

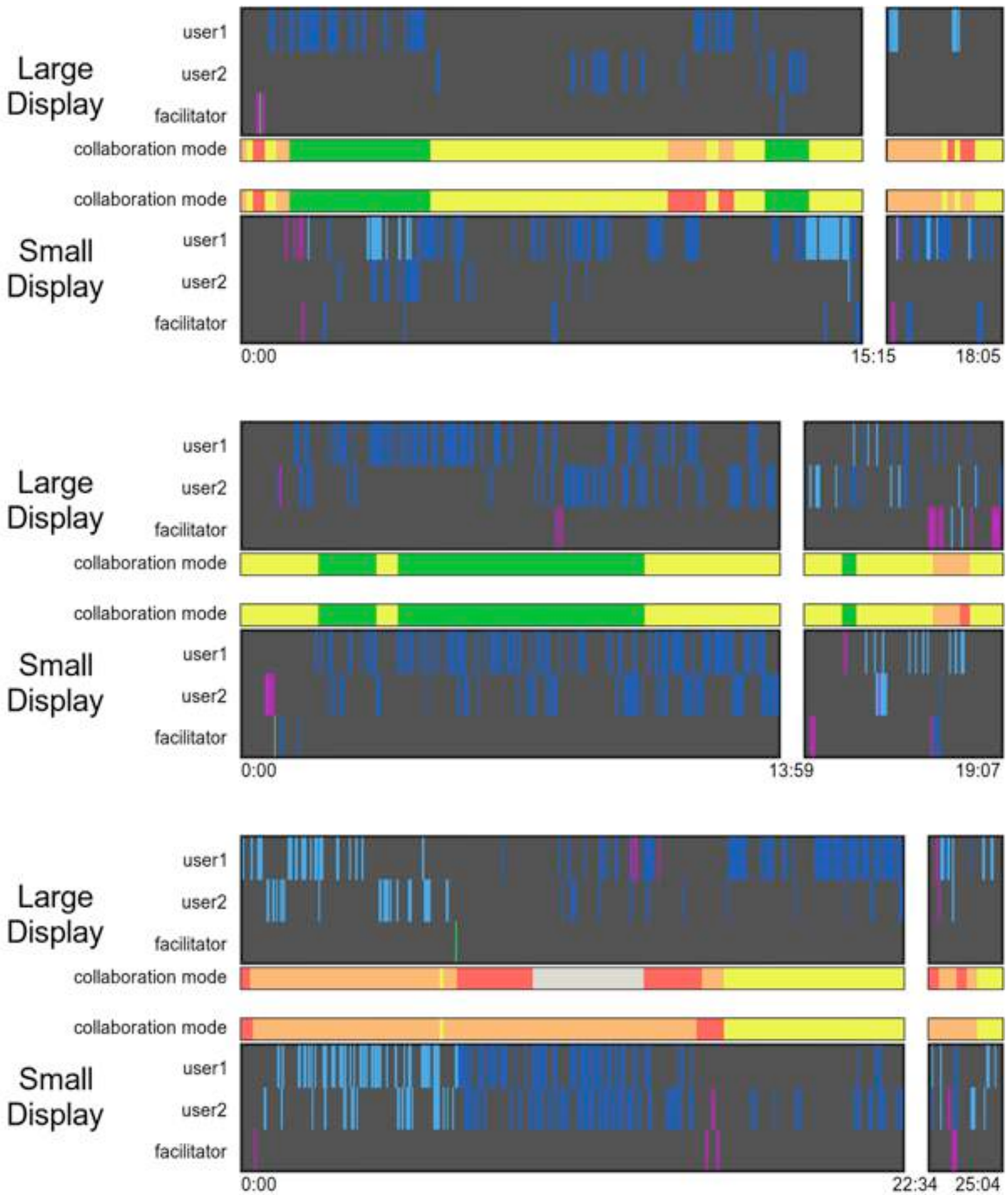






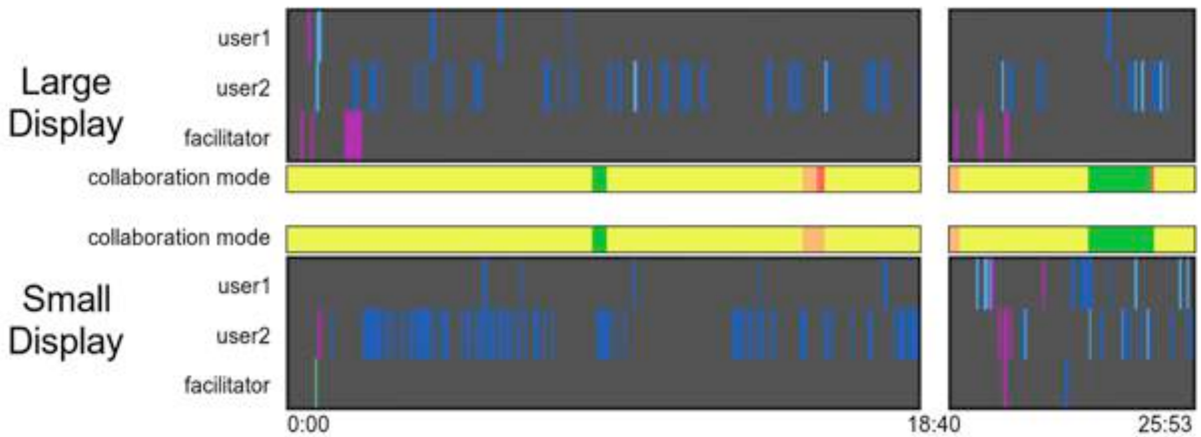
APPENDIX F (Continued)

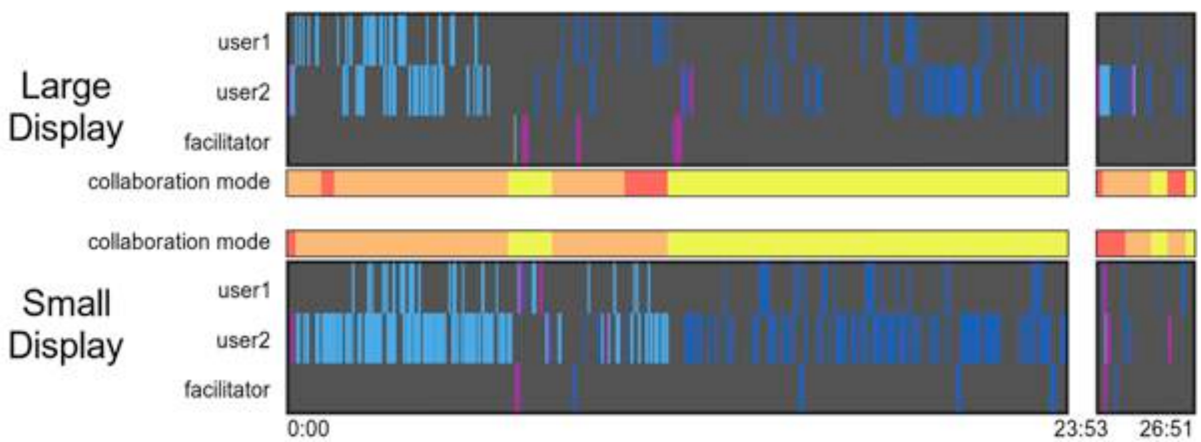
Normalized *advanced data-synchronization options* timelines (continued):



APPENDIX F (Continued)

Normalized *advanced data-synchronization options* timelines (continued):

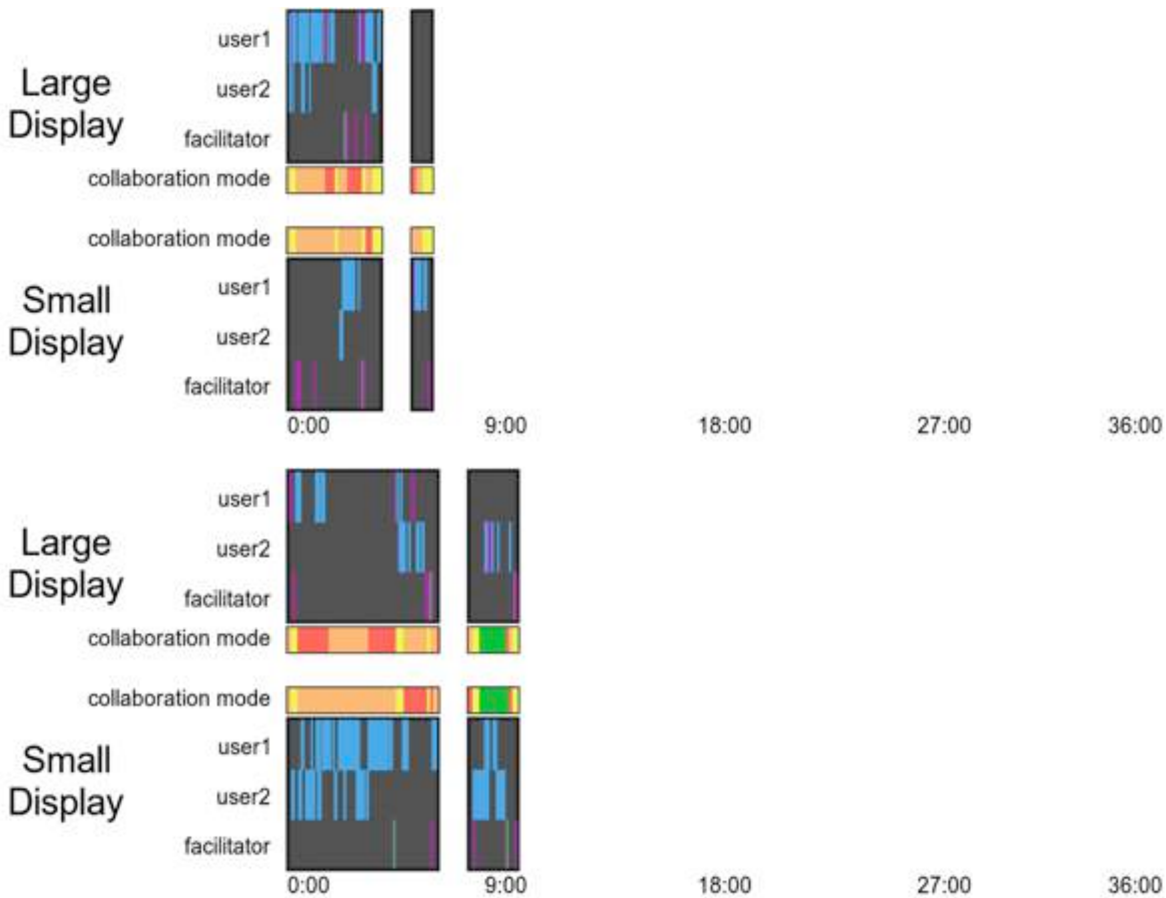




APPENDIX F (Continued)

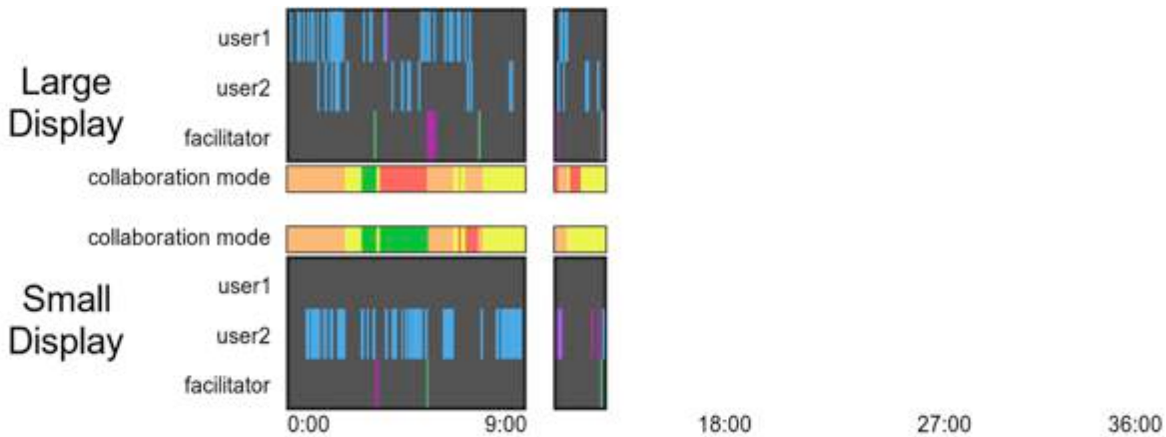
The second set shows absolute timeline visualizations that can be easily compared between visualizations. The timelines are ordered from quickest to slowest for each data sharing technique.

