Multimodal Situated Analytics (MuSA) for Analyzing Conversations in Extended Reality

by

Ashwini Ganapati Naik
M.S. in Computer Science, University of Illinois Chicago, 2011
B.E. in Information Science, Visvesvaraya Technological University, 2007

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, 2024

Chicago, Illinois

Defense Committee:
Andrew Johnson, Chair and Advisor
Robert Kenyon
Debaleena Chattopadhyay
Nikita Soni
Steve Jones, Communication

To my late father whose resolute faith in the transformative power of education to brighten the world has continually inspired me.

ACKNOWLEDGMENT

I am deeply grateful to my parents and brother for their unwavering faith in me and motivation to always aim higher and strive for excellence. Your love, wisdom, and support have been foundational to my accomplishments. I am equally thankful to my parents-in-law and sister-in-law for their endless encouragement and support of my aspirations and dreams. I am grateful to our son Agastya, whose presence adds unparalleled joy and beauty to every moment, enriching our lives beyond measure.

I'm profoundly thankful to my advisor, Dr. Andrew Johnson, for lighting the path that led me through this academic journey. You have inspired me and have been a critical element in my decision to pursue this path. I deeply appreciate your boundless patience, guidance, and support. I am also deeply grateful to my co-advisor, Dr. Robert Kenyon, for instilling in me the genuine essence of research, providing a solid foundation for this endeavor, and fostering a spirit of resilience. My heartfelt gratitude goes to my committee, whose critical insights have significantly enhanced the quality of my work and made me a better researcher.

To Lance, I want to express my gratitude for the countless times you've assisted me with technical challenges in the lab. Thank you so much for your continued encouragement, invaluable guidance, and unwavering support. Your positive outlook and influence have made a great impact on my journey. Special thanks to Luc for his patience, flexibility, guidance, and support. I would also like to thank Dana for her constant support, encouragement, and care. I feel truly fortunate to be surrounded by such incredible people, and the thought of moving on is almost

daunting. Thank you all! Your contributions have left an indelible mark on my journey; I am forever grateful for that.

I am also grateful to all my friends at EVL, whose support has been a constant source of strength throughout these years. The extended family at the Electronic Visualization Laboratory has been a bedrock of support and guidance, for which I am immensely grateful.

Finally, to my beloved husband, Abhishek, thank you for making this milestone possible for us. Your belief in my success, unconditional love, patience, and unwavering support have been my greatest strengths. Thank you for being by my side on this incredible journey. I couldn't imagine doing it without you!

AGN

CONTRIBUTIONS OF AUTHORS

This dissertation incorporates portions of previously published work by myself and other authors. All portions contained within in this dissertation represent my direct contributions to these publications. The prior publications included in this work are the following:

In chapters 1, 2, 3, and 6 - A. Naik and A. Johnson, "PSA: A Cross-Platform Framework for Situated Analytics in MR and VR," in 2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Sydney, Australia, 2023 pp. 92-96.

In chapters 1, 2, 3, and 6 - In chapters 1, 2, 3, and 6 - A Naik and A Johnson. 2023. Using Personal Situated Analytics (PSA) to Interpret Recorded Meetings. In Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23 Adjunct). Association for Computing Machinery, New York, NY, USA, Article 42, 1–3. https://doi.org/10.1145/3586182.3616697

Room and furniture models used for Phase 1 User Study in Chapter 3 - Nishimoto, A. (2019). VirtualUIC-EVL (Version 1.0) [Computer software].
https://bitbucket.org/arthurnishimoto/virtualuic-evl

Kreimer, B (2022). Allensworth-Building-Models (Version 1.0) [Data set].
https://sketchfab.com/benkreimer/collections/allensworth-building-models

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

CHAPTER **PAGE**

TABLE OF CONTENTS (Continued)

CHAPTER **PAGE**

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Continued)

FIGURE **PAGE**

LIST OF FIGURES (Continued)

FIGURE **PAGE**

LIST OF FIGURES (Continued)

FIGURE **PAGE**

LIST OF FIGURES (Continued)

FIGURE **PAGE**

LIST OF FIGURES (Continued)

FIGURE **PAGE**

LIST OF ABBREVIATIONS

AR Augmented Reality

CA Conversation Analysis

HCI Human Computer Interaction

MR Mixed Reality

MuSA Multimodal Situated Analytics

UIC University of Illinois Chicago

VR Virtual Reality

XR Extended Reality

SUMMARY

Analyzing conversations and interactions that involve multiple participants spans several disciplines, including linguistics, communication studies, and human-computer interaction (HCI). Among the methodologies employed in these fields, Multimodal Analysis stands out for its comprehensive approach. This method examines a broad spectrum of semiotic codes—ranging from verbal language and non-verbal cues to visual elements, gestures, facial expressions, and even the spatial arrangement of participants. Given the multimodal nature of modern communications, which now integrate text, video, images, voice, AI agents, and immersive technologies, there is a growing need to understand these complex interactions deeply.

Recognizing this need, we introduce Multimodal Situated Analytics (Multimodal Situated Analytics (MuSA)), a framework designed to enable individuals to immerse themselves in recorded conversations across various levels of the Reality-Virtuality spectrum. This immersion allows users to analyze the dynamics of conversations and their own exploration patterns within these dialogues. MuSA merges the concepts of embodied cognition, situated analytics, and multimodal analysis to create a unique environment for studying conversations as they occurred, using immersive technologies to provide a context-rich exploration experience.

Our development process for MuSA encompassed several critical stages: from tracking and capturing data to cleaning, synchronizing, and finally building a prototype for deployment on end-user hardware. Initially, we focused on conversations among seated participants to refine our approach. Following this, a pilot user study involving 12 participants was conducted

to explore the efficacy of using Mixed Reality (MR) and Virtual Reality (VR) technologies, specifically through the HoloLens2 and Quest2 devices, for analyzing recorded conversations.

This initial exploration was further extended through a second user study with 13 participants, focusing on non-seated, moving individuals to understand the usability, adoption, and spatial interaction within multimodal conversations. Feedback from both user studies, complemented by expert evaluations in linguistics and communication, provided invaluable insights into the strengths and challenges of the MuSA framework.

These insights inform our development pipeline and outline a path for future research. The lessons learned from these exploratory studies and the feedback received are instrumental in refining our approach to multimodal conversation analysis. As we evolve the MuSA framework, we aim to enhance the capabilities and applications of embodied situated analytics, thereby contributing to the broader field of multimodal analysis.

