

Multi-person Spatial Interaction in a Large Immersive Display Using Smartphones as Touchpads

Gyanendra Sharma and Richard J. Radke

Rensselaer Polytechnic Institute
Troy, NY
gyanendra.sharma870@gmail.com, rjradke@ecse.rpi.edu

Abstract. In this paper, we present a multi-user interaction interface for a large immersive space that supports simultaneous screen interactions by combining (1) user input via personal smartphones and Bluetooth microphones, (2) spatial tracking via an overhead array of Kinect sensors, and (3) WebSocket interfaces to a webpage running on the large screen. Users are automatically, dynamically assigned personal and shared screen sub-spaces based on their tracked location with respect to the screen, and use a webpage on their personal smartphone for touchpad-type input. We report user experiments using our interaction framework that involve image selection and placement tasks, with the ultimate goal of realizing display-wall environments as viable, interactive workspaces with natural multimodal interfaces.

Keywords: spatial intelligence; immersive spaces; interaction design; multi person interaction; smart phones



Fig. 1: Multiple users simultaneously interacting with a large screen using their smartphones and voices, coupled with spatial information about their physical locations. This panoramic image shows a 5m tall 360-degree display wall with a 44m perimeter. Please refer to the video hosted at https://youtu.be/CYqXKyHTO_U for detailed interaction examples.

1 Introduction

Designing interactive multi-user interfaces for large-scale immersive spaces requires accommodations that go beyond conventional input mechanisms. In recent years, incorporating multi-layered modalities such as personal touchscreen

devices, voice commands, and mid-air gestures have evolved as viable alternatives [1,2,3,4]. Especially in projector-based displays like the one discussed here, distant interaction via smartphone-like devices plays a pivotal role [5].

Apart from the input modes of interaction, the size and scale of such spaces greatly benefit from contextualizing user locations within the space for interaction design purposes [6,7,8]. This is especially true for large enclosed displays such as CAVE [9,10], CAVE2 [11], CUBE [12] and CRAIVE [13]. Representing physical user locations on such screen spaces presents considerable challenges due to spatial ambiguities compared to flat display walls.

In this paper, we present mechanisms for multiple users to simultaneously interact with a large immersive screen by incorporating three components: users' physical locations obtained from external range sensors, ubiquitous input devices such as smartphones and Bluetooth microphones, and automatic contextualization of personal vs. shared screen areas. As shown in Figure 1, discrete personal interaction regions appear on two sides of a rectangular enclosed screen, where users freely move to make spatial selections and manipulate or generate relevant images. The shared screen region between the two sides can be simultaneously manipulated by multiple users to create a desired layout based on combinations of pre-selected text with user-curated images.

Our method and overall architecture allows multiple users to interface with the large visually immersive space in a natural way. Integrating personal devices and voice along with spatial intelligence to define personal and shared interaction areas opens avenues to use the space for applications such as classroom learning, collaboration, and game play.

We designed controlled laboratory experiments with 14 participants to test the usability, intuitiveness and comfort of this multimodal, multi-user-to-large-screen interaction interface. Based on the results, we observe that the designed mechanism is easy to use and adds a degree of fun and enjoyment to users while in the space.

The rest of the paper is organized as follows. In the next section, we discuss prior work on spatial intelligence and interaction mechanisms in the context of large immersive spaces. We then introduce the overall system, consisting of our proposed multi-modal spatial interaction design along with the overall system architecture. Next we describe user studies to validate our system design for interaction and present the results. We conclude by discussing our findings and potential future work.

2 Background and Related Work

Our system is inspired by a diverse body of prior work, generally related to spatial sense-making in large immersive spaces, personal vs. shared spaces in large screens, multi-user support, and interactions using ubiquitous devices such as smartphones.

Spatial intelligence in immersive spaces Microsoft Kinects and similar 3-D sensors have been widely used for user locations or gestural interpretation in the context of various large screens [14,15,16] and common spaces [17]. Research has primarily been focused on developing mid-air gestures and other interaction mechanisms using methods similar to ray-casting, which require knowledge of spatial layout and users’ physical locations [18]. A unique aspect of our system is an overhead Kinect array that allows many users to be simultaneously tracked and their locations to be correlated to screen coordinates and workspaces.

Personal vs. shared spaces In terms of demarcating public vs. personal spaces within large screens, Vogel and Balakrishnan [1] discussed how large displays can accommodate and transition between public and personal interaction modes based on several factors. This thread of research extends to privacy-supporting infrastructure and technologies [19,20]. Wallace et al. [21] recently studied approaches to defining personal spaces in the context of a large touch screen display, which we cannot directly incorporate in our system but inspired our design considerations.

Multi-user support Realizing large immersive spaces as purposeful collaboration spaces through multi-user interaction support remains an active area of research [22]. Various approaches such as visualization of group interaction [23], agile team collaboration [24], along with use cases such as board meeting scenarios [25], have been proposed. The Collaborative Newspaper by Lander et al. [26] and Wordster by Luo et al. [27] showed how multiple users can interact at the same time with a large display. Doshi et al. presented a multi-user application for conference scheduling using digital “sticky notes” on a large screen [28].