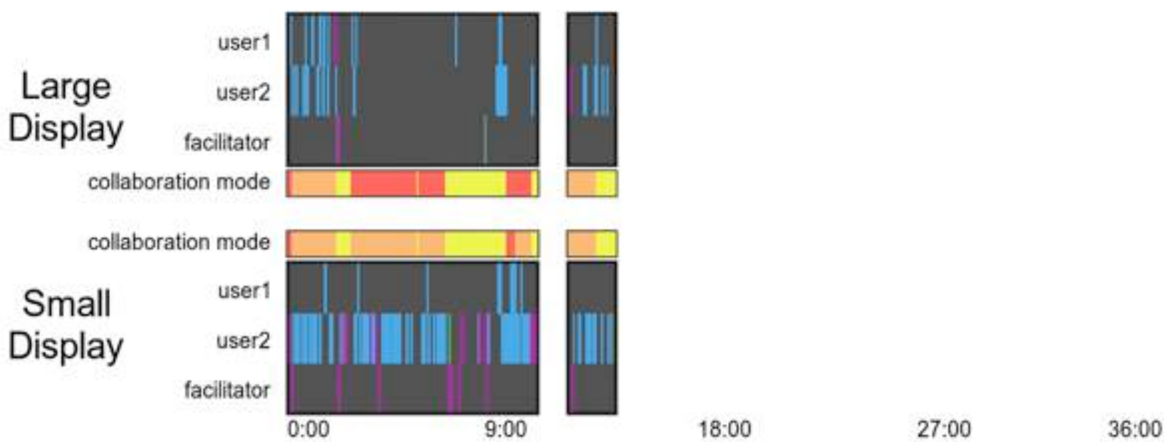
Absolute *data-pushing* timelines:

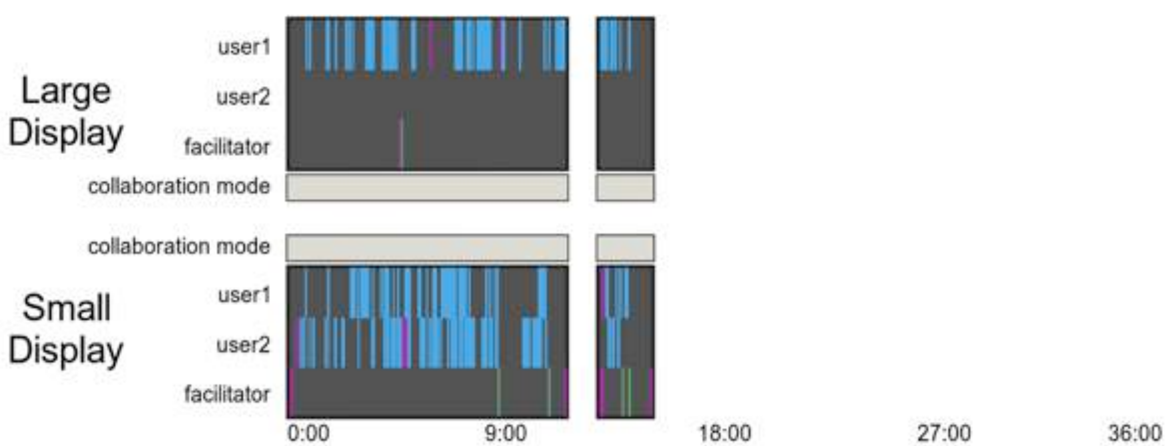


APPENDIX F (Continued)

Absolute *data-pushing* timelines (continued):

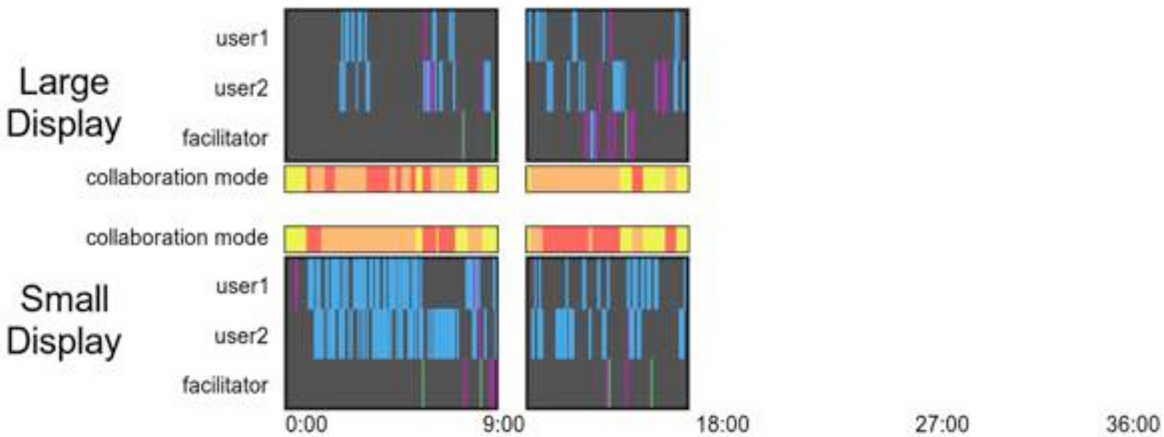


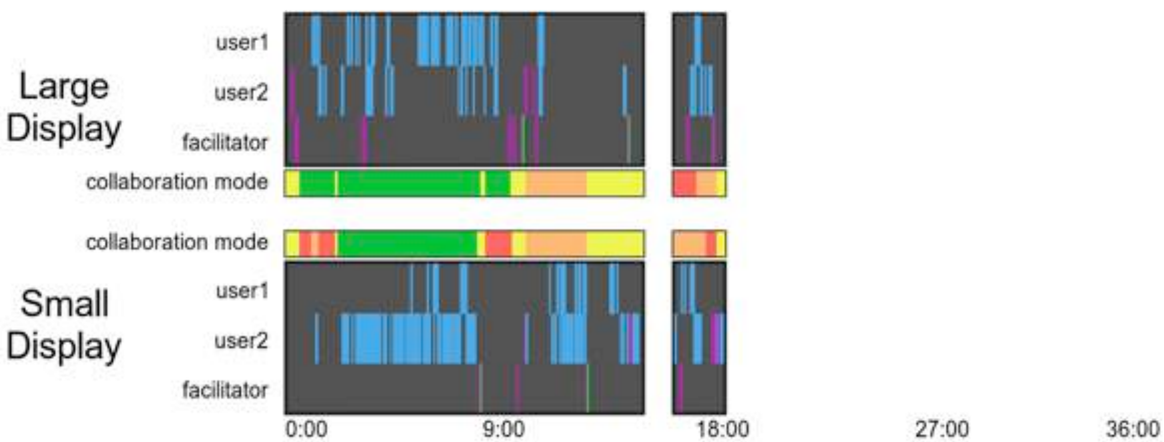


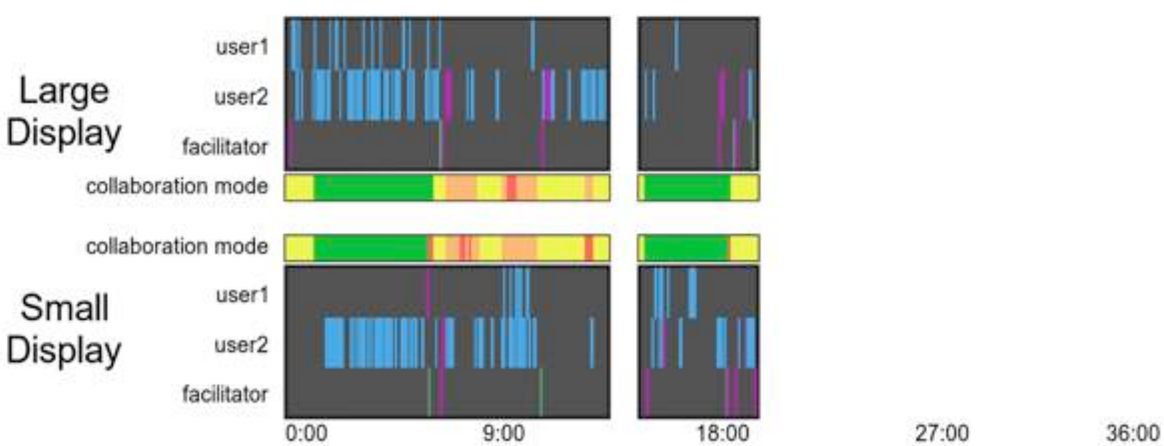


APPENDIX F (Continued)

Absolute *data-pushing* timelines (continued):

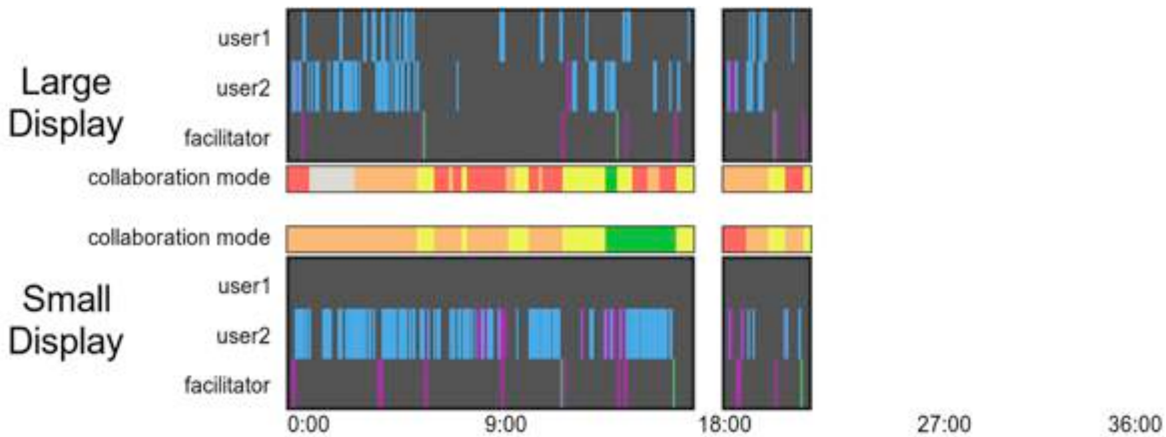


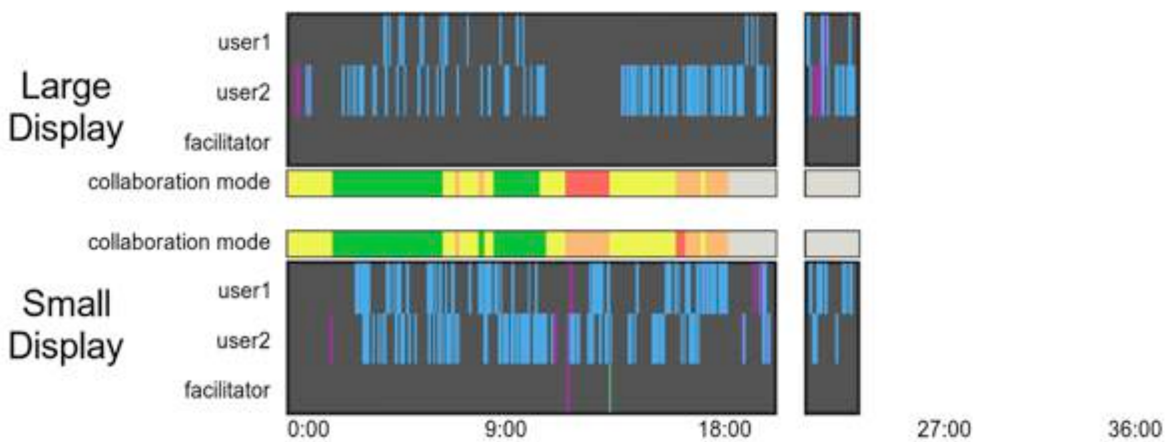


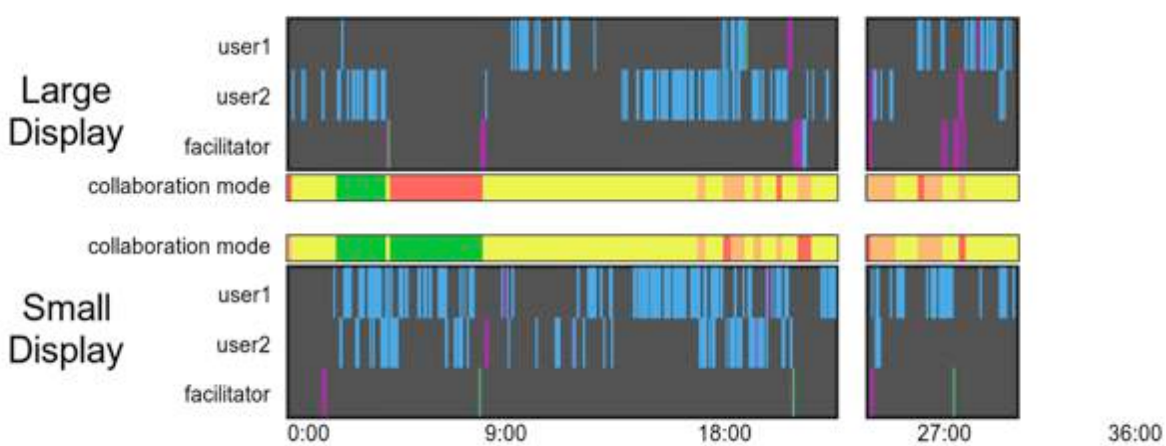


APPENDIX F (Continued)