CHAPTER 1

BACKGROUND

Parts of this chapter have been published in the proceedings of ISMAR 2023 [90] and UIST 2023 [91].

Multimodal Analysis involves an intricate pipeline and encompasses the examination of multiple communication modes, including verbal language, gestures, body language, visuals, and non-verbal cues. This approach addresses the multifaceted nature of human communication by capturing a wide spectrum of interactive elements. However, executing such comprehensive analyses presents several challenges. Key issues include the complexity inherent in multi-layered interactions, the difficulty of generalizing findings across different settings, challenges in integrating diverse data types, subjective interpretation of data, and the technological demands of synthesizing this information [59, 60].

Research in multimodal analysis has revealed significant advantages of this method for analyzing communication among individuals and groups. This approach integrates various communication modes, enhancing the understanding of how people interact both verbally and non-verbally. A multimodal approach offers a comprehensive perspective, taking into account not just what is spoken but also the broader context and more subtle, often non-verbal forms of communication. By integrating various modes—such as gestures, facial expressions, and spatial positioning—it allows for a deeper understanding of how people communicate and interact. This holistic view is crucial because it recognizes that effective communication is influenced

by a complex mix of verbal and non-verbal cues, all of which contribute to the meaning and dynamics of a conversation. This method enhances contextual awareness and provides deeper insights into how individuals engage with and respond to various communicative cues. Such analysis reveals the complex interplay of different modalities that are essential to understanding nuanced human interactions [33].

Moreover, insights derived from multimodal analysis can elucidate the dynamics of interaction at a granular level. By examining how different modes of communication intersect and influence one another, researchers can gain a more comprehensive understanding of communication processes. This is particularly valuable in fields like Human Computer Interaction Human Computer Interaction (HCI), where understanding the subtleties of user interaction can inform more effective design and implementation of interactive systems [58].

In recent years, substantial research has focused on using immersive environments to develop user-friendly applications for data visualization and analysis. This field aims to make complex data more accessible and interactive through immersive technology. However, this area of study often overlooks the potential applications of embodied cognition and situated analytics. These concepts explore how spatial context can enhance our understanding and navigation of environments, yet they remain underutilized in current research.

One key area where these ideas are being integrated is in the analysis of conversations. Recent studies have begun to acknowledge the embodied nature of communication, how people physically gather and interact, and the influence of their activities' ecological, material, and spatial contexts. For instance, Mondada [59] highlights these aspects in her examination of

conversation analysis challenges, emphasizing the physical and situational elements of communication.

Situated Analytics, as an emerging field within Human Computer Interaction (HCI), offers a novel approach to data analysis by linking data representations directly to relevant objects, places, and persons. This method is gaining traction across various disciplines, including visualization, Human Computer Interaction (HCI), and augmented reality, due to its potential for enhancing sensemaking and decision-making processes. Thomas (2018) discusses the foundational concepts and applications of Situated Analytics in detail, illustrating its relevance and applicability in multiple contexts [89].

Our current research focuses on applying embodied situated analytics specifically to the analysis of recorded conversations. By creating immersive environments that replicate the original setting of a conversation, we aim to improve strategic planning and sensemaking. We utilize virtual avatars and dynamic elements such as conversation snippets appearing as rising bubbles alongside these avatars, enhancing the multimodal analysis experience.

However, the application of situated analytics has predominantly been within Augmented/Mixed Reality (Augmented Reality (AR)/Mixed Reality (MR)) settings, which naturally integrate spatial context by overlaying digital information onto the real world. Virtual Reality Virtual Reality (VR), in contrast, often lacks this integration due to the occlusive nature of VR headsets, which block out the physical environment. To address this, our research uses detailed 3D models of the original conversation settings within Virtual Reality (VR) to simulate and maintain spatial context. We then compare user experiences across both Mixed Reality

(MR) and Virtual Reality (VR) platforms to assess the effectiveness of these different immersive approaches.

1.1 Hypothesis

We hypothesize that incorporating the elements of spatial context and immersiveness through the use of situated analytics and embodied cognition can assist in the analysis of multimodal data in conversations. We gather the location data of each user with the help of OptiTrack tracking systems in a large collaborative environment. The dataset in our study includes the conversation data, tracking data of the participants of the conversation, and the environment of the conversation along with supporting visuals that may aid analysis. Hence the data visualized are both physically and temporally situated with respect to most of the factors influencing the conversation. A visualization is physically situated in space if its physical presentation is physically close to the data's physical referent. A visualization is temporally situated if the data's temporal referent is close to the moment in time the physical presentation is observed[89]. Research has shown that the multimodality of the data in conversations can be further expanded by considering not only embodied resources for interacting but also embodied practices for sensing the world in an intersubjective way [1]. The increasing attention given to conversational aspects in human-computer interaction and artificial intelligence reflects a trend toward emphasizing the social dimension, particularly communicative interaction and mutual understanding between people. Gaze, posture, body movement, spatial distance as well and the arrangement of participants and objects in space are important semiotic codes in conversation and influence how we organize and make sense of our activities [39]. The data used in the experiment incor-

porates all these semiotic codes thereby potentially aiding the analysis of the conversation in question.

1.1.1 Exploratory Questions

Our initial research questions were more exploratory in nature and were intended to understand the capabilities of such a system to conduct multimodal analysis.

1. RQ1: Could we use MuSA to analyze Multimodal meetings in Extended Reality (XR)?
2. RQ2: Do users prefer Mixed Reality (MR) environments over Virtual Reality (VR) environments or vice versa for in-situ multimodal analysis?

1.2 Contributions

Our contributions can be summarized as follows:

1. Design space for Multimodal Situated Analytics pipeline and immersive application: We establish a workflow that acts as a sample design space for embedding individuals in recorded conversations in immersive environments (Figure 9). The workflow consists of multiple steps such as tracking, capturing, data cleaning, synchronization, prototype development and deployment to commercially available hardware. We then developed immersive VR and MR applications for exploring and analyzing conversations. Figure 1 depicts an analyst exploring a seated conversation within MuSA.
2. Empirical Results from User Study 1(n=12) for seated participants: We conducted a pilot study with 12 users to explore their preferences for exploration and analysis in terms of choice of interactions, attributes, viewpoints, and space usage. The study also explored

the changes in users behavior when getting close to objects of interest while navigating in the environment. This assessment primarily served as a formative stage, guiding us to refine our boundaries and directing us toward more focused research (Figure 2 (a)(b)).

3. Empirical results from User Study 2(n=13) for standing and moving participants: We conducted a second user evaluation after modifying our application to accommodate standing and moving participants in MR (Figure 2(c)(d)).