Smartphones as interaction devices The limitations of conventional input devices for natural interactions with pervasive displays have led to several innovations, for example allowing ubiquitous devices such as smartphones to be used as interaction devices. Such touchscreen devices allow for greater flexibility and diversity in how interaction mechanisms with pervasive displays are materialized. Earlier concepts such as the one proposed by Ballagas et al. [29] have evolved towards more native web-based or standalone application-based interfaces. For instance, Baldauf et al. developed a web-based remote control to interact with public screens called *ATREUS* [30]. Beyond the touchscreen element of smartphones, researchers have investigated combining touch and air gestures [31], 3D interaction mechanisms [32] and using built-in flashlights for interaction [33].

3 System Design

Our system was designed and implemented in the CRAIVE-Lab, a large immersive display wall environment with a 5m tall 360-degree front-projected screen enclosing a 12m × 10m walkable area [13]. The screen is equipped with 8 1200 × 1920

resolution projectors, resulting in an effective 1200×14500 pixel display, and contains a network of 6 overhead downward-pointed Kinect sensors for visual tracking of multiple participants.

3.1 Spatial sense-making

Large immersive spaces have exciting potential to support simultaneous multi-user interactions. Flat 2D displays can support such functions simply by using multiple input devices with minimal consideration for physical user locations. However, to instrument large immersive environments for multi-person usage, it is necessary to demarcate personal vs. collaborative or shared sub-spaces within the context of the large screen. Contextualizing physical user locations in the space plays an important role.

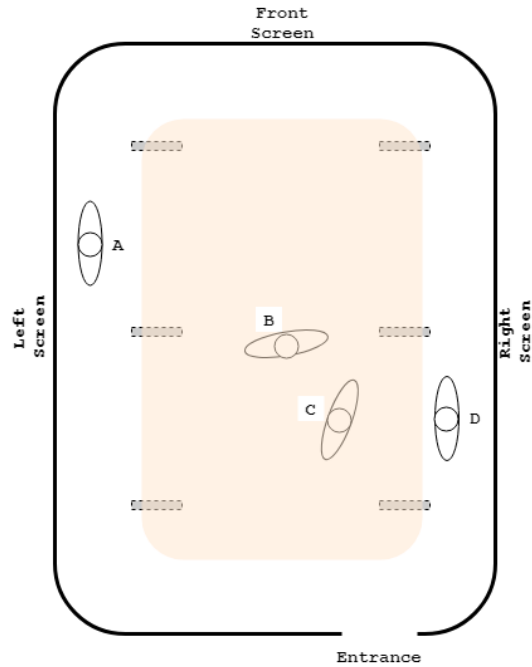


Fig. 2: Users A and D are able to interact with the screen whereas users beyond 2 meters distance to the screen (B and C) are in the inactive (light red) region and thus cannot interact with the screen.

To allow multiple users to interact with the screen at the same time, the large screen is subdivided into dynamic sub-spaces based on physical user locations. The existing ceiling-mounted Kinect tracking system returns the (x,y) location of

each user in a coordinate system aligned to the rectangular floor space. Although users are tracked wherever they are in the space, we enabled display interactions only for users that are located within 2 meters of the screen, as shown in Figure 2. In this way, the center of the room acts as an inactive zone, where users can look around and decide on their next steps instead of actively participating at all times.

In order to make this behavior clear to the users, we carefully calibrated the floor (x,y) positions to corresponding screen locations. A key element of our design is a continuous visual feedback mechanism shown at the bottom of the screen when a user is in range, appearing as animated circular rings, as shown in Figures 5b, 5c and 5d. This feedback serves a two-fold purpose. It makes users aware that their movements and physical locations are being automatically interpreted by the system, and it also allows them to adjust their movements to accomplish interactions with small sub-screens or columns on the large screen. Beyond the continuous feedback, we create discrete interaction spaces on the screen that change dynamically based on user locations. Thus, at a given point in time, users are able to visualize how the system continuously interprets their physical locations in real time, and also the column or sub-screen with which they are able to interact.

3.2 Input Modes of Interaction

Early Iterations 3-D sensing devices such as the Leap Motion have been previously instrumented to act as interaction devices, albeit in the context of 2-D large displays [34]. In previous work on supporting mid-air gestures, we experimented with the Leap Motion device as an input interaction method that leverages the underlying spatial context in the CRAIVE-Lab [35,36]. However, instrumenting the Leap Motion as a wrist-worn device to perform gestures such as grab, swipe, tap, scroll etc., led to user fatigue as it required both hands to be extended outwards for a considerable duration. Moreover, the device works best when it is secured and sits on a stable support. However, in our initial experiments, users freely moved around the space while wearing the device on one hand and performing interaction tasks with the other, which led to shakiness and poor results.

We also explored the feasibility of using voice commands as an exclusive means of interaction but quickly discarded this approach due to discouraging results from early testing and previous recommendations on usability [37,38,39]. Major challenges resulted from problems with the speech transcription technology due to, e.g., variation in accents. However, instead of getting rid of the voice control entirely, we selectively allowed the users to populate the content on the screen based on their voice commands. This was designed to enable users to generate visible content instead of limiting them only to manipulating existing elements on the screen. We hoped that instead of typing on a keyboard, speaking to the attached Bluetooth microphone would be more natural.

We also instrumented a smart watch as an input mechanism. However, translating screen gestures on a very small watch surface to the large screen that spans

a height of 5 meters was not very appealing. Beyond the factor of novelty, we concluded that it would not serve our aim of having an intuitive interaction mechanism.