Absolute *data-pushing* timelines (continued):

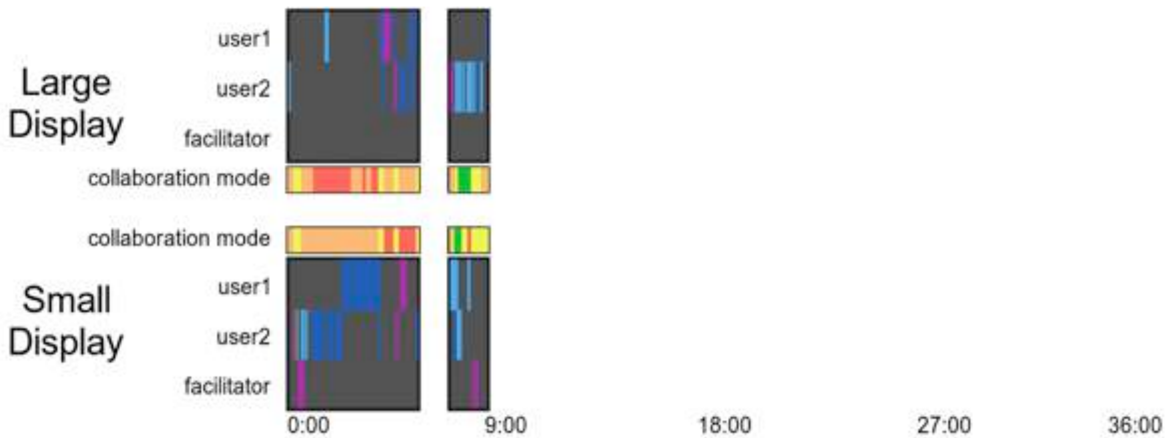


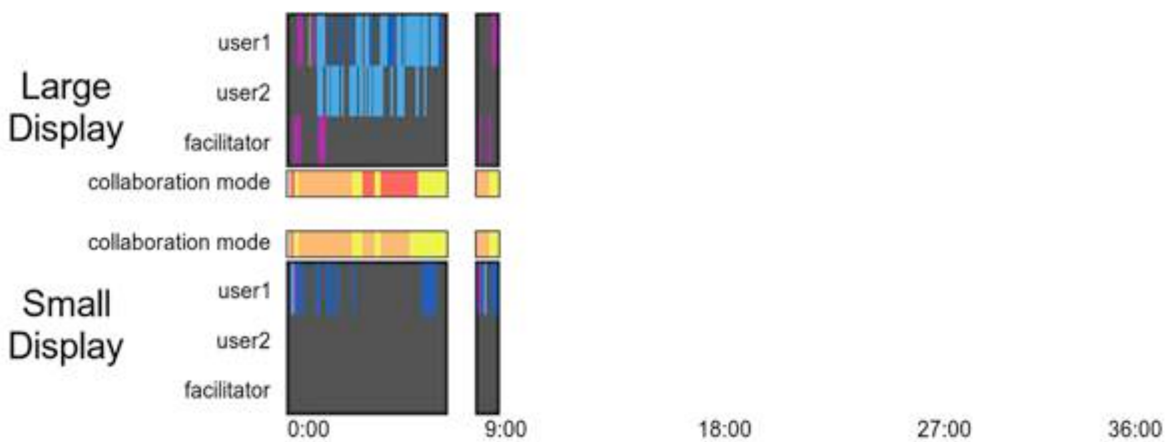




APPENDIX F (Continued)

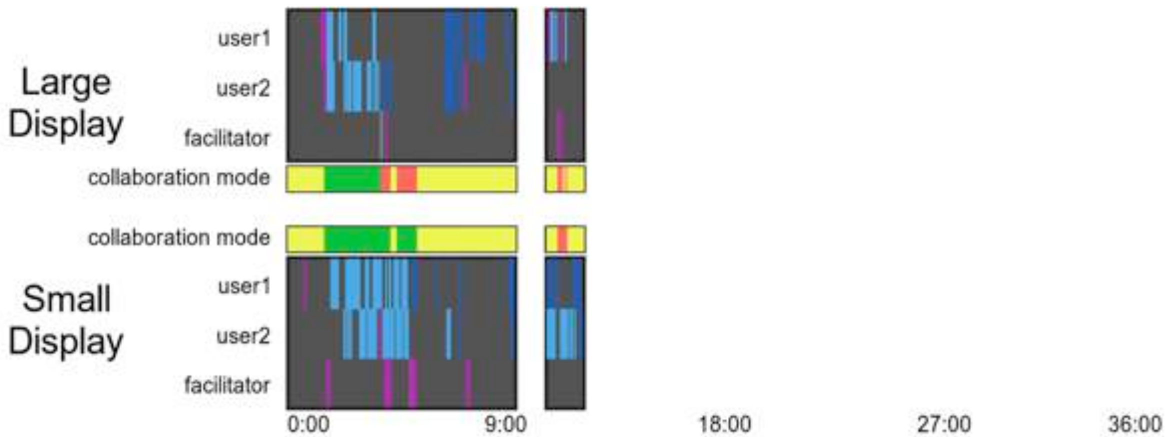
Absolute *data-duplication* timelines:

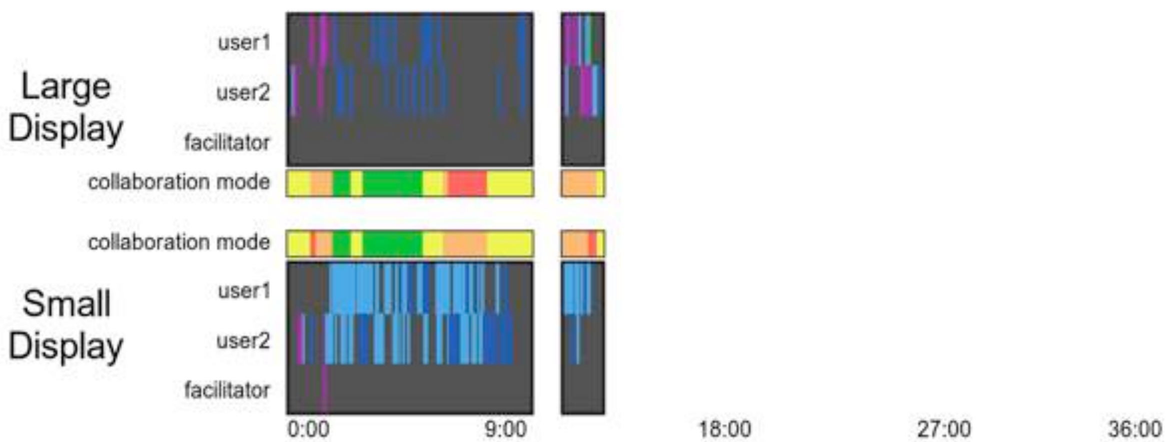


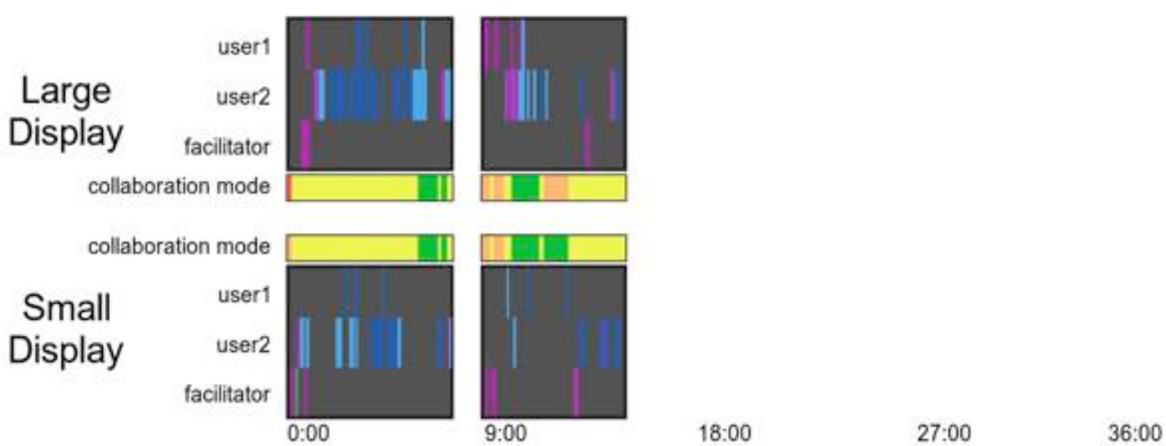


APPENDIX F (Continued)

Absolute *data-duplication* timelines (continued):

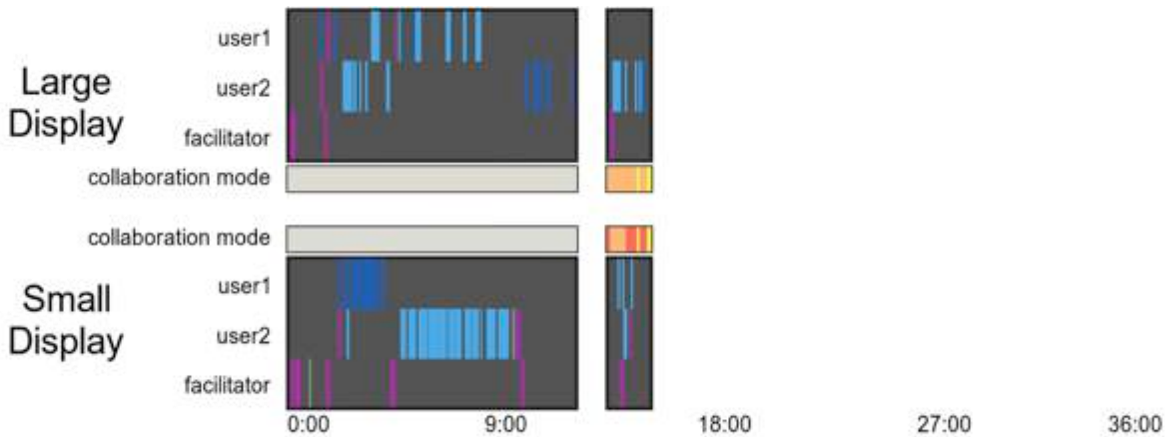


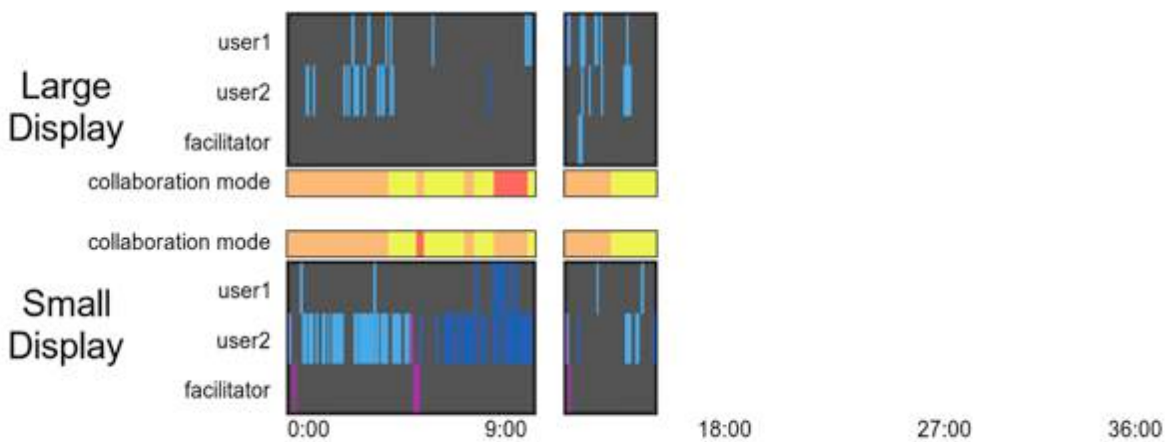


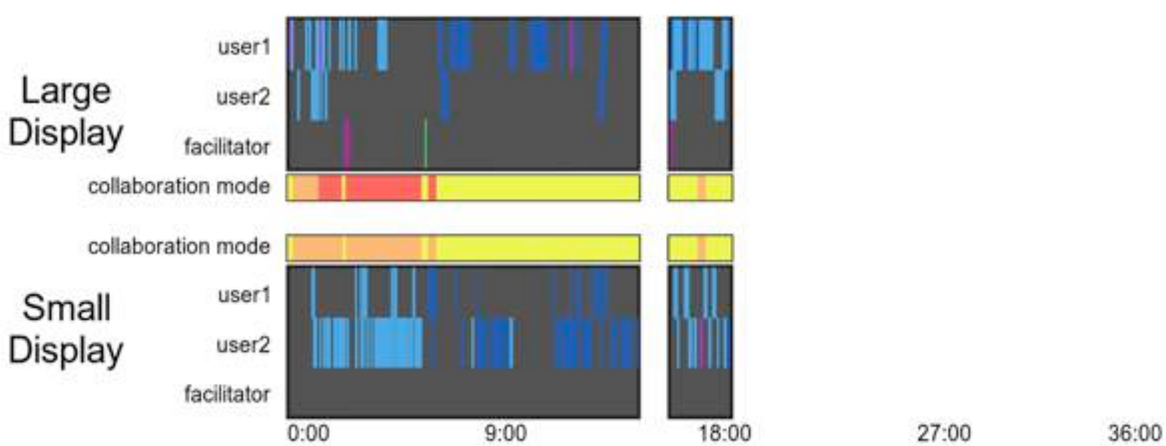


APPENDIX F (Continued)