4. Lessons Learned: Drawing from our experience in the user evaluations and the feedback received from the participants, we present a set of lessons learned that could be useful for the research community and potentially enhance the multimodal analysis pipeline. These lessons learned include insights into the design of immersive environments, user engagement, and the use of immersive technologies for multimodal analysis.

Our research has contributed to systematically analyzing multimodal meetings within immersive environments. By shedding light on the advantages and drawbacks of employing immersive technologies for multimodal analysis, it paves the way for further investigation and innovation in this field. For the first user evaluation, we used a dataset that involved two seated individuals using a visual conversational AI agent to analyze COVID-19 data. In the second evaluation, we used a dataset with non-seated moving individuals exploring augmented reality models of historical buildings through their phones. We will refer to the individuals in the datasets as **"participants"** and the individuals analyzing the data in user evaluations as **"analysts"** in the rest of the paper. Figure 3 shows the difference between analysts and participants.

Figure 1: (a) Depiction of an analyst embedded in a conversation using Personal Situated Analytics in Virtual Reality (VR) in Phase 1. It shows two seated participants represented as virtual avatars engaged in a conversation with a visual AI agent on the display.

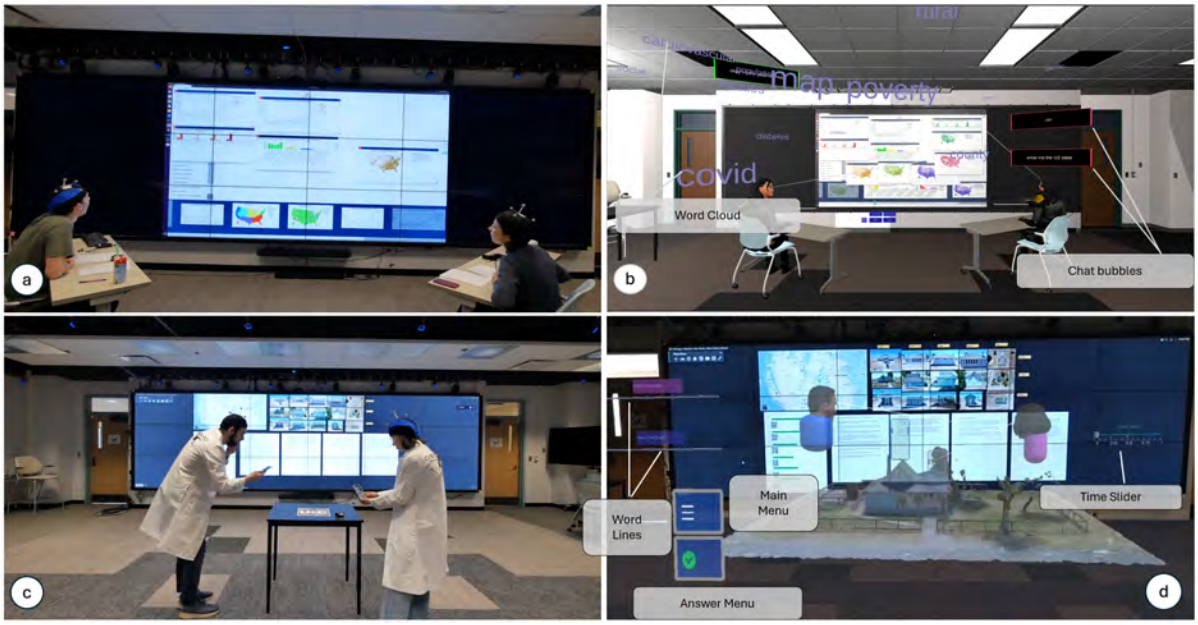


Figure 2: (a)Phase I (Data Collection) - Two seated participants exploring talking with a visual conversational AI agent on the display wall. (b) Phase I (User Study) Replaying (a) in Virtual Reality through MuSA in Quest2. (c) Phase II (Data Collection) - Two non-seated participants exploring AR building assets (d) Phase II (User Study) Replaying (c) in Mixed Reality through MuSA in HoloLens2.

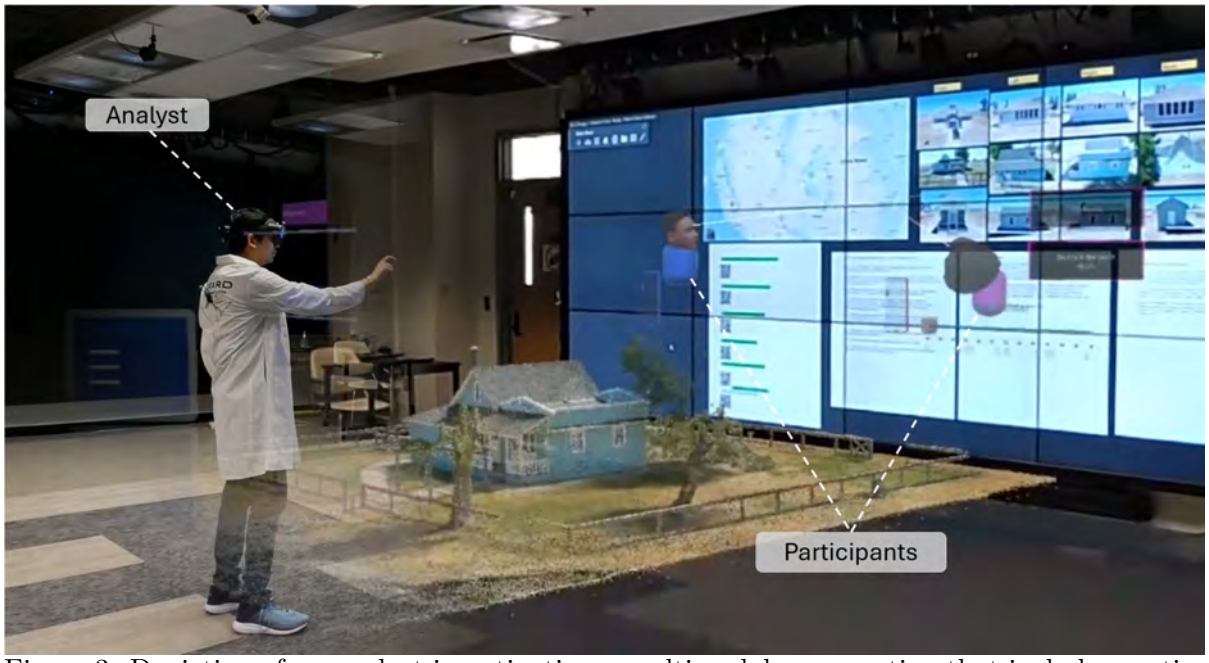


Figure 3: Depiction of an analyst investigating a multimodal conversation that includes participants and other entities of the conversation through MuSA in Mixed Reality (HoloLens2)

CHAPTER 2

RELATED WORK

Parts of this chapter have been published in the proceedings of ISMAR 2023 [90] and UIST 2023 [91].