Smartphones as Input Devices Ultimately, we developed our approach around users' personal smartphones, with which they are deeply comfortable and familiar. In addition, as discussed in earlier sections, prior research shows that instrumenting a touch enabled smart device as an interaction device for large screens has met with considerable success.

We developed a web application that can run on any touch screen device connected to the internet. Users were provided with QR codes or a short link, which allowed them to navigate to the web page. The web page was designed to run as a track pad, where regular touch screen gestures such as tap, swipe, scroll, double tap, pinch, drag and zoom were supported. Developing on a web platform removes the cumbersome process of having to download and install a standalone application.

3.3 System Architecture

The system architecture of the overall system is shown in Figure 3. It is primarily comprised of 3 components: (1) user input via smartphone and Bluetooth microphone, (2) spatial tracking via overhead Kinect sensors, and (3) the webpage running on the large immersive screen for visualization and output. All components communicate with each other in real time using the WebSocket protocol. The users' smartphone gestures are sent via WebSocket to the web application running on the large screen, as well as any voice input, which is passed through the Google speech-to-text transcription service. The user tracking system is located in a different node, which sends the (x,y) location of all users to the screen. The web application running on the large screen receives all the data, and displays dynamic feedback and visualizations accordingly.

3.4 Overall System

Combining all the components discussed in the previous sections, we designed a multi-user spatial interaction mechanism for the large immersive space, using smartphones as input interaction devices and voice control for content generation. As shown in Figure 5, two users can walk into the space and scan a QR code located near the entrance to launch the web application on their personal devices. The two QR codes correspond to the left and right sides of the large screen. As the users move towards their respective screens and come within the defined threshold of 2 meters, the location feedback and interaction mechanisms are activated, allowing them to interact with the individual columns as they see fit. In the experiments we describe next, we populate each of the columns with images from the public Flickr API. As users move around the space, the continuous spatial feedback appears at the bottom of the large screen and the

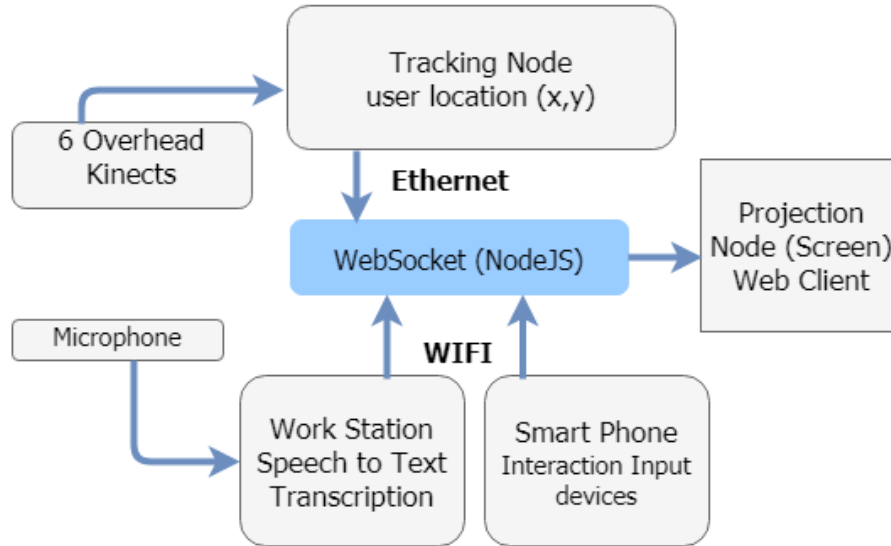


Fig. 3: Overall system architecture.

column with which each user can interact is highlighted in bright red. Since the interaction column is tied to the spatial location of the user, it can be viewed as exclusive or personal to the user standing in front of it.

A red cursor dot that appears on the spatially selected column can be moved using the web application on the phone and acts similar to a mouse pointer. Table 1 shows the list of supported gestures and how they translate to the big screen. An important aspect of our systems is that users need not look at their phone to manipulate the screen elements (in fact, the “touchpad” webpage is just a blank screen). The idea is for the users to keep their eyes on the screen while using a finger or thumb on the phone in a comfortable arm position. Images that users select on the left and right screens can be moved to the front screen, which supports a personal column for each user as well as a large shared usage area. Users can move their personally curated images to the shared area on the front screen. In our particular case, we designed an application in which users can simultaneously drag their personal images around the shared screen to design a simple newspaper-article-like layout.

4 User Studies

We gathered 14 participants to test the usability of and gain feedback about our overall system. Only 3 participants had extensive prior experience with working in immersive display environments. We designed two experiments. The first experiment was designed to gain a quantitative understanding of how long individual users take to perform various tasks on the screen using our system. The

Phone Gestures	Screen Result
Move	Move red pointer/drag image on shared screen
Tap	Select image
Swipe (Left or Right)	Move image/s to front screen
Swipe (Up or Down)	Scroll up/down personal column
Pinch	Shrink selected image
Zoom	Enlarge selected image
Double Tap	Enlarge/shrink selected image
Long Tap	Activate/deactivate drag on shared screen
Other input	Screen Result
Move (Physical user locations)	Select different column/continuous circular visualization
Voice input (“Show me pictures of X”)	Populate column with pictures of “X”

Table 1: List of phone gestures, physical user movements, and voice inputs, and their corresponding screen results.

second experiment was designed as a simple game, where two users simultaneously work using both their personal and shared screens to come up with a final layout. This was largely designed to understand how comfortable users felt in the space and how intuitive they felt the system to be. For this experiment, users were mostly left on their own to complete the tasks based on their understanding of how the system works.