Absolute *data-duplication* timelines (continued):

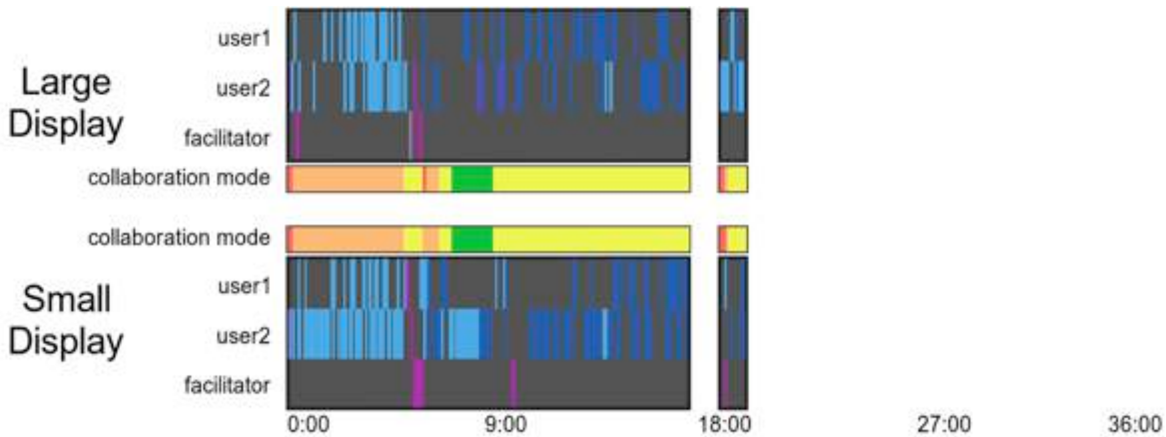


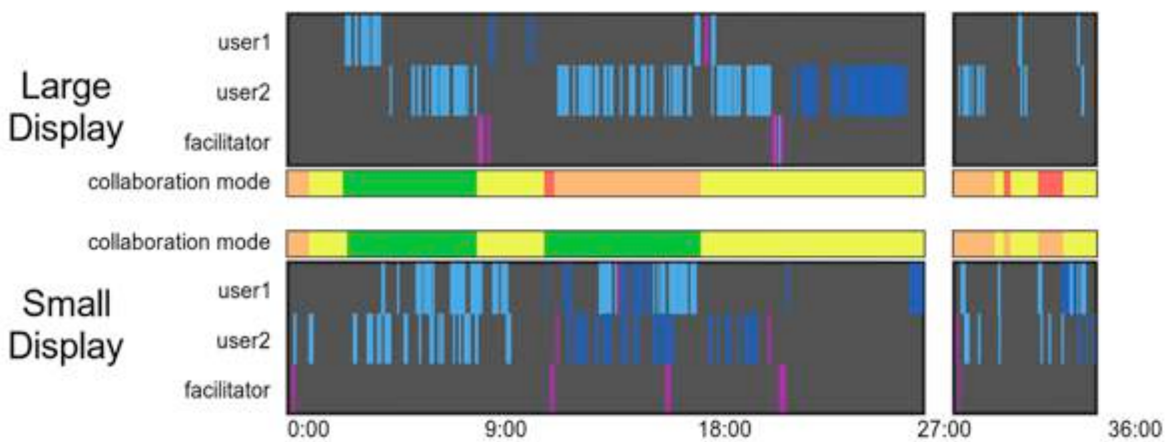


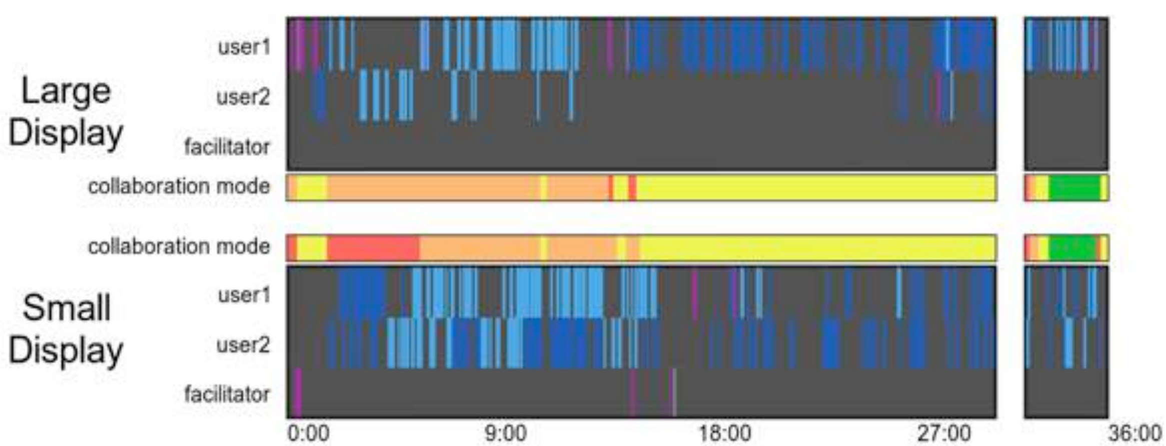


APPENDIX F (Continued)

Absolute *data-duplication* timelines (continued):

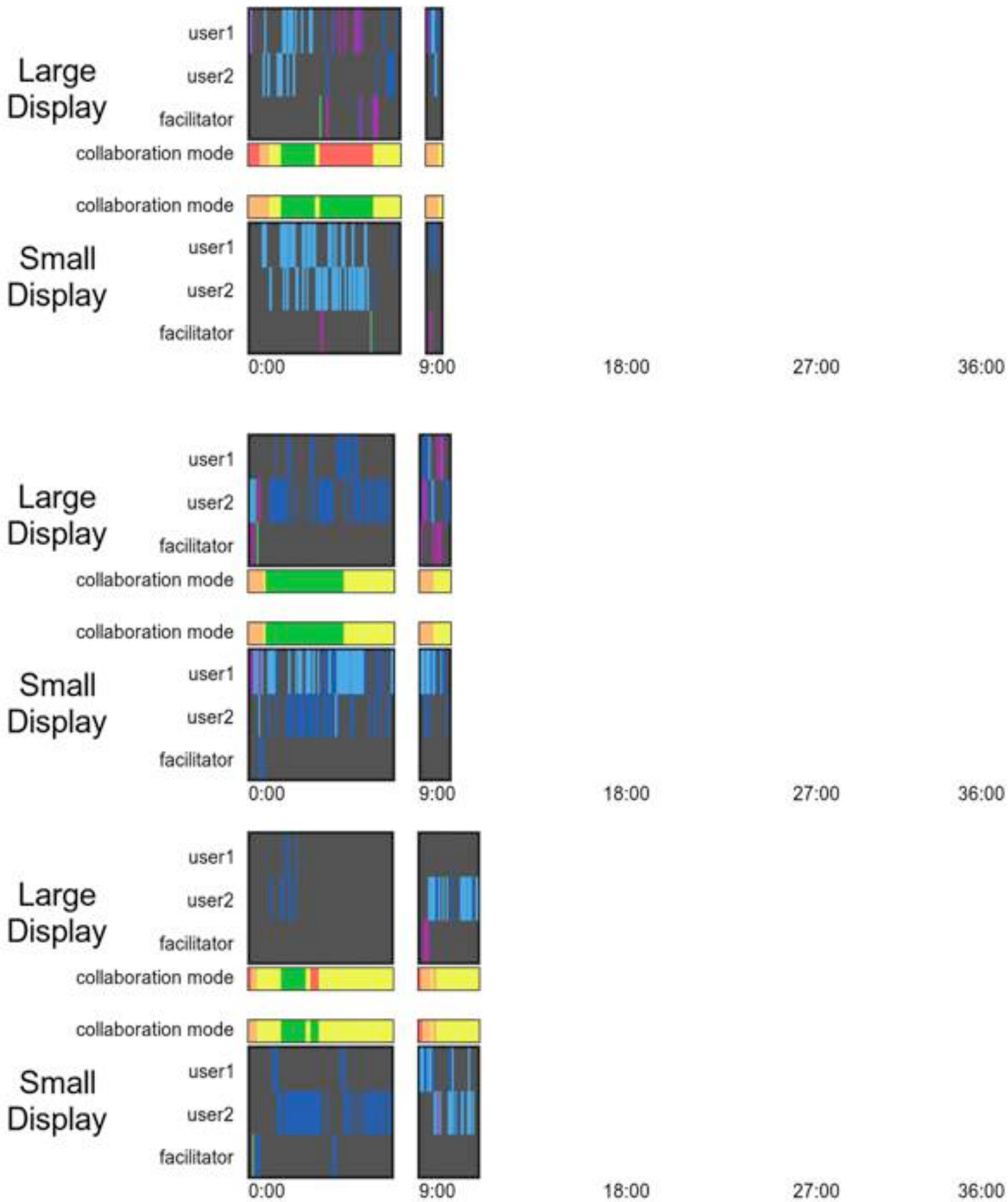






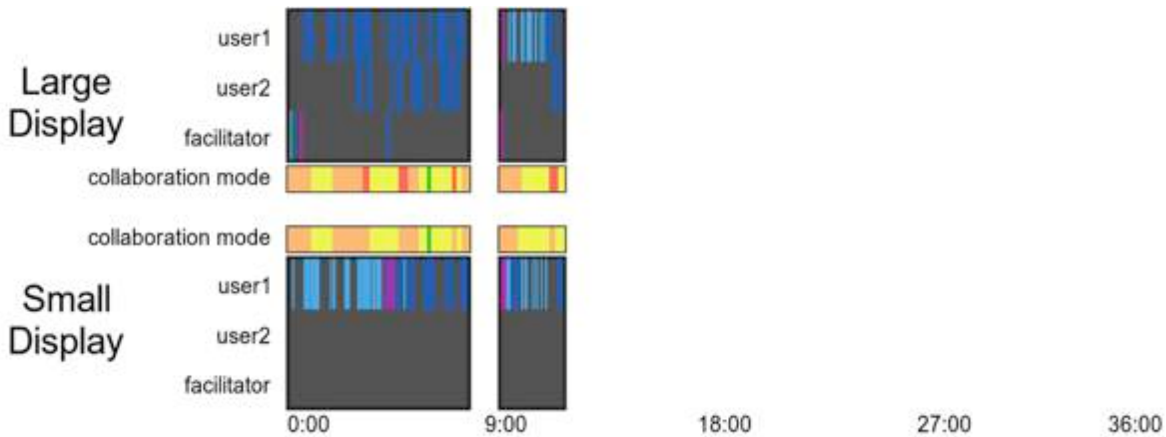
APPENDIX F (Continued)

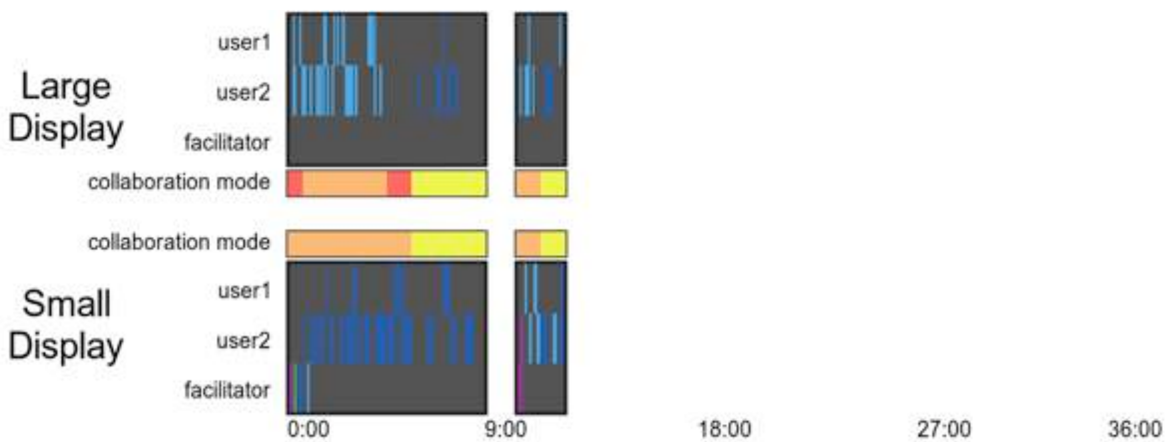
Absolute *advanced data-synchronization options* timelines:

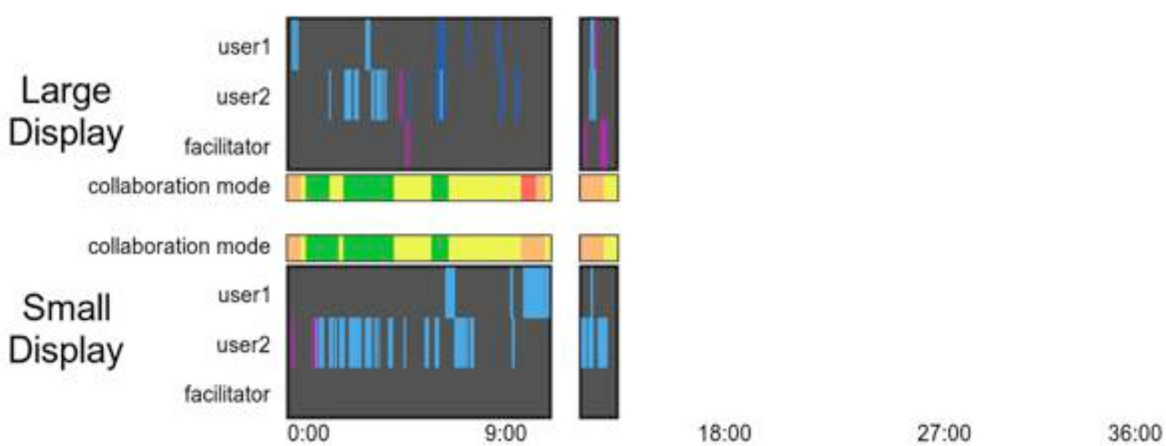


APPENDIX F (Continued)

Absolute *advanced data-synchronization options* timelines (continued):

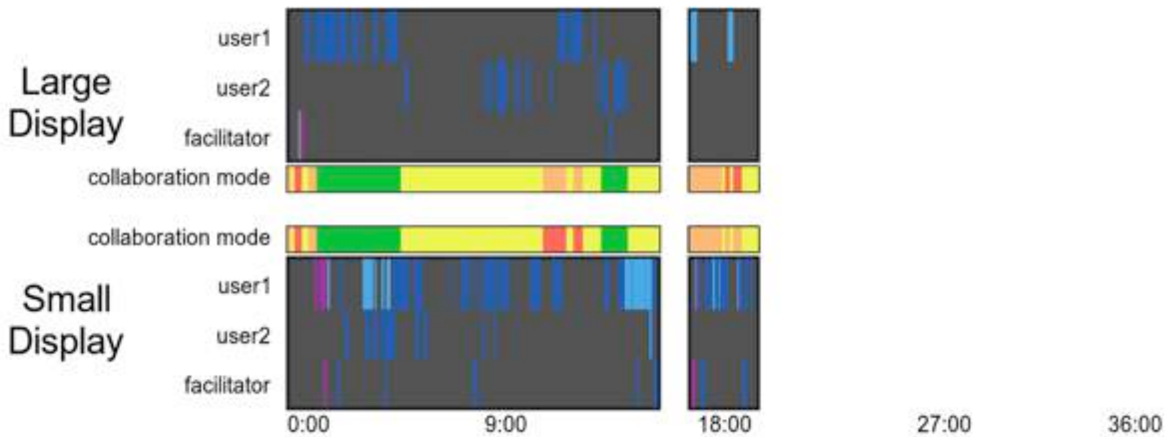


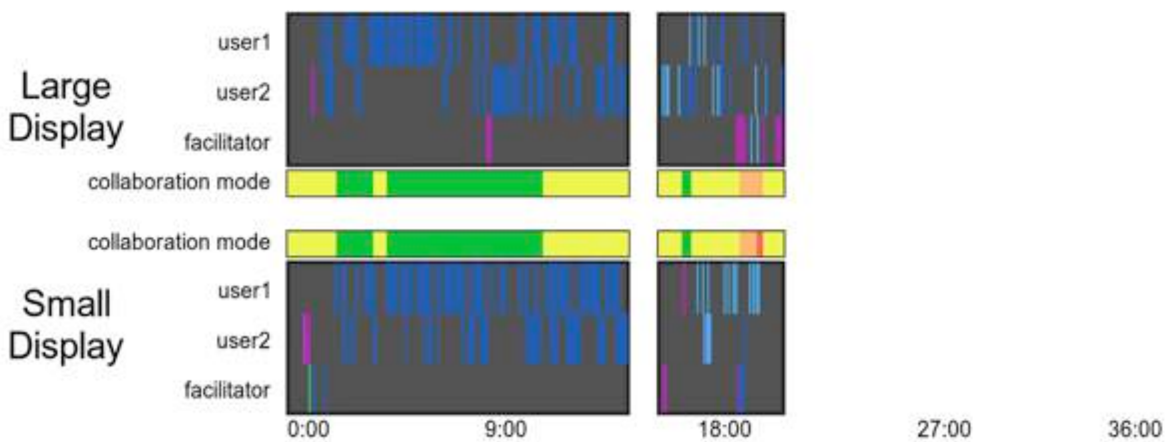


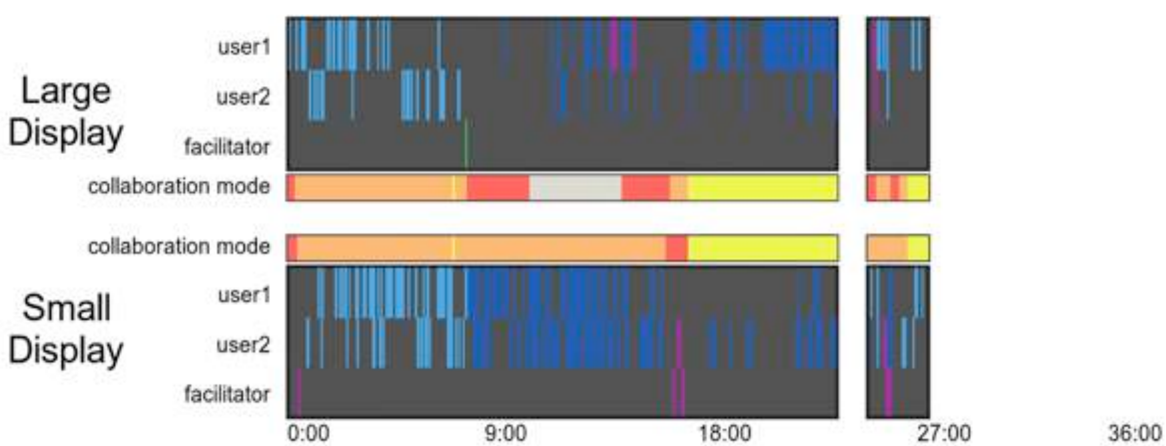


APPENDIX F (Continued)

Absolute *advanced data-synchronization options* timelines (continued):

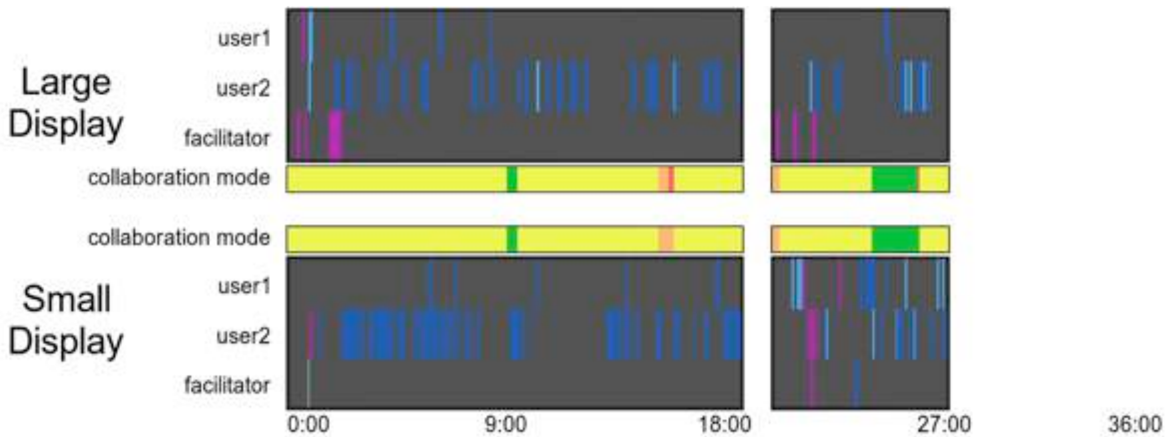


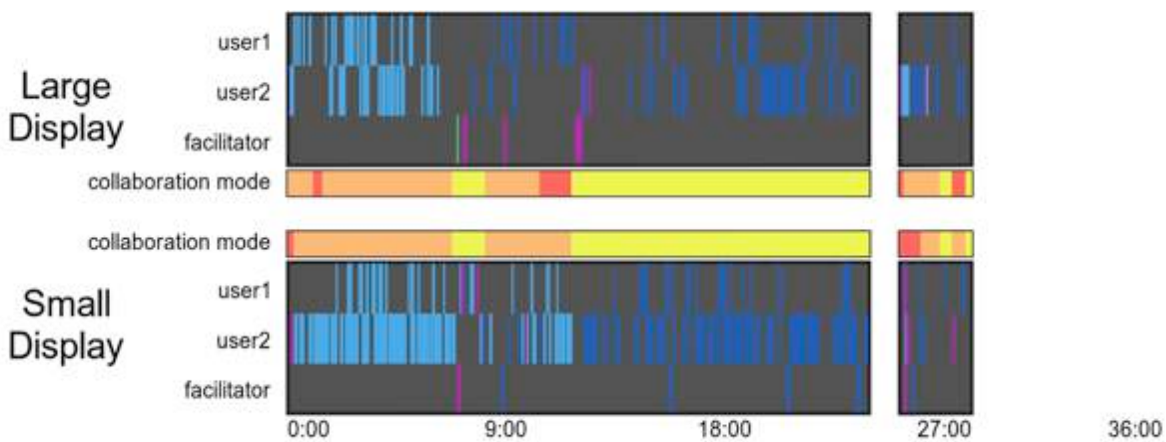




APPENDIX F (Continued)

Absolute *advanced data-synchronization options* timelines (continued):



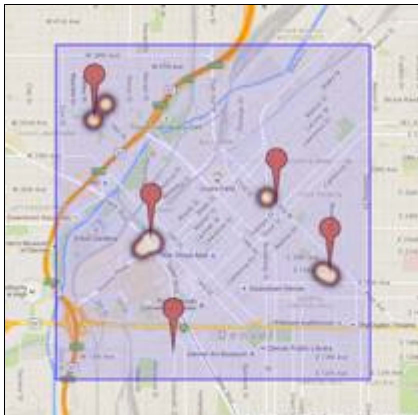
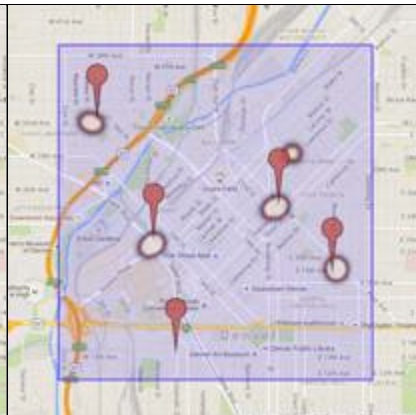
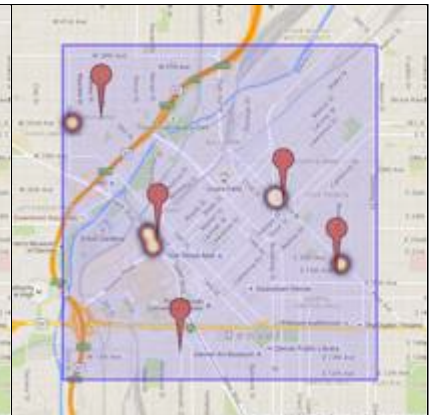


APPENDIX G

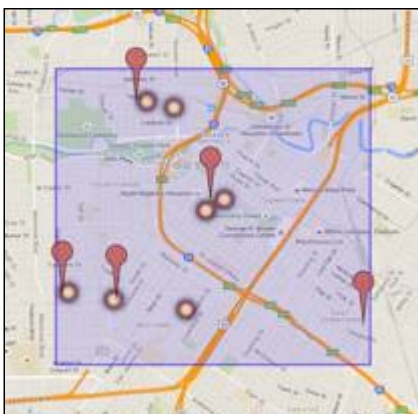
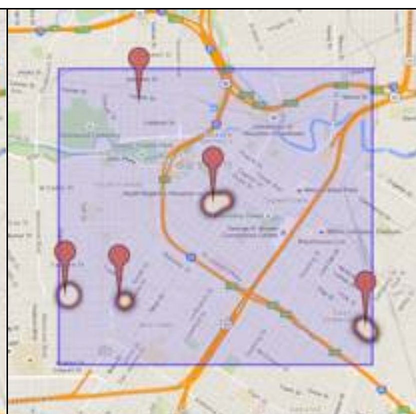
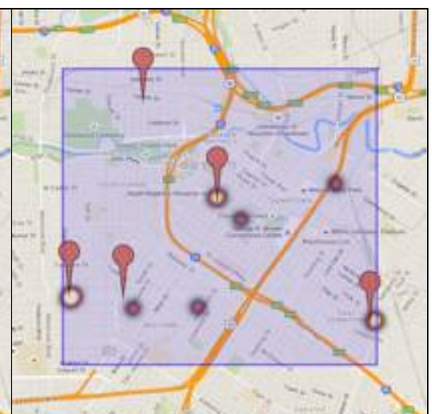
Heatmap visualizations of answers groups gave for the two tasks in the formal user study.