In this chapter, we explore several research areas that intersect with our work, beginning with the concept of Immersive Analytics. This field encompasses various subdomains including Human-Computer Interaction (HCI), Augmented Reality (AR), Virtual Reality (VR), and the integration of Visual Analytics within immersive environments. We discuss how Immersive Analytics facilitates strategic immersion for efficient data exploration and how these environments are leveraged for effective data sensemaking.

We then examine the role of embodiment in visual analytics, highlighting its application across diverse fields such as education, performing arts, and visual analytics research. Following this, we delve into Proxemics to analyze how spatial arrangements between individuals and groups during interactions can enhance Multimodal Analysis.

Finally, we review the tools available in HCI for conversation and multimodal analysis, comparing them to demonstrate how the Multimodal Usage Space Analysis (MuSA) tool excels in various metrics against other similar tools. This comparison aims to underscore the unique contributions and advantages of MuSA within the field.

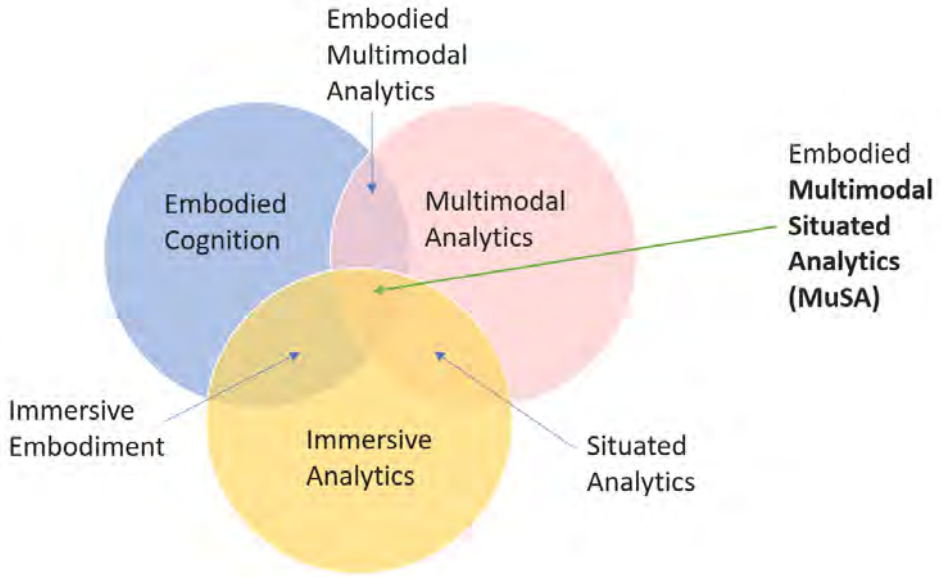


Figure 4: shows the intersection of 3 related fields - Multimodal Analytics, Embodied Cognition, and Immersive Analytics which encompasses the spectrum of AR, MR, and VR. At the intersection of 3 domains lies Embodied Multimodal Situated Analytics.

Our research draws on knowledge from multiple fields of research such as Embodied Cognition, Immersive Analytics, and Immmersive Embodiment. A Venn Diagram (Figure 4) illustrates how the intersection of these fields serves as the foundation for our work.

2.1 Immersive Analytics

Visualization in immersive environments or Immersive Analytics is the use of engaging, embodied analysis tools to support data understanding and decision-making. It combines paradigms from multiple fields such as virtual and augmented reality, data visualization, visual analytics, computer graphics, and human-computer interaction. In the literature, visual analytics is defined as "the science of analytical reasoning facilitated by interactive visual inter-

faces". Researchers often use various visualization techniques and tools to improve performance time and productivity for software analysis, data analysis, and information retrieval. The technologies used for immersive display come in various forms such as room-sized CAVE-like projections [23, 32], Virtual reality Head Mounted Displays (e.g. the HTC Vive, Meta Quest) [25], interactive tables, walls, multi-display environments [88] and portable Augmented Reality head-mounted displays such as HoloLens and ARGlasses.

Research has shown that understanding data visualizations through immersive analytics can be significantly more effective when compared to traditional interfaces. Sawyer et al. demonstrated that collocated collaboration can provide significant benefits by showing through their experiments that team rooms supporting face-to-face activities helped focus the activities of work groups and removed them from interruptions [76]. In a similar work, Teasley et al. showed that "war rooms" with access to tools such as computers, whiteboards, and flip charts were twice as productive as similar teams working in a traditional office environment [87]. In their review of three Immersive Analytics projects undertaken by research teams using the CAVE2 immersive projection environment Marai et al. found significant benefits from teams working together in an Immersive Analytics setting [51, 52].

2.2 Strategic Immersion & SenseMaking in Immersive Analytics

The field of Immersive Analytics encompasses various aspects aimed at eliminating obstacles between individuals, their data, and the analytical tools they employ. This term typically pertains to both technological and psychological immersion and serves as a focal point in Immersive Analytics research [53, 16]. To understand the role of humans in human-machine

cooperative analysis Stuerzlinger et al. [82] explore both strategic immersions through accessible systems as well as enhanced understanding and control through immersive interfaces that enable rapid workflow. Strategic immersion is closely related to high-level problem-solving that involves a state of deep engagement and absorption in an activity, where the individual focuses on employing effective strategies and problem-solving skills to achieve success. The process of sensemaking involves gathering information from the world around us and interpreting it to create understanding. Throughout this process, we collect data, form hypotheses, and reassess our conclusions based on new information as it becomes available.