4.1 Experiment 1

Individual users were directed to use only the left screen, where they were asked to complete tasks based on the prompts appearing on the large screen. There were 9 columns, each filled with random images, as illustrated in Figure 4. Screen prompts would appear randomly on any of the 9 columns asking the user to complete various tasks, one after the other. We tested all the gestures and inputs by asking the user to perform tasks shown in the second column of Table 1, except for the pinch, zoom, and double tap. Each task is completed once the user performs the correct input that corresponds to the displayed prompt. For instance, if a user at a given point in time is in front of the 2nd column, a prompt might appear in the 9th column indicating “*Select a picture from this column and move it to the front screen*”. Then, the user would physically move until the system highlights the 9th column, and perform the corresponding scroll, tap, and swipe gestures. This would successfully complete the task and another prompt would appear on a different column, such as “*Populate this column with pictures of dogs*”, which would require a voice command. We recorded the time it took for



Fig. 4: Experiment 1 setup. The red dot (appearing next to the silhouette of the user) acts as a cursor to interact with the spatially selected column. The yellow text box, shown in the column that the user is interacting with, directs user tasks, in this case “Move an image from this column.”

the user to accomplish each task, including both the time it took to make spatial selections by moving between the columns and the time it took to successfully perform phone or voice input.

4.2 Experiment 2

We designed this experiment to be completed in pairs. Both users had completed Experiment 1 before taking part in this experiment. Our aim was to make sure users understand all input mechanisms and are comfortable to freely use the system.

On the front screen, where the shared screen is located, we presented a simple layout with two short paragraphs of text and image placeholders. Each paragraph consisted of a heading indicating a recipe name and text below describing the ingredients and preparation. Each user was responsible for finding an appropriate image for “their” recipe. Initially, the users independently move along the left and right sides of the screen, selecting one or more images and moving them to their personal columns on the front screen. Then, they move to the front screen and select the most likely candidate image from the refined set of images and move it to the shared screen with the recipe. A screen prompt on the large screen notifies the user whether a correct image was selected (i.e., a picture of the dominant ingredient in the recipe, such as an avocado picture for a guacamole recipe).

Once the correct image is moved to the shared screen, users can perform a long-tap gesture on their phone to activate dragging on the shared screen. This

allows the users to simultaneously drag their answer images to an appropriate location, which is generally next to the corresponding text. A screen prompt notifies the user once the target image has been moved to the required location on the shared screen. When both users complete their tasks on the shared screen, the full task is complete. The steps of the process are illustrated in Figure 5. Each

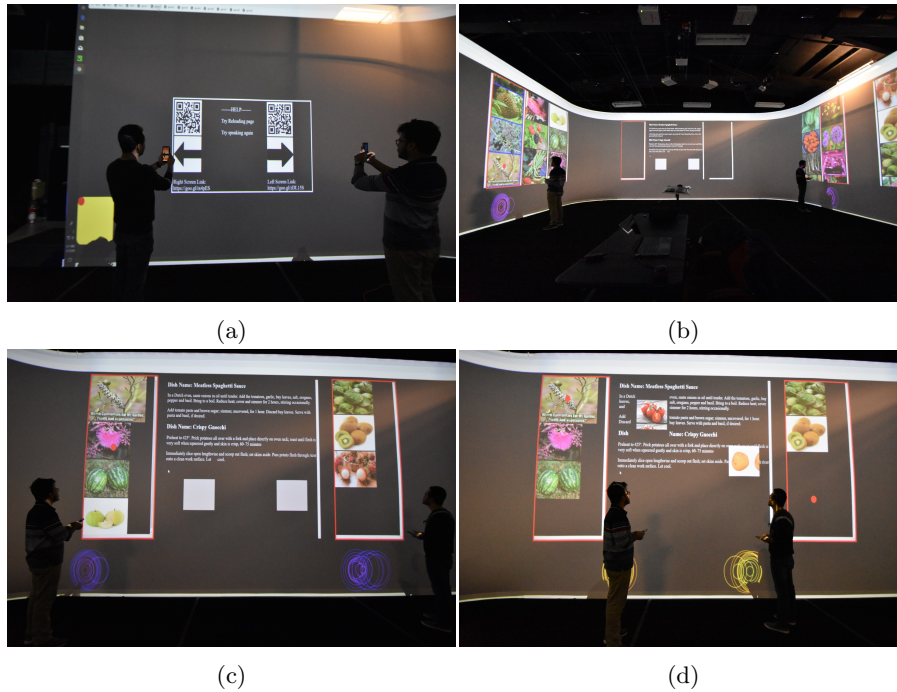


Fig. 5: Multiple users during Experiment 2. (a) Each user scans a QR code. (b) Users work in their personal spaces using their smartphones and/or voice control. (c) Users move to the front screen to view their curated list of images. (d) Both users manipulate the shared space to complete a full task. For better visualization, please refer to the video at https://youtu.be/CYqXKyHTO_U.

user pair was presented with six sets of recipe “games”. Three of the recipe pairs had the correct images already placed in one of the pre-populated columns and the users had to move around, scroll the columns, and locate the correct image. The other three pairs did not have the answer images in any of the columns and this required the users to generate content on their own by verbally requesting the system to populate a blank column with images of what they thought was the main ingredient in the recipe, one of which the user had to select and move to the front screen to verify.