The first set shows answers for task 1 (finding 2-4 potential locations for coffee shops). The red markers show the 5 possible ground truth answers.

Denver

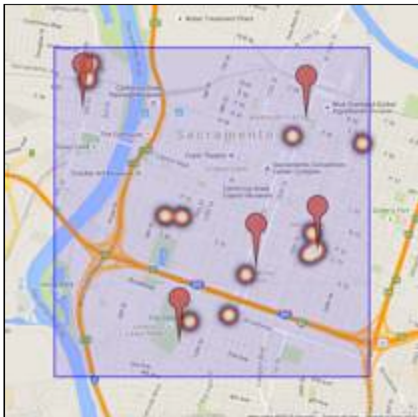
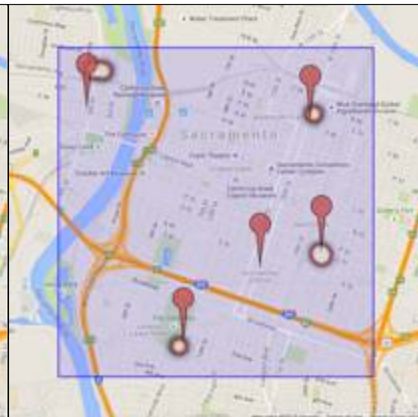
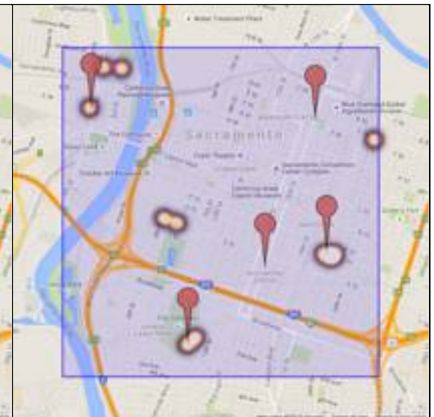
		
Data-pushing	Data-duplication	Advanced data-synchronization options

Houston

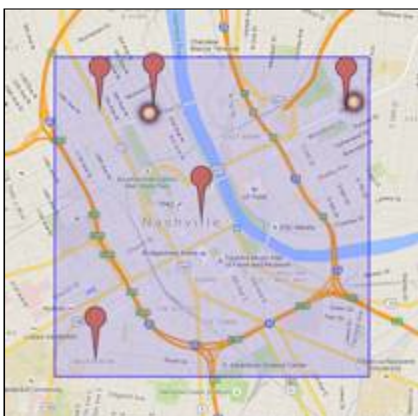
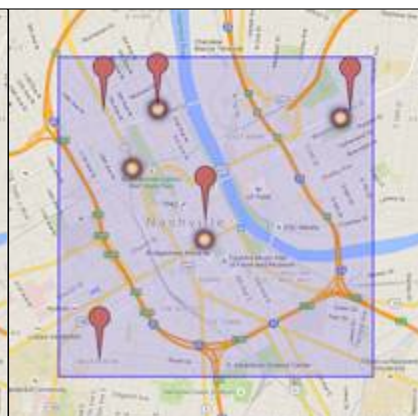
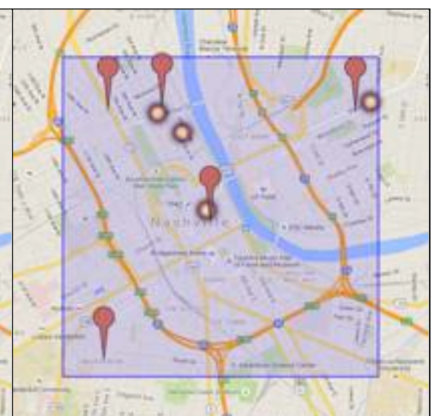
		
Data-pushing	Data-duplication	Advanced data-synchronization options

APPENDIX G (Continued)

Sacramento

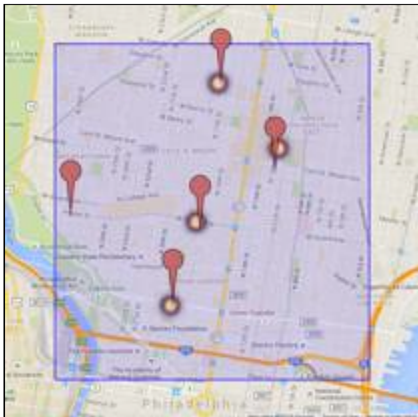
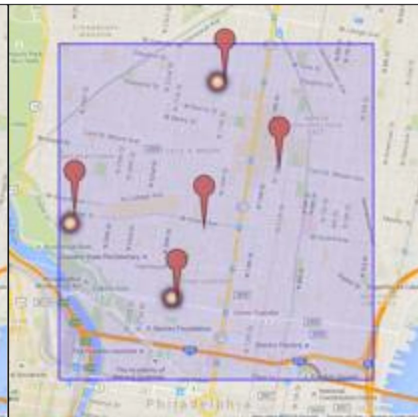
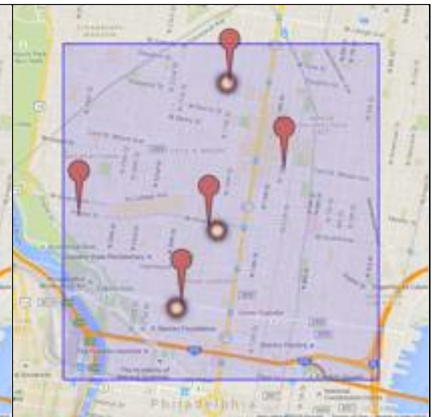
		
Data-pushing	Data-duplication	Advanced data-synchronization options

Nashville

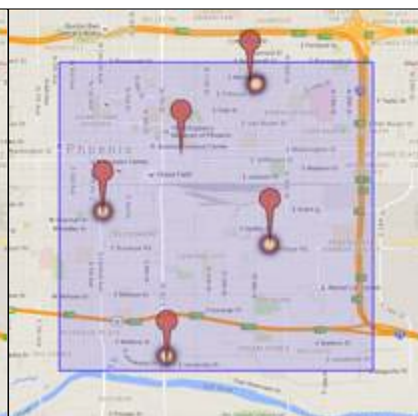
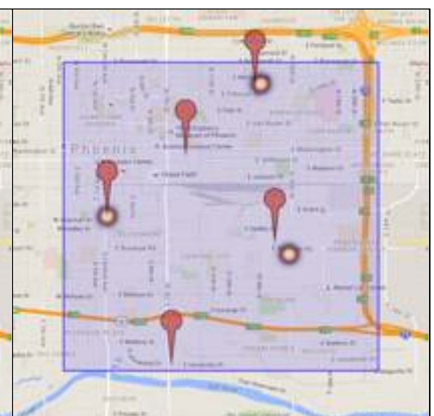
		
Data-pushing	Data-duplication	Advanced data-synchronization options

APPENDIX G (Continued)

Philadelphia

		
Data-pushing	Data-duplication	Advanced data-synchronization options

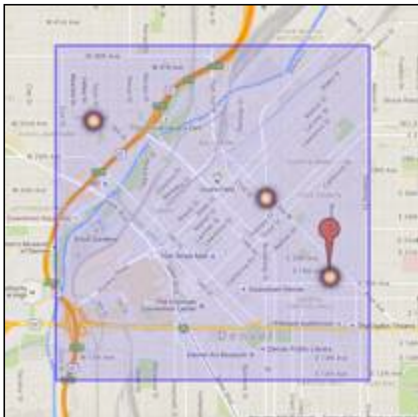
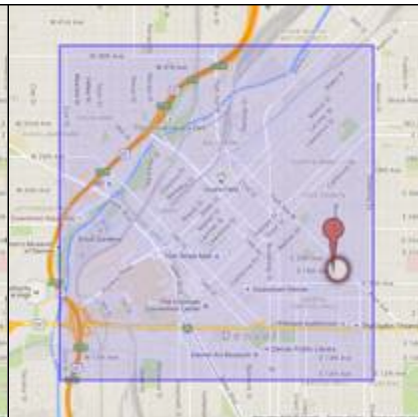
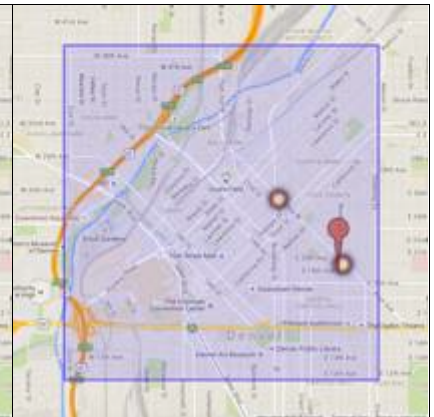
Phoenix

		
Data-pushing	Data-duplication	Advanced data-synchronization options

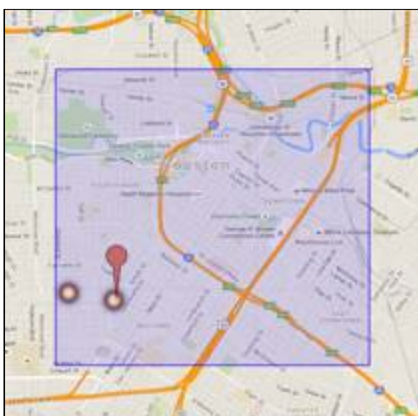
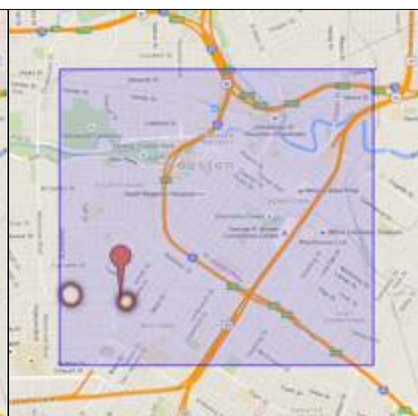
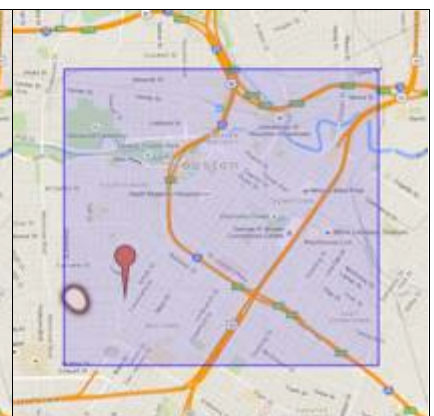
APPENDIX G (Continued)

The second set shows answers for task 2 (finding 1 final location for a coffee shop). The red marker shows the best possible ground truth answer.