Sensemaking is a challenging cognitive task that demands creativity, comprehension, mental modeling, and situational awareness [24, 41]. Applications like Immersive Space to Think have been utilized to examine how analysts employ 3D immersive environments for information organization and externalizing their thinking process [6, 49]. These tools can be used to author and modify visualizations thereby providing flexibility for the dynamic nature of the sensemaking process [48, 24]. Greater immersion can offer advantages such as improved perception of depth, reduced visual complexity, enhanced spatial orientation, heightened peripheral awareness, increased information absorption, broader bandwidth, and enhanced engagement [11, 53, 3]. According to Skarbez et al. [78], abstract data visualizations should be incorporated into most existing immersive analytics systems to facilitate knowledge generation. A considerable amount of research has been conducted on developing prototypes for immersive analytics. Cordeil et al.'s ImAxes [22] is one such approach that emphasizes interactive data visualizations and employs embodied interactions to generate visualizations within an immersive analytic system. The

imAxes technique allows users to explore data points by simultaneously moving and rotating multiple axes using a set of intuitive gestures, which facilitates the discovery of patterns and relationships that may be difficult to detect using traditional 2D scatterplots. To utilize the surrounding space to organize semi-structured information [37, 42, 50] have created systems that exploit the benefits of distributed cognition (Figure 5 (a)). These studies demonstrate how presenting and laying out data in an efficient manner within a user's spatial environment in immersive settings can enhance the efficiency of analysis and sensemaking of vast amounts of information.

Immersive analytics enables interaction with large amounts of data at various levels of scale. For instance, it allows manipulation of numerous data objects through multi-touch [68] or physical-navigation aware cone-casting [69] techniques. Moreover, multiple input devices can be utilized to take advantage of the most suitable interactive affordances for each sensemaking task [21]. Another approach, explored by Satriadi et al., examines geospatial data analysis within an immersive analytic prototype [75] (Figure 5 (b)). Yang et al.'s map-based approach employs Tilt Map [97], which uses both 2D and 3D visualizations based on users' interactions with the system to provide on-demand details. These approaches highlight the importance of utilizing multiple input devices to provide customized interactive affordances for different sensemaking tasks. These works offer strategies for effectively presenting and manipulating data to enhance the sensemaking process in immersive environments.

The previous studies mentioned provide significant contributions to the area of sensemaking and strategizing through immersive analytics. However, they do not have the capability to be

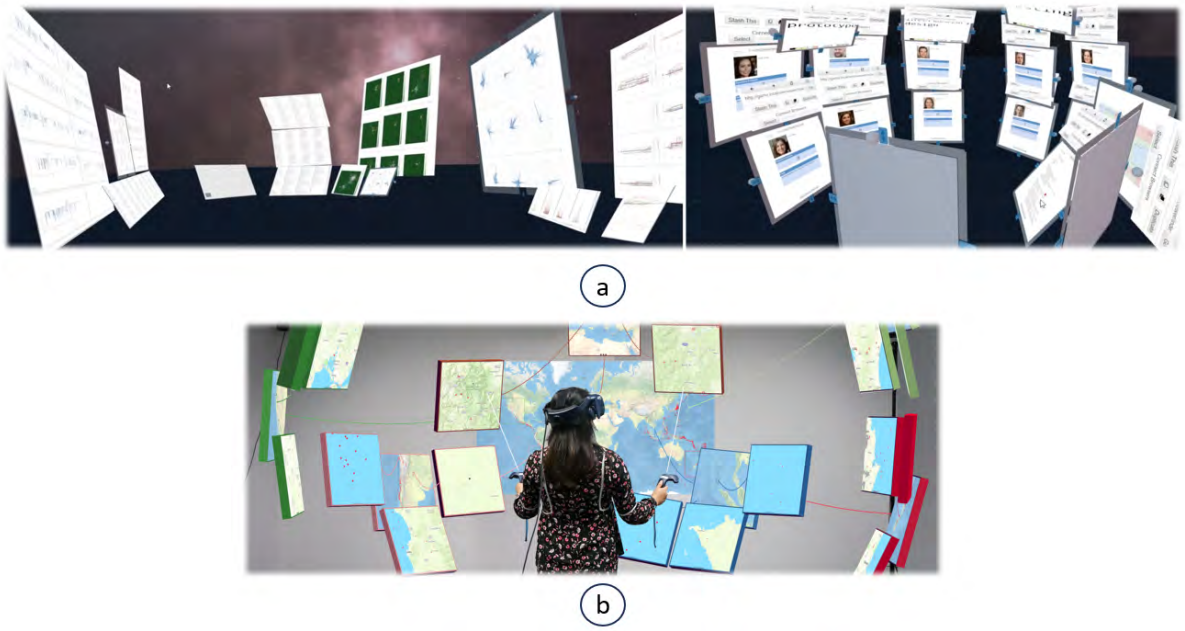


Figure 5: (a) shows Sensemaking through interpretation and organizing - a layout showing utilization of the surrounding space to organize semi-structured information[42] (b) shows Sensemaking and Strategic Immersion in the data - a user examines geospatial data analysis within an immersive analytics prototype [75]

a flexible component in analyzing recorded conversations along with the spatial context. Our research, on the other hand, incorporates the use of embodied situated analytics to provide analysts with the necessary tools for effective sensemaking and strategic gameplay, all embedded within the system. With our approach, we examine the workflow of conversation analysis that allows us to compare behaviors in both mixed and virtual immersive environments.

2.3 Immersive Embodiment & Embodied Visual Analytics

Immersive embodiment refers to the use of technology to create a virtual reality experience that allows a person to feel fully present and embodied in a digital environment. Immersiveness

can prove to be a great tool when it comes to teaching and learning practices. Performing arts is one main domain where immersive embodiment has been extensively researched [10]. Digital performances, particularly Shakespearean productions, have also been using immersive experiences for the past decade for experimenting with students and for theorizing how experiences of presence, liveness, immersion, and interactivity work in such settings [95, 83]. It can also be used in cataloging and managing knowledge bases for collaborations and interventions in occupational safety, and health management systems and for providing an understanding of symptoms and providing a basis for treatment. [34, 31, 98] (Figure 6 (b)). These works provide evidence that immersive embodiment can be an effective method for enhancing user experiences and facilitating a variety of tasks across different domains.

Embodiment in Virtual Reality has also been explored in other domains such as storytelling [92] and exploring experiments involving magic tricks [4] relating the sense of immersion to the science of magic. These works show how embodiment can increase the sense of immersion and help to convey a more engaging and memorable experience. Embodied teaching with a focus on the body of the teacher can demonstrate the relationship between teachers' and online presence using attributes such as face and voice along with the use of silence [20]. Through this work, we understand how embodied teaching highlights the importance of nonverbal cues, such as facial expressions and silence for conveying meaning and creating a strong presence. Thornett et al., in their work, talk about how augmented and virtual reality can be used in scenographic practice to create effective audience experiences [90]. In summary, these examples indicate that embodiment can be a powerful tool to enhance understanding and engagement.