We designed this setup to study whether users felt comfortable completing tasks based on the interaction mechanisms we designed for our display envi-

ronment. We also wanted to find out if the users, most of whom had no prior experience with these kinds of spaces, found interacting with an unconventional immersive space such as this one to be fun and intuitive. Therefore, we asked the users to fill out a NASA-TLX questionnaire along with an additional questionnaire based on a 5 point Likert scale to obtain feedback on specific spatial, gestural, and voice input mechanisms that we designed.

5 Results

On average, each of the 14 participants performed 27 tasks during Experiment 1, where each of the 5 tasks appeared at random. Users were required to perform at least 20 and at most 35 tasks depending on the randomness of the distributed tasks as well as their speed at completing them. All tasks were assigned equal probability of appearing, except for voice control tasks, which appeared less often, according to the design considerations discussed earlier. The average number of tasks per user was distributed as follows: spatial selection (7.35), scrolling image columns (4.93), selecting an image (6.43), moving images to the center screen (6.65), and populating with voice (2.36). Timing results are reported in Figure 6.

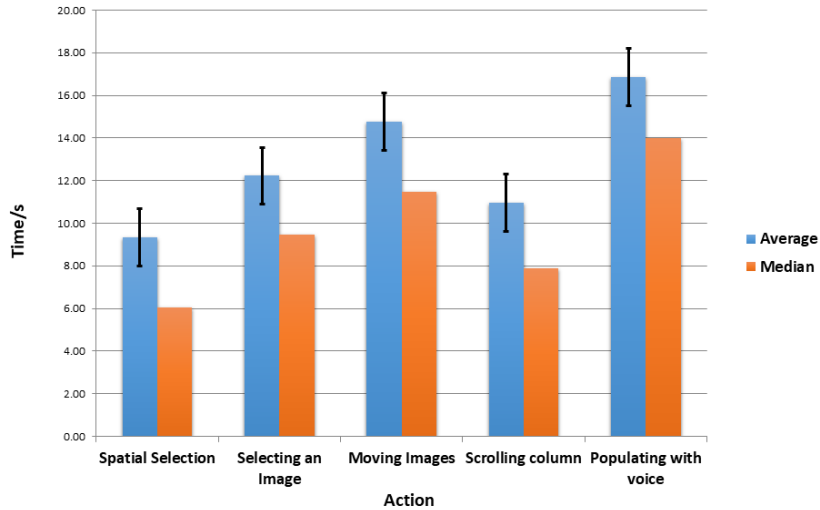


Fig. 6: Average and median times for users to complete each action in Experiment 1. Error bars represent standard error (σ/\sqrt{N}). All actions take longer duration than spatial selection as their completion requires spatial selection as a prerequisite.

Even though we report both the average and median time for each of the actions, we believe that the median times for each of the tasks are more reflective

of typical user performance. We observed many cases in which a certain user would take a lot of time to internalize one particular action, while completing other similar actions quickly. This varied significantly from one user to another and therefore led to some higher average values than expected. Unsurprisingly, voice input was the most time consuming action as can be seen in Figure 6.

For Experiment 2, where multiple users worked simultaneously on their personal screens and came together on the shared screen space to complete the full task, we recorded the time of completion. Since there were 3 games for the touch-only interface and 3 for the voice interface, each pair of participants played 6 games. Out of the 21 games for each type of input (7 participant pairs \times 3 games per input type), participants completed 17 of each. 4 games for each input were not completed for various reasons, typically a system crash or one of the participants taking too long to figure out the answer and giving up. The average time taken for a pair of participants to complete the touch-only based game and voice-based game were 2.31 minutes and 1.67 minutes respectively. Even though experiment 1 revealed that voice input generally takes longer, we note that for touch input the user has to physically move and search for the correct image among a wide array of choices, while for the voice input, users can quickly generate for pictures of their guessed ingredient and move one to the shared screen area.

We asked participants to fill out a NASA-TLX questionnaire after completing both experiments to investigate how comfortable and usable our overall system is, and present the results in Figure 7. We added an extra question regarding the intuitiveness of the overall system, where on the 21 point scale, a higher number indicates a higher degree of intuitiveness. Overall, participants rated their mental, physical, and temporal demand, along with effort and frustration in using the system, to be low. Performance and intuitiveness were highly rated. In addition, users filled out another questionnaire related to how well they liked/disliked particular interaction mechanisms such as phone gestures, spatial interactions, voice input, and so on, using a 5 point Likert scale. As shown in Figure 8, median values for most of these components are rated very highly. The ratings were also high for whether the overall tasks were fun and enjoyable. Users also highly rated the user interface and other feedback on the screen, including the constant localization feedback. Among the 14 participants, 3 were previously familiar with the physical space. However, the interaction interface was completely new to them, the same as the rest of the users. We observed that the users familiar with the large immersive space performed 25% and 30% faster than the overall average for the touch and voice games respectively.