Denver

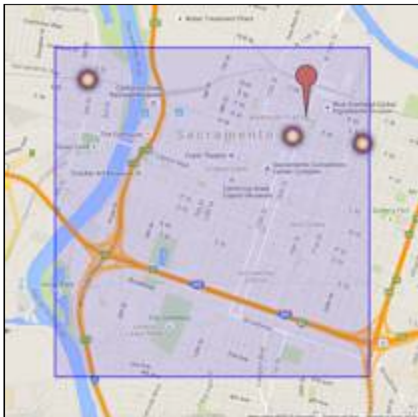
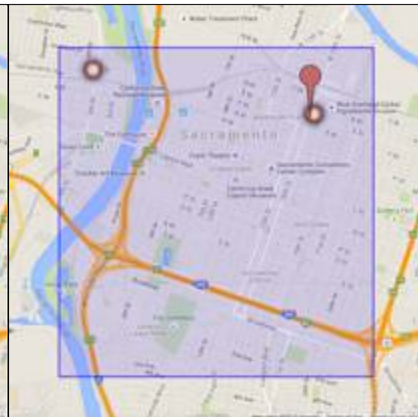
		
Data-pushing	Data-duplication	Advanced data-synchronization options

Houston

		
Data-pushing	Data-duplication	Advanced data-synchronization options

APPENDIX G (Continued)

Sacramento

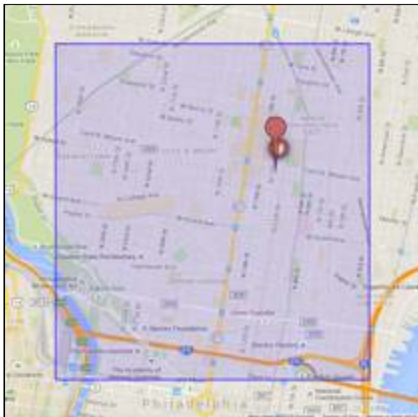
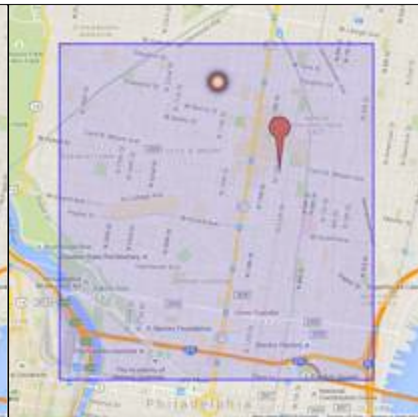
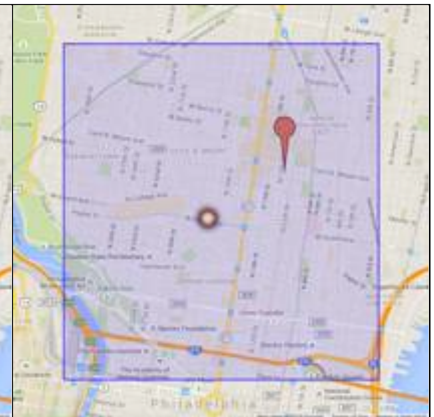
		
Data-pushing	Data-duplication	Advanced data-synchronization options

Nashville

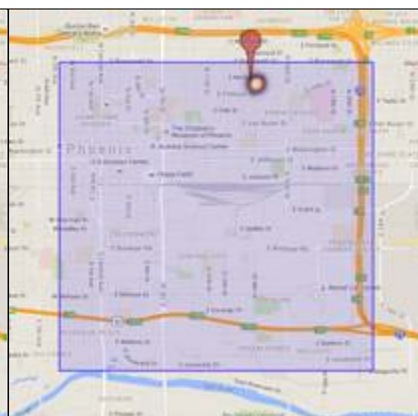
		
Data-pushing	Data-duplication	Advanced data-synchronization options

APPENDIX G (Continued)

Philadelphia

		
Data-pushing	Data-duplication	Advanced data-synchronization options

Phoenix

		
Data-pushing	Data-duplication	Advanced data-synchronization options

APPENDIX H

Institutional Review Board approval letter from the University of Illinois at Chicago for the Longitudinal User Study.

UNIVERSITY OF ILLINOIS
AT CHICAGO

Office for the Protection of Research Subjects (OPRS)
Office of the Vice Chancellor for Research (MC 672)
203 Administrative Office Building
1737 West Polk Street
Chicago, Illinois 60612-7227

Exemption Granted

February 7, 2015

Thomas Marrinan, BS, BA
Computer Science
851 S. Morgan Street
SEO Room 218, M/C 152
Chicago, IL 60607
Phone: (312) 996-3002 / Fax: (312) 413-7585

RE: Research Protocol # 2015-0056
"Data-Intensive Remote Collaboration using Scalable Visualizations in Heterogeneous Display Spaces"

Sponsors: None

Upon receipt, please submit – via amendment - a copy of the University of Hawaii IRB approval letter or exemption determination.

Dear Mr. Marrinan:

Your Claim of Exemption was reviewed on February 7, 2015 and it was determined that your research meets the criteria for exemption. You may now begin your research.

Exemption Period:	**February 7, 2015 – February 7, 2018**
Performance Site(s):	UIC, see text box above
Subject Population:	Adult (18+ years) subjects only
Number of Subjects:	15

The specific exemption category under 45 CFR 46.101(b) is:
(2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: (i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.

Please note the Review History of this submission:

Receipt Date	Submission Type	Review Process	Review Date	Review Action
01/13/2015	Initial Review	Exempt	01/27/2015	Modifications Required
01/28/2015	Response to Modifications	Exempt	02/07/2015	Approved

Phone: 312-996-1711 http://www.uic.edu/depts/ovcr/oprs/ Fax: 312-413-2929

APPENDIX H (Continued)

You are reminded that investigators whose research involving human subjects is determined to be exempt from the federal regulations for the protection of human subjects still have responsibilities for the ethical conduct of the research under state law and UIC policy. Please be aware of the following UIC policies and responsibilities for investigators:

1. Amendments You are responsible for reporting any amendments to your research protocol that may affect the determination of the exemption and may result in your research no longer being eligible for the exemption that has been granted.

2. Record Keeping You are responsible for maintaining a copy all research related records in a secure location in the event future verification is necessary, at a minimum these documents include: the research protocol, the claim of exemption application, all questionnaires, survey instruments, interview questions and/or data collection instruments associated with this research protocol, recruiting or advertising materials, any consent forms or information sheets given to subjects, or any other pertinent documents.

3. Final Report When you have completed work on your research protocol, you should submit a final report to the Office for Protection of Research Subjects (OPRS).

4. Information for Human Subjects UIC Policy requires investigators to provide information about the research protocol to subjects and to obtain their permission prior to their participating in the research. The information about the research protocol should be presented to subjects in writing or orally from a written script. When appropriate, the following information must be provided to all research subjects participating in exempt studies:
   a. The researchers affiliation; UIC, JBVMAC or other institutions,
   b. The purpose of the research,
   c. The extent of the subject's involvement and an explanation of the procedures to be followed,
   d. Whether the information being collected will be used for any purposes other than the proposed research,
   e. A description of the procedures to protect the privacy of subjects and the confidentiality of the research information and data,
   f. Description of any reasonable foreseeable risks,
   g. Description of anticipated benefit,
   h. A statement that participation is voluntary and subjects can refuse to participate or can stop at any time,
   i. A statement that the researcher is available to answer any questions that the subject may have and which includes the name and phone number of the investigator(s).
   j. A statement that the UIC IRB/OPRS or JBVMAC Patient Advocate Office is available if there are questions about subject's rights, which includes the appropriate phone numbers.

Please be sure to:

→ Use your research protocol number (2015-0056) on any documents or correspondence with the IRB concerning your research protocol.

APPENDIX H (Continued)

We wish you the best as you conduct your research. If you have any questions or need further help, please contact the OPRS office at (312) 996-1711 or me at (312) 355-2908. Please send any correspondence about this protocol to OPRS at 203 AOB, M/C 672.

Sincerely,

Charles W. Hoehne, B.S., C.I.P.
Assistant Director
Office for the Protection of Research Subjects

cc: Robert Sloan, Computer Science, M/C 152
Andrew Johnson, Computer Science, M/C 154

APPENDIX H (Continued)

Institutional Review Board approval letter from the University of Hawai'i at Mānoa for the Longitudinal User Study.

UNIVERSITY of HAWAI'I® MĀNOA

Office of Research Compliance
Human Studies Program

September 25, 2015

TO: Dylan Kobayashi
Jason Leigh
Principal Investigators
Information & Computer Science

FROM: Denise A. Lin-DeShetler, MPH, MA
Director

SUBJECT: CHS #23396- "Data-Intensive Remote Collaboration Using Scalable Visualizations in Heterogeneous Display Spaces"

This letter is your record of the Human Studies Program approval of this study as exempt.

On September 25, 2015, the University of Hawai'i (UH) Human Studies Program approved this study as exempt from federal regulations pertaining to the protection of human research participants. The authority for the exemption applicable to your study is documented in the Code of Federal Regulations at 45CFR 46.101(b)(Exempt Category 2).

Exempt studies are subject to the ethical principles articulated in The Belmont Report, found at http://www.hawaii.edu/irb/html/manual/appendices/A/belmont.html.

Exempt studies do not require regular continuing review by the Human Studies Program. However, if you propose to modify your study, you must receive approval from the Human Studies Program prior to implementing any changes. You can submit your proposed changes via email at uhirb@hawaii.edu. (The subject line should read: Exempt Study Modification,) The Human Studies Program may review the exempt status at that time and request an application for approval as non-exempt research.

In order to protect the confidentiality of research participants, we encourage you to destroy private information which can be linked to the identities of individuals as soon as it is reasonable to do so. Signed consent forms, as applicable to your study, should be maintained for at least the duration of your project.

This approval does not expire. However, please notify the Human Studies Program when your study is complete. Upon notification, we will close our files pertaining to your study.

If you have any questions relating to the protection of human research participants, please contact the Human Studies Program at 956-5007 or uhirb@hawaii.edu. We wish you success in carrying out your research project.

1960 East-West Road
Biomedical Sciences Building B104
Honolulu, Hawai'i 96822
Telephone: (808) 956-5007
Fax: (808) 956-8683
An Equal Opportunity/Affirmative Action Institution

APPENDIX H (Continued)

Institutional Review Board approval letter from the University of Illinois at Chicago for the amendment to the Longitudinal User Study.

University of Illinois
at Chicago

Office for the Protection of Research Subjects (OPRS)
Office of the Vice Chancellor for Research (MC 672)
203 Administrative Office Building
1737 West Polk Street
Chicago, Illinois 60612-7227

Exemption Determination
Amendment to Research Protocol – Exempt Review
UIC Amendment #1

October 12, 2015

Thomas Marrinan, BS, BA
Computer Science
851 S. Morgan Street
Room 1120, M/C 152
Chicago, IL 60607
Phone: (312) 996-3002 / Fax: (312) 413-7585

RE: Protocol # 2015-0056
"Data-Intensive Remote Collaboration using Scalable Visualizations in Heterogeneous Display Spaces"

Dear Mr. Marrinan:

The OPRS staff/members of Institutional Review Board (IRB) #7 have reviewed this amendment to your research, and have determined that your research protocol continues to meet the criteria for exemption as defined in the U. S. Department of Health and Human Services Regulations for the Protection of Human Subjects [(45 CFR 46.101(b)].

The specific exemption category under 45 CFR 46.101(b) is:
(2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: (i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.

You may now implement the amendment in your research.

Please note the following information about your approved amendment:

Exemption Period:	**October 12, 2015 – October 12, 2018**
Amendment Approval Date:	October 12, 2015

Amendment:
Summary: UIC Amendment #1 dated February 2, 2015 and submitted to OPRS on October 2, 2015 is an investigator-initiated amendment adding the University as a performance site.

You are reminded that investigators whose research involving human subjects is determined to be

APPENDIX H (Continued)

exempt from the federal regulations for the protection of human subjects still have responsibilities for the ethical conduct of the research under state law and UIC policy. Please be aware of the following UIC policies and responsibilities for investigators:

1. Amendments You are responsible for reporting any amendments to your research protocol that may affect the determination of the exemption and may result in your research no longer being eligible for the exemption that has been granted.

2. Record Keeping You are responsible for maintaining a copy all research related records in a secure location in the event future verification is necessary, at a minimum these documents include: the research protocol, the claim of exemption application, all questionnaires, survey instruments, interview questions and/or data collection instruments associated with this research protocol, recruiting or advertising materials, any consent forms or information sheets given to subjects, or any other pertinent documents.

3. Final Report When you have completed work on your research protocol, you should submit a final report to the Office for Protection of Research Subjects (OPRS).

4. Information for Human Subjects UIC Policy requires investigators to provide information about the research protocol to subjects and to obtain their permission prior to their participating in the research. The information about the research protocol should be presented to subjects in writing or orally from a written script. When appropriate, the following information must be provided to all research subjects participating in exempt studies:

    a. The researchers affiliation; UIC, JB VAMC or other institutions,
    b. The purpose of the research,
    c. The extent of the subject's involvement and an explanation of the procedures to be followed,
    d. Whether the information being collected will be used for any purposes other than the proposed research,
    e. A description of the procedures to protect the privacy of subjects and the confidentiality of the research information and data,
    f. Description of any reasonable foreseeable risks,
    g. Description of anticipated benefit,
    h. A statement that participation is voluntary and subjects can refuse to participate or can stop at any time,
    i. A statement that the researcher is available to answer any questions that the subject may have and which includes the name and phone number of the investigator(s).
    j. A statement that the UIC IRB/OPRS or JB VAMC Patient Advocate Office is available if there are questions about subject's rights, which includes the appropriate phone numbers.

Please be sure to:

→Use your research protocol number (2015-0056) on any documents or correspondence with the

APPENDIX H (Continued)

IRB concerning your research protocol.