Embodied visual analytics is a term used to describe a data analysis approach that combines interactive visualization techniques with bodily engagement to enhance the sense-making process. Since exploring multidimensional data can be challenging for students, Chen et al. [18] demonstrated an embodied approach for visual analytics designed to teach students about exploring alternative 2D projections of high dimensional data points using weighted multidimensional scaling (Figure 6 (a)). Embodied concepts can be used to engage and assist the general audience in the exploration of data and thus facilitate discovery and insight. [43] (Figure 6 (c)). In this case, the researchers use the user's position and movement to control the content of the exploration space. Although the above-mentioned approaches help in analysis like tasks and attempt to explore sensemaking in immersive environments, these explorations do not occur in situ. Our research aims to explore the impact of spatial situatedness on the comprehension, analysis and embodied sense-making of recorded conversations, as well as the cognitive and perceptual processes that underlie these effects. We extend this work by conducting our study in two distinct immersive environments i.e. MR and VR, and examining their usefulness and effectiveness in analyzing multimodal conversations.

2.4 Situated Analytics

Significant advancements have been made in the field of situated analytics, which has facilitated researchers and analysts in gaining better comprehension, interpretation, and decision-making based on complex data representations. Real-time exploration and analysis of data in the user's physical environment has been made possible through situated analytics [27, 28, 29]. It can be used to create AR and VR authoring tools that leverage information from reality to

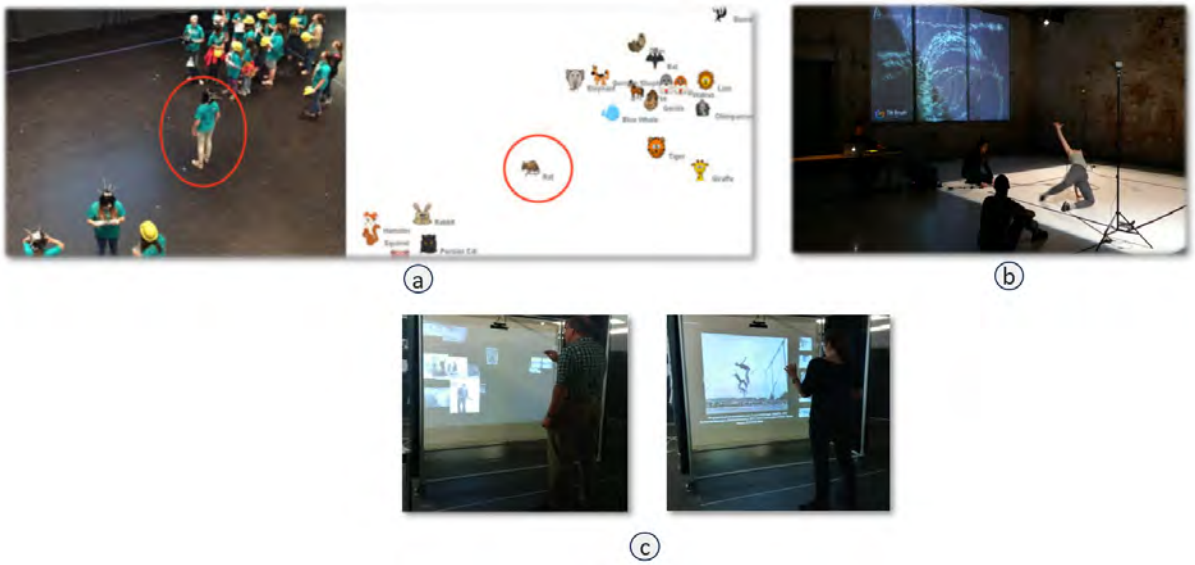


Figure 6: (a) shows Embodiment for teaching students about exploring alternative 2D projections of high dimensional data points [18] (b) Embodiment to generate traces as an effective method for enhancing user experiences [34] (c) Using embodiment to assist general audience in data exploration to facilitate discovery and insight [43]

assist non-experts in addressing relationships between data and pertinent objects [19, 57, 72]. The systems provide users with the contextual information necessary to design embedded and situated data visualizations in a safe and convenient remote setting [71]. Researchers have also explored how in situ analysis may be used for visual search tasks, information retrieval, and exploration and manipulation tasks for information visualized in its semantic and spatial context [15, 47, 14]. Xu et al. [96] study the use of identifying products and displaying detailed information to help consumers make purchasing decisions that fulfill their needs while decreasing the decision-making time. It has also been explored to study human-robot handover tasks using situatedness and gaze modalities [62] (Figure 7 (a)). Mapping data on 3d spatial terrains to pro-

vide insight into data through immersive interactive applications using head-mounted displays has also gained interest. Nguyen et al. [66] use such an interface for understanding the drifting behavior of bees in their natural habitat while Zheng et al. [99] use this for implementing a fieldwork navigation tool to increase the arable land use efficiency. In summary, these studies illustrate the capacity of situated analytics to enable real-time exploration and analysis of data within the user's physical environment. These works demonstrate that situated analytics has the potential to effectively facilitate data exploration tasks across different domains.

Situated analytics has also been used for a wide variety of topics, including in situ interactive exploration of mineralogy spatially co-located and embedded with rock surfaces [30], exploring graphs with node-link structures [46] and for storytelling using (SLAM) enabled augmented reality [40]. These applications show the range and applicability of situated analytics in various fields. It has also been used to train students and professionals for the industry [73] and for scientific visualization of volumes using density-based haptic vibration technique and an adaptation of a cutting plane for 3D scatterplot [71]. Multiview (MV) representations along with situated analytics (Figure 7 (b)) can be used to potentially address complex analytic tasks in immersive visualization [93]. Although in situ data analysis can be promising, the complexity of the data may also necessitate the use of more traditional non-immersive visual analytics setups. Hubenschmid et al.'s work in ReLive [38] demonstrates a pipeline for exploration and analysis by bridging the gap between in situ and ex situ analysis. ReLive offers immersive VR for in-situ analysis and non-immersive desktop view for ex-situ analysis, providing an interactive spatial recording of the original study in the VR view, while the desktop view enables malleable

Figure 7: (a) Using situatedness and gaze modalities for human-robot handover [62] (b) Using situatedness for displaying analytics in sports and board game [93] (c) Using situatedness for reliving experience, and combining it with ex-situ analysis [38]

analysis of aggregated data (Figure 7 (c)). These works show how situated analytics can be used to build context-aware applications for exploration, manipulation, training, and navigation in immersive environments. However, they lack the embedding of situated analytics as a part of an analysis pipeline. They also do not compare the user experiences in AR and VR environments. Our work addresses both these areas thereby adding research insights to this emerging field.