We observed that in Experiment 2, the average time for completion with voice input was less than that for the smart phone input, even though Experiment 1 revealed that voice input takes a longer time on average. This can be explained due to the time-consuming nature of search required in the phone input subtask. On the other hand, for the voice input task, upon knowing the key ingredient, users were quickly able to ask for valid pictures and move them to the shared screen area.

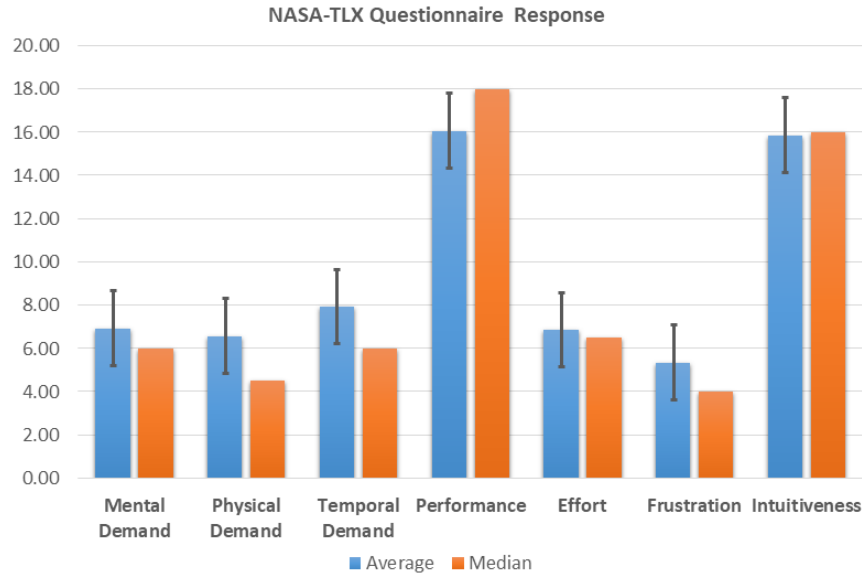


Fig. 7: Average and median values of user responses to the NASA-TLX questionnaire rated on a 21 point scale. Error bars represent standard error.

6 Discussion

Based on our observations and post-task interviews, many users appreciated the constant spatial feedback, allowing them to understand their impact on the space. Some users appreciated the automatic demarcation of personal vs. shared within the scope of the same large screen. We observed many issues related to the automatic speech understanding and transcription. Non-native English speakers had more difficulty populating their columns with desired input. Thus it was unsurprising that the average time for actions to be completed using voice was the largest, as shown in Figure 6. Users were divided on the usefulness and comfort of voice input; one user wished he could carry out the entire task using his voice while another was completely opposed to using voice as any kind of input mechanism.

Many participants gave high marks to the system’s approach of mapping the horizontal location of their screen cursor to their physical location and the vertical location to their smartphone screen. However, technical difficulties in which some users had to repeatedly refresh the webpage on their phone due to lost WebSocket connections contributed to a certain level of annoyance.

One of the major usage challenges that many participants commented on was the appropriate appearance of screen prompts and other feedback at eye height. Designing user interfaces/feedback for large displays without blocking screen content is a continuing challenge for this type of research.

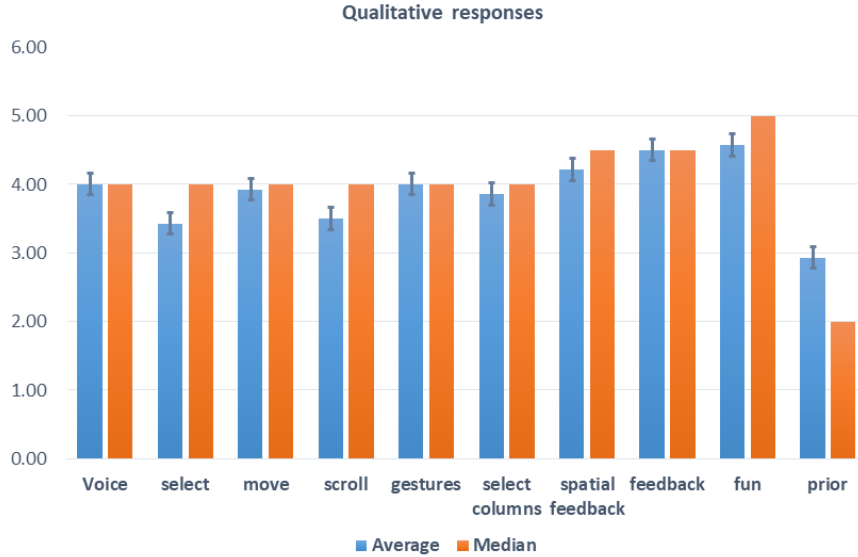


Fig. 8: Average and median values of user responses to the second questionnaire on a 5 point Likert scale. Error bars represent standard error.

7 Conclusion and Future Work

In terms of overall performance, the results were very encouraging with regard to the usefulness of the overall interaction interface. Using standalone methods such as cross-device platforms or voice only methods have shown limited usability in the past [39]. However, in our case, we see that users adapt well to multi-modal inputs; touch screen and voice are used in conjunction with automatic interaction mechanisms based on spatial tracking.

Large-display-rich immersive spaces such as the one presented in this work draw significant amount of user attention. When designing interaction interfaces that do not overwhelm users, it is important to devise methods that require minimal attention. We found that allowing users to work with their own familiar smartphones was a successful approach. Furthermore, allowing users to move freely and using the voice commands selectively, only for content generation, helped users to continuously focus on the screen and the task at hand instead of having to repeatedly glance at the phone screen or manually type input commands.