We wish you the best as you conduct your research. If you have any questions or need further help, please contact me at (312) 355-2908 or the OPRS office at (312) 996-1711. Please send any correspondence about this protocol to OPRS at 203 AOB, M/C 672.

Sincerely,

Charles W. Hoehne
Assistant Director, IRB #7
Office for the Protection of Research Subjects

cc: Robert Sloan, Computer Science, M/C 152
Andrew Johnson, Computer Science, M/C 154

APPENDIX H (Continued)

Institutional Review Board approval letter from the University of Illinois at Chicago for the Formal User Study.

UNIVERSITY OF ILLINOIS
AT CHICAGO

Office for the Protection of Research Subjects (OPRS)
Office of the Vice Chancellor for Research (MC 672)
203 Administrative Office Building
1737 West Polk Street
Chicago, Illinois 60612-7227

Exemption Granted

June 3, 2015

Thomas Marrinan, BS, BA
Computer Science
851 S. Morgan Street
Room 1120, M/C 152
Chicago, IL 60607
Phone: (312) 996-3002 / Fax: (312) 413-7585

RE: Research Protocol # 2015-0556
"Data-Intensive Remote Collaboration using Scalable Visualizations in Heterogeneous Display Spaces"
Sponsor(s): None

Dear Mr. Marrinan:

Your Claim of Exemption was reviewed on June 1, 2015 and it was determined that your research protocol meets the criteria for exemption as defined in the U. S. Department of Health and Human Services Regulations for the Protection of Human Subjects [(45 CFR 46.101(b)]. You may now begin your research.

Exemption Period:	**June 1, 2015 – June 1, 2018**
Performance Site:	UIC
Subject Population:	Adult (18+ years) subjects only
Number of Subjects:	96

The specific exemption category under 45 CFR 46.101(b) is:
(2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: (i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.

You are reminded that investigators whose research involving human subjects is determined to be exempt from the federal regulations for the protection of human subjects still have responsibilities for the ethical conduct of the research under state law and UIC policy. Please be aware of the following UIC policies and responsibilities for investigators:

1. Amendments You are responsible for reporting any amendments to your research protocol that may affect the determination of the exemption and may result in your research no longer being eligible for the exemption that has been granted.

APPENDIX H (Continued)

2. Record Keeping You are responsible for maintaining a copy all research related records in a secure location in the event future verification is necessary, at a minimum these documents include: the research protocol, the claim of exemption application, all questionnaires, survey instruments, interview questions and/or data collection instruments associated with this research protocol, recruiting or advertising materials, any consent forms or information sheets given to subjects, or any other pertinent documents.
3. Final Report When you have completed work on your research protocol, you should submit a final report to the Office for Protection of Research Subjects (OPRS).
4. Information for Human Subjects UIC Policy requires investigators to provide information about the research protocol to subjects and to obtain their permission prior to their participating in the research. The information about the research protocol should be presented to subjects in writing or orally from a written script. When appropriate, the following information must be provided to all research subjects participating in exempt studies:
   a. The researchers affiliation; UIC, JBVMAC or other institutions,
   b. The purpose of the research,
   c. The extent of the subject's involvement and an explanation of the procedures to be followed,
   d. Whether the information being collected will be used for any purposes other than the proposed research,
   e. A description of the procedures to protect the privacy of subjects and the confidentiality of the research information and data,
   f. Description of any reasonable foreseeable risks,
   g. Description of anticipated benefit,
   h. A statement that participation is voluntary and subjects can refuse to participate or can stop at any time,
   i. A statement that the researcher is available to answer any questions that the subject may have and which includes the name and phone number of the investigator(s).
   j. A statement that the UIC IRB/OPRS or JBVMAC Patient Advocate Office is available if there are questions about subject's rights, which includes the appropriate phone numbers.

Please be sure to:

→Use your research protocol number (listed above) on any documents or correspondence with the IRB concerning your research protocol.

We wish you the best as you conduct your research. If you have any questions or need further help, please contact me at (312) 355-2908 or the OPRS office at (312) 996-1711. Please send any correspondence about this protocol to OPRS at 203 AOB, M/C 672.

Sincerely,

Charles W. Hoehne, B.S., C.I.P.
Assistant Director
Office for the Protection of Research Subjects

cc: Robert Sloan, Computer Science, M/C 152
Andrew Johnson, Computer Science, M/C 152

VITA

THOMAS J. MARRINAN

EDUCATION

University of Illinois at Chicago (UIC), Chicago, IL (2010 – 2015 anticipated)

Doctor of Philosophy – Electronic Visualization Laboratory (EVL) and the Department of Computer Science. Dissertation *Data-Intensive Remote Collaboration using Scalable Visualizations in Heterogeneous Display Spaces* identifies factors that contribute to successful big data collaboration across distance using large-scale tiled display walls. The research involves designing and implementing multi-user interaction and data synchronization schemes within the web-based framework of SAGE2™. After developing prototypes, longitudinal observations and a focused user study will be utilized to uncover collaboration paradigms for partially distributed teams. It is believed that this research will lower the barriers for coordinating group work amongst remotely located collaborators and lead to the better utilization of large-scale tiled display walls.

Selected Course Projects

- **Computer Graphics II – Planetary Systems**
  Created a hybrid visualization for 2D small-multiples and interactive 3D graphics specifically designed for ultra high-resolution displays. This project investigated the ever-increasing number of exoplanets that are being discovered in order to compare other planetary systems with our own solar system.
- **Visualization and Visual Analytics II – Lake Michigan**
  Created an interactive 3D application of Lake Michigan incorporating water temperature, flow, and wind velocity from 2006-2010. This application uses volume visualization and an array of GUI elements to allow users to gain insight about the dynamics of a vast body of water.
- **Multimedia Systems – Red Light / Green Light**
  Worked with a team to design a virtual Red Light / Green Light game with the goal of encouraging children to be active. I was in charge of streaming accelerometer data from smartphones and creating a pedometer app to infer footsteps. The phones were then placed in the pockets of users during the game.
- **Video Game Design – Siege Breakers**
  Led a development team in creating a castle siege video game using the Unity engine. The process incorporated storyboarding, 3D modeling and animation, artificial intelligence, and gameplay design.

GPA – 3.91 / 4.00

Drake University, Des Moines, IA (2006 – 2010)

Bachelor of Science – Computer Science.

Bachelor of Arts – Graphic Design

GPA – 3.76 / 4.00

PROFESSIONAL EXPERIENCE

Research Assistant, EVL, UIC (2011 – Present)

- Worked with a team to develop the SAGE2™ Collaborative Multi-user Operating Environment, a project funded by the National Science Foundation that has been awarded $5 million over five years. SAGE2™ is a next generation collaborative platform for large ultra high-resolution shared displays. I integrated cloud-based and web-browser technologies (Node.js, HTML, CSS, JavaScript, WebGL, D3, WebRTC) into an environment suited for data intensive problem solving in authentic scenarios and created advanced remote collaboration features for partially distributed teams to share and synchronize data in real-time.

- Visualization and high-performance computing expert (using Pthreads, MPI, PETSc, C++, OpenGL, VTK) for simulating the human cerebral vascular system. This research was a collaboration between the UIC Departments of BioEngineering, Neurosurgery, and Computer Science. I developed graphical stereoscopic visual applications for a standard single monitor as well as an immersive CAVE platform. The visualization focused on viewing 4D-dimensional (space and time) medical images from angiography, MRI, CT, and blood flow simulations for individual patients.
- Co-led demonstrations and laboratory tours for CS Open Houses, Chicago Ideas Week, prospective collaborators, and high school and middle school student groups.

Teaching Assistant, UIC (2011 – 2012)

- Assisted with graduate courses in computer graphics and compiler construction. Gave lectures, helped create project assignments, and evaluated student written code.

Research Intern, Accenture (2011)

- Developed a database resource identifier to convert transactions into Petri Net Markup Language models. Worked with a research team to identify and prevent SQL database deadlock.

Course Grader, UIC (2010)

- Evaluated student homework assignments and course examinations.

Research Assistant, Drake University (2009 – 2010)

- Developed a multi-dimensional data visualization tool. This research led to more efficient identification of chromatography systems used for modifying the selectivity of the separation in complex chemical mixtures.
- Developed a tool for volumetric visualization of data from the Hubble Space Telescope in order to help astronomers understand the kinematics of ionized gas in the nuclear regions of Seyfert galaxies, the most common active galactic nuclei.

SELECTED HONORS AND AWARDS

Best Paper (2014)

- *SAGE2: A New Approach for Data Intensive Collaboration Using Scalable Resolution Shared Displays*. IEEE CollaborateCom 2014.

Physics Today – Cover (2013)

- My research on visualizing the human cerebral vascular system in the CAVE2™ Hybrid Reality System and I are depicted on the cover of the *Physics Today* journal.

NSF Highlighted Project (2013)

- NSF named my research as one of its twelve highlighted projects for the year in its *Budget Request to Congress* with a short description about how our “State-of-the-Art Virtual Reality System is the Key to Medical Discovery.”

Best Poster Honorable Mention (2012)

- *Whole-Brain Vascular Reconstruction, Simulation, and Visualization*. IEEE VisWeek 2012.

The Images of Research – 1st Place (2012)

- *Artificially Created Cortical Functional Blood Unit*. UIC annual interdisciplinary exhibit competition that showcases the breadth and diversity of research.

Outstanding Student in Computer Science (2010)

- Drake University College of Arts and Sciences Awards Ceremony.

PROFESSIONAL ACTIVITIES

Conferences

- **Supercomputing** 2014. SAGE BoF Presentation, "Scalable Adaptive Graphics Environment (SAGE) for Global Collaboration."
- **CollaborateCom** 2014. Research Paper Presentation, "SAGE2: A New Approach for Data Intensive Collaboration Using Scalable Resolution Shared Displays."
- **VisWeek** 2012. Poster Presentation, "Whole-Brain Vascular Reconstruction, Simulation, and Visualization."

Grant Writing

- **NIH R21**. "Stereoscopic 4D Modeling and Interactive Virtual Exploration of Cerebral Vasculature." Impact/Priority Score: 28, Percentile: 19%, Funding Rate: 14%.

PUBLICATIONS

Conference Proceedings

- **T. Marrinan**, J. Aurisano, A. Nishimoto, K. Bharadwaj, V. Mateevitsi, L. Renambot, L. Long, A. Johnson, and J. Leigh, "SAGE2: A New Approach for Data Intensive Collaboration Using Scalable Resolution Shared Displays," in *Proceedings of the IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing*, 2014 (to appear).
- T. Urness, **T. Marrinan**, A. Johnson, and M. Vitha, "Multivariate Visualization of Chromatographic Systems," in *Proceedings of SPIE-IS&T Electronic Imaging*, 2011, vol. 7868.

Journals

- L. Renambot, **T. Marrinan**, J.Aurisano, A. Nishimoto, V. Mateevitsi, K. Bharadwaj, L. Long, A. Johnson, M. Brown, and J. Leigh, "SAGE2: A Collabora6on Portal for Scalable Resolu6on Displays," to be published in *Future Generation Computer Systems*, 2015
- A. Linninger, I. Gould, **T. Marrinan**, C. Hsu, M. Chojecki, and A. Alaraj, "Cerebral Microcirculation and Oxygen Tension in the Human Secondary Cortex," in *Annals of Biomedical Engineering*, 2013, vol. 41, no. 11, pp. 2264-2284.
- **T. Marrinan**, T. Urness, C. Nelson, K. Kreimeyer, and J. Mirocha, "Understanding and Interpreting Multivalued Astronomical Data," in *IEEE Computer Graphics and Applications*, 2010, vol. 30, no. 5, pp. 12-17.
- A. Johnson, M. Vitha, T. Urness, and **T. Marrinan**, "System Selectivity Cube: A 3D Visualization Tool for Comparing the Selectivity of Gas Chromatography, Supercritical-Fluid Chromatography, High-Pressure Liquid Chromatography, and Micellar Electrokinetic Capillary Chromatography Systems," in *Analytical Chemistry*, 2010, vol. 82, no. 14, pp. 6251-6258.

Posters

- **T. Marrinan**, I. Gould, C. Hsu, and A. Linninger, "Whole-Brain Vascular Reconstruction, Simulation, and Visualization," in *IEEE VisWeek*, 2012.

PROFESSIONAL MEMBERSHIPS

Association for Computer Machinery (ACM)

European Alliance for Innovation (EAI)

EXTRACURRICULAR ACTIVITIES

Recreational Sports Leagues

- Flag Football
- Beach Volleyball

SCUBA Diving

- NAUI Passport Certification

PERSONAL

Image Generation / Manipulation – Adobe Creative Suite: photo editing and compositing, icon design

Animation – Blender: modeling, rigging, key frame and motion capture animation

Video Editing – Final Cut Pro / iMovie: editing, transitions, audio/video mixing