2.5 Proxemics and Multimodal Analysis

The study of proxemics, which examines the significance of spatial relationships in human interactions, has been a key focus in multimodal analysis research. [35]. Extrinsic and intrinsic sensory interference requires such spatial behavior to be dynamic [56]. Human head orientation estimation has been of interest in proxemics because head orientation serves as a cue to directed

social attention. Currently, most approaches rely on visual and high-fidelity sensor inputs and deep learning strategies that do not consider the social context of unstructured and crowded mingling situations. However, alternative inputs, such as speech status, body location, orientation, and acceleration, may also contribute to head orientation estimation [86]. The proxemics of social interactions (e.g., body distance, relative orientation) influence many aspects of our everyday life: from patients' reactions to interaction with physicians, successes in job interviews, to effective teamwork. Tools like Protractor have been developed for measuring interaction proxemics as part of non-verbal behavior cues with fine granularity. [61]. This method takes a new approach to studying interactional proxemics by using automated ways to monitor distance and relative body orientation thereby making the method more reliable. Research has also been done on how proxemic features can be used to understand relationships in product development teams providing evidence that social signals are related to team performance [44]. These works illustrate how multimodal analysis has been utilized to examine and study proxemics.

2.6 Multimodal Analysis in HCI and Tools

Multimodal Analysis has proven to be useful in several areas of HCI, including but not limited to conversational agents, human-robot interactions, and Extended Reality (XR) [62, 80]. Although automated analysis tools have been developed, some researchers still use manual analysis techniques due to the limited generalizability of automated solutions across different research areas and use cases [17, 13]. Video recordings are commonly employed to study conversations, with some studies using a combination of interviews, usage logs, and observation alongside video recordings for a comprehensive analysis, such as Chattopadhyay et al. [17]

who studied group behavior, and Brown et al. [13] who analyzed mobile search in everyday conversations. Gaze direction has also been utilized to examine the effects of video calls on face-to-face conversations and in group settings [54, 8]. Several tools have been developed to mitigate the complexity associated with analyzing conversations [55, 77, 79]. NOVA, as discussed in several studies [7, 36], is a tool that primarily focuses on annotating multimodal behavior and offers features for collaborative human-machine annotation (Figure 8 (a)). Wagner et al. [91] developed SSI (Social Signal Interpretation) framework to facilitate the analysis of behavior by utilizing synchronized data collection from multiple sensors and plug-in detection algorithms. More specialized tools have been built to address specific functionalities. TARDIS [2] was created to assist job interview training in human-avatar interactions, and has features embedded to playback webcam, Kinect, and audio recordings together with visualizations of annotations. MultiSense [81] focuses on the analysis of psychological distress in dyadic interactions, offering both online and offline feedback, and Opensense [79] offers a customizable pipeline editor for choosing use-case-specific modalities. ConAn [70] is another easy-to-deploy and use cross-platform tool that allows users to conduct conversation analysis across multiple modalities (Figure 8 (b)). HuCETA [26] enables human-driven data storytelling interfaces for reflection and decision-making for teachers and students in healthcare (Figure 8 (c)). Backchannels are short interjections of the listener, that serve important meta-conversational purposes like signifying attention or indicating agreement. The MultiMediate challenge addresses, for the first time, the tasks of backchannel detection and agreement estimation from backchannels in group conversations [63]. Bodily Behaviors in Social Interaction (BBSI [5]) is another such

tool that is developed for automatic analysis of Body Language. Table I shows a comparison of conversation analysis tools modified and extended from [70] for comparison with MuSA. The current state of research in the field of conversation analysis has seen the development of various automated tools that can be used for different use cases. However, these tools have not yet explored the possibilities of being used in extended reality environments. While some of these existing tools are more generic and versatile than others, none of them have specifically focused on the use of mixed and virtual reality for conducting conversation analysis. Therefore, with the development of PSA, we aim to provide a framework that is more generic and flexible in its applicability, and which can be employed in any conversational setting in mixed and virtual reality environments.

This chapter has explored how various research fields, including situated analytics, strategic immersion and sensemaking, proxemics, multimodal analysis, embodiment, and conversation analysis, contribute to our understanding of using immersive environments for analyzing conversations thereby informing our research. We also drew comparisons between our work and the related research and discussed how MuSA addresses some of the gaps in existing approaches in analyzing multimodal conversations.

In the next chapter, we will detail the development of our system and describe the initial phase of our user study. This phase focuses on analyzing conversations among seated participants, providing insights into how our system performs in practical settings.

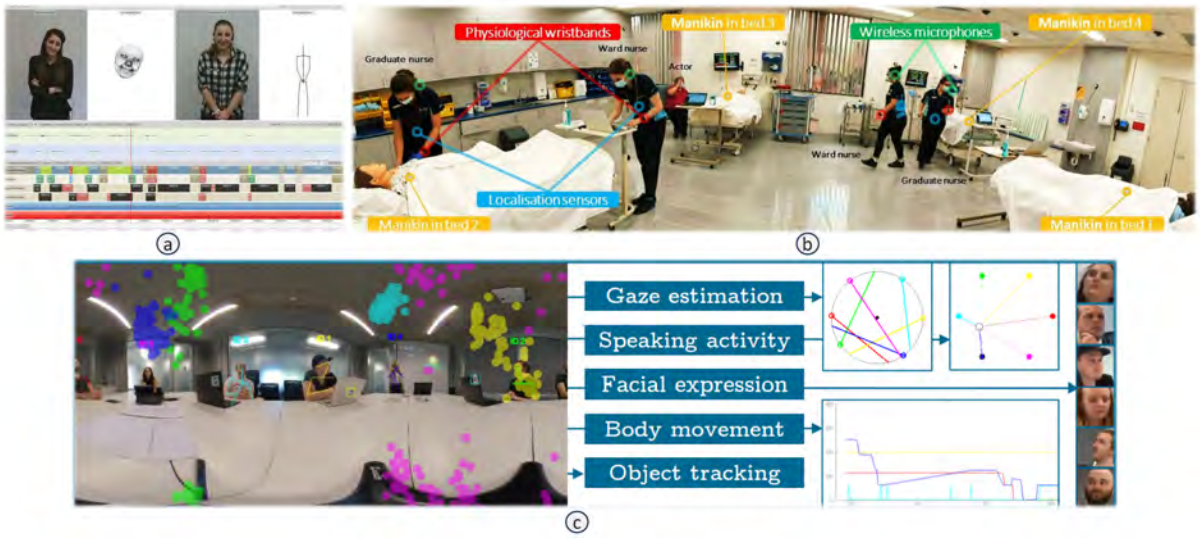


Figure 8: (a) NOVA for annotating multimodal behavior and collaborative human-machine annotation [36] (b) ConAN for conducting Conversation Analysis (CA) using gaze estimation, speaker and facial action unit [70] (c) HuCETA for capturing team activity and to enable human-driven data storytelling [26]