The multi-modal user interface presented in this work and its success led us to work towards building new use cases for our large-display environments. We are working towards building a language learning classroom use case, where students match language characters to images. The image selection and placement tasks discussed here can be re-purposed to support classroom activities, where the

room in itself is a teaching tool in contrast to conventional classrooms [40]. Assessed learning outcomes based on student feedback and the overall success of our interface will be important in furthering interaction design choices going forward.

While we only reported 2-user studies here, our immediate next step is to accommodate 3–6 users simultaneously, to fully realize the potential of our immersive environment as a multi-user space. In addition to direct extensions of the experiments we discussed here, we are investigating how the screen space for each user can be dynamically defined based on their location rather than constrained to one side of the screen.

We are also working to replace the worn Bluetooth microphones with an ambient microphone array that uses beamforming, along with the users’ known locations, to extract utterances for verbal input. Finally, we hope to conduct more systematic eye-tracking experiments to explore where the users look on the big screen and how often/under what circumstances they glance down at their phone “touchpad”.

8 Acknowledgements

The research in this thesis was partially supported by the US National Science Foundation under awards CNS-1229391 and IIP-1631674, and by Cognitive and Immersive Systems Laboratory, a research collaboration between Rensselaer Polytechnic Institute and IBM through IBM’s AI Horizon Network. Thanks to Devavrat Jivani for his contribution in setting up the multi-tracking system described in this work.

References

1. Vogel, D., Balakrishnan, R.: Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In: Proceedings of the 17th annual ACM symposium on User interface software and technology, pp. 137–146. ACM (2004)
2. Malik, S., Ranjan, A., Balakrishnan, R.: Interacting with large displays from a distance with vision-tracked multi-finger gestural input. In: Proceedings of the 18th annual ACM symposium on User interface software and technology, pp. 43–52. ACM (2005)
3. Kister, U., Reipschläger, P., Dachsel, R.: Multilens: Fluent interaction with multi-functional multi-touch lenses for information visualization. In: Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces, pp. 139–148. ACM (2016)
4. Bragdon, A., DeLine, R., Hinckley, K., Morris, M.R.: Code space: touch+ air gesture hybrid interactions for supporting developer meetings. In: Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces, pp. 212–221. ACM (2011)
5. Langner, R., Kister, U., Dachsel, R.: Multiple coordinated views at large displays for multiple users: Empirical findings on user behavior, movements, and distances. *IEEE transactions on visualization and computer graphics* **25**(1), 608–618 (2019)

6. Liu, C.: Leveraging physical human actions in large interaction spaces. In: Proceedings of the adjunct publication of the 27th annual ACM symposium on User interface software and technology, pp. 9–12. ACM (2014)
7. Wolf, K., Abdelrahman, Y., Kubitzka, T., Schmidt, A.: Proxemic zones of exhibits and their manipulation using floor projection. In: Proceedings of the 5th ACM International Symposium on Pervasive Displays, pp. 33–37. ACM (2016)
8. Kister, U., Klamka, K., Tominski, C., Dachselt, R.: Grasp: Combining spatially-aware mobile devices and a display wall for graph visualization and interaction. In: Computer Graphics Forum, vol. 36, pp. 503–514. Wiley Online Library (2017)
9. Cruz-Neira, C., Leigh, J., Papka, M., Barnes, C., M. Cohen, S., Das, S., et al.: Scientist in wonderland: A report on visualization applications in the cave virtual reality environment. pp. 59–66. IEEE Symposium on Research Frontiers in Virtual Reality (1993)
10. Cruz-Neira, C., J. Sandin, D., A. Defanti, T., V. Kenyon, R., C. Hart, J.: The cave: Audio visual experience automatic virtual environment. *Communication of the ACM* **35**, 64–72 (1992)
11. Febretti, A., Nishimoto, A., Thigpen, T., Talandis, J., Long, L., Pirtle, J., Peterka, T., Verlo, A., Brown, M., Plepys, D., et al.: Cave2: a hybrid reality environment for immersive simulation and information analysis. In: IS&T/SPIE Electronic Imaging, pp. 864,903–864,903. International Society for Optics and Photonics (2013)
12. Rittenbruch, M., Sorensen, A., Donovan, J., Polson, D., Docherty, M., Jones, J.: The cube: a very large-scale interactive engagement space. In: Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces, pp. 1–10. ACM (2013)
13. Sharma, G., Braasch, J., Radke, R.J.: Interactions in a human-scale immersive environment: the CRAIVE-Lab. In: Cross-Surface Workshop at ISS2016. Niagara Falls, Canada (2016)
14. Ackad, C., Clayphan, A., Tomitsch, M., Kay, J.: An in-the-wild study of learning mid-air gestures to browse hierarchical information at a large interactive public display. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 1227–1238. ACM (2015)
15. Yoo, S., Parker, C., Kay, J., Tomitsch, M.: To dwell or not to dwell: an evaluation of mid-air gestures for large information displays. In: Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction, pp. 187–191. ACM (2015)
16. Ackad, C., Tomitsch, M., Kay, J.: Skeletons and silhouettes: Comparing user representations at a gesture-based large display. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 2343–2347. ACM (2016)
17. Ballendat, T., Marquardt, N., Greenberg, S.: Proxemic interaction: designing for a proximity and orientation-aware environment. In: ACM International Conference on Interactive Tabletops and Surfaces, pp. 121–130. ACM (2010)
18. Kopper, R., Silva, M.G., McMahan, R.P., Bowman, D.A.: Increasing the precision of distant pointing for large high-resolution displays (2008)
19. Brudy, F., Ledo, D., Greenberg, S., Butz, A.: Is anyone looking? mitigating shoulder surfing on public displays through awareness and protection. In: Proceedings of The International Symposium on Pervasive Displays, p. 1. ACM (2014)
20. Hawkey, K., Kellar, M., Reilly, D., Whalen, T., Inkpen, K.M.: The proximity factor: impact of distance on co-located collaboration. In: Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work, pp. 31–40. ACM (2005)

21. Wallace, J.R., Weingarten, A., Lank, E.: Subtle and personal workspace requirements for visual search tasks on public displays. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 6760–6764. ACM (2017)
22. Anslow, C., Campos, P., Jorge, J.: Collaboration Meets Interactive Spaces. Springer (2016)
23. von Zadow, U., Dachselt, R.: Giant: Visualizing group interaction at large wall displays. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 2639–2647. ACM (2017)
24. Kropp, M., Anslow, C., Mateescu, M., Burkhard, R., Vischi, D., Zahn, C.: Enhancing agile team collaboration through the use of large digital multi-touch card-walls. In: International Conference on Agile Software Development, pp. 119–134. Springer, Cham (2017)
25. Horak, T., Kister, U., Dachselt, R.: Presenting business data: Challenges during board meetings in multi-display environments. In: Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces, pp. 319–324. ACM (2016)
26. Lander, C., Speicher, M., Paradowski, D., Coenen, N., Biewer, S., Krüger, A.: Collaborative newspaper: Exploring an adaptive scrolling algorithm in a multi-user reading scenario. In: Proceedings of the 4th International Symposium on Pervasive Displays, pp. 163–169. ACM (2015)
27. Luoju, P., Koskela, J., Ollila, K., Mäki, S.M., Kulpa-Bogossia, R., Heikkinen, T., Ojala, T.: Wordster: collaborative versus competitive gaming using interactive public displays and mobile phones. In: Proceedings of the 2nd ACM International Symposium on Pervasive Displays, pp. 109–114. ACM (2013)
28. Doshi, V., Tuteja, S., Bharadwaj, K., Tantillo, D., Marrinan, T., Patton, J., Marai, G.E.: Stickyschedule: an interactive multi-user application for conference scheduling on large-scale shared displays. In: Proceedings of the 6th ACM International Symposium on Pervasive Displays, p. 2. ACM (2017)
29. Ballagas, R., Borchers, J., Rohs, M., Sheridan, J.G.: The smart phone: a ubiquitous input device. *IEEE Pervasive Computing* (1), 70–77 (2006)
30. Baldauf, M., Adegeye, F., Alt, F., Harms, J.: Your browser is the controller: advanced web-based smartphone remote controls for public screens. In: Proceedings of the 5th ACM International Symposium on Pervasive Displays, pp. 175–181. ACM (2016)
31. Chen, X., et al.: Air+ touch: interweaving touch and in-air gestures. In: Proceedings of the 27th annual ACM symposium on User interface software and technology. ACM (2014)
32. Du, Y., Ren, H., Pan, G., Li, S.: Tilt & touch: Mobile phone for 3d interaction. In: Proceedings of the 13th international conference on Ubiquitous computing, pp. 485–486. ACM (2011)
33. Shirazi, A.S., Winkler, C., Schmidt, A.: Flashlight interaction: a study on mobile phone interaction techniques with large displays. In: Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services, p. 93. ACM (2009)
34. Dingler, T., Funk, M., Alt, F.: Interaction proxemics: Combining physical spaces for seamless gesture interaction. In: Proceedings of the 4th International Symposium on Pervasive Displays, pp. 107–114. ACM (2015)
35. Sharma, G., Jivani, D., Radke, R.J.: Manipulating screen elements in an immersive environment with a wrist-mounted device and free body movement. In: Living Labs Workshop, CHI 2018. Montreal, Canada (2018)

36. Jivani, D., Sharma, G., Radke, R.J.: Occupant location and gesture estimation in large-scale immersive spaces. In: Living Labs Workshop, CHI 2018. Montreal, Canada (2018)
37. Nutsi, A., Koch, M.: Multi-user usability guidelines for interactive wall display applications. In: Proceedings of the 4th International Symposium on Pervasive Displays, pp. 233–234. ACM (2015)
38. Nutsi, A.: Usability guidelines for co-located multi-user interaction on wall displays. In: Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces, pp. 433–438. ACM (2015)
39. Sarabadani Tafreshi, A.E., Soro, A., Tröster, G.: Automatic, gestural, voice, positional, or cross-device interaction? comparing interaction methods to indicate topics of interest to public displays. *Frontiers in ICT* **5**, 20 (2018)
40. Divekar, R.R., Zhou, Y., Allen, D., Drozdal, J., Su, H.: Building human-scale intelligent immersive spaces for foreign language learning. *iLRN 2018 Montana* p. 94 (2018)