Name	Target Use Case	Modalities	360 Support	Extended Reality
MutualEyeContact[77]	Dyadic Interaction Analysis	Gaze, Facial Expressions	No	No
SSI [91]	Multimodal data recording and feature extractions	Extendable multi-sensor recording framework	No	No
NOVA [36]	Annotation & cooperative machine learning	Extendable annotation framework	No	No
MultiSense [81]	Analysis of dyadic counseling interactions	Speech, Body, Gaze, Face	No	No
TARDIS [2]	Job interview training	Speech, Body, Gaze, Face	No	No
OpenSense [79]	Multimodal data recording and feature extraction	Gaze, Speech, Body Pose, Head Gestures, Facial Expressions, Music	No	No
ConAn [70]	Group Interaction Analysis	Gaze, Speaking Status, Facial Expressions, Body Pose, Object Tracking	Yes	No
HuCETA [26]	Hybrid human-machine multimodal sensing, human-driven data storytelling	Body Pose, Speech, Face and physiological data	Yes	No
MutiMediate'22 [63]	Backchannel detection and agreement estimation from backchannels	Bodily Gestures, Head and hand movement, Face & Gaze	No	No
BBSI [5]	Annotations of complex Bodily Behaviors embedded in continuous Social Interactions	Body pose, gesture, social signals, behavior detection	No	No
MuSA	Generalizable In-Situ Multimodal Analysis	Gaze, Speech, Body Pose, Object Tracking, and Extendable multi-sensor recording framework	Yes	Yes

TABLE I: Overview of Multimodal Analysis Tools

CHAPTER 3

PRELIMINARY STUDY - PHASE I

Parts of this chapter have been published in the proceedings of ISMAR 2023 [64] and UIST 2023 [65].

In this chapter, we delve into the initial version of the MuSA system, starting with an overview of the data collection process which lays the foundation for our analysis. We then shift our focus to the various elements that make up the prototype, providing insight into the core components that define MuSA, as well as its overall design and user interface. Next, we discuss the methodologies involved in recruiting participants, the specifics of the prototype used in the study, and the protocol followed during the user study. This section aims to give a comprehensive view of how we engaged with participants and the structured approach we used for gathering data. Finally, we conclude the chapter by examining the outcomes of our initial user study. We analyze the results, drawing important conclusions, and discuss the key lessons learned, which will guide future improvements and research directions for the MuSA system.

3.1 Methods

3.1.1 Data

We captured 13 live conversations from an approved user study (#2022-0354) where two participants interacted with Articulate+ [84, 85] an always-listening AI agent built to disambiguate requests while also spontaneously presenting informative visualizations on an 18-screen

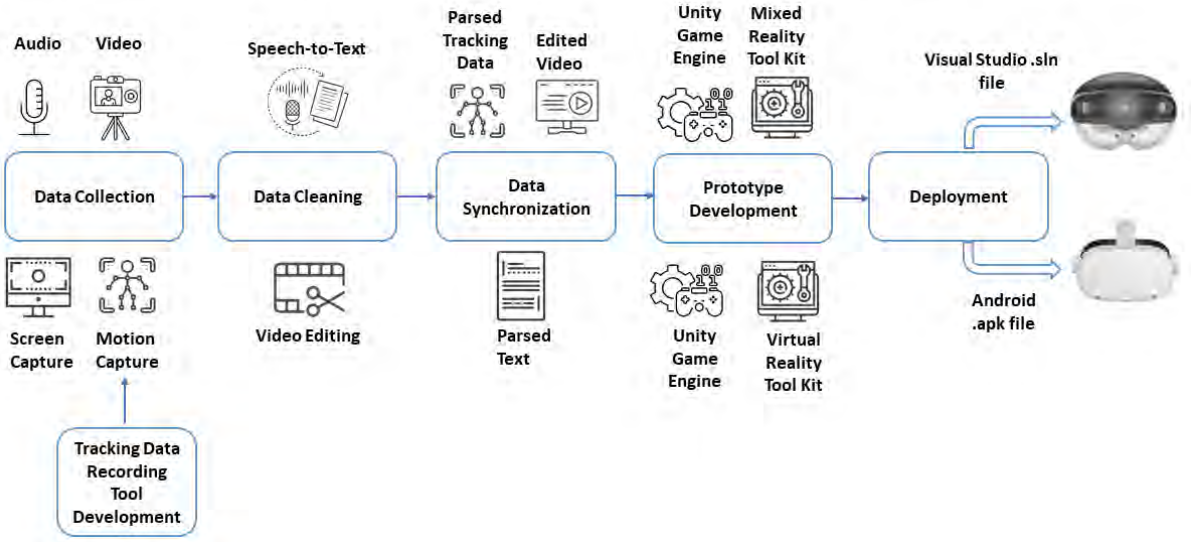


Figure 9: Workflow employed to implement PSA. We start by creating a tracking data recording tool, followed by data collection. This step is followed by Data Cleaning involving speech-to-text conversion, and manual editing of transcription and video. The next step involves the synchronization of all the edited data. This is followed by prototype development in Unity using MRTK and VRTK v4 SDKs. The final step involves deploying the software to HoloLens2 and Quest2 devices.

tiled display wall. Each study lasted for about 1.5 hours and the main goal of the participants was to arrive at answers for a series of questions related to the dataset in question (a COVID-19 dataset in this case) with the assistance of Articulate+. We chose to capture this conversation as it captures the essence of real-time collaborative data analysis of a real-world dataset. Their video, audio, screen usage, and head and body movements were captured. To track the participants' head movements, the participants were asked to wear a hat embedded with OptiTrack markers. Additionally, OptiTrack markers were attached to the chairs used by the participants to capture body movements, while hand movements were not tracked. The room was equipped with an OptiTrack motion capture system, comprised of 24 cameras. Out of the 13 datasets we chose the two best conversation datasets to be explored in our study, one each for the Mixed Reality and Virtual Reality sessions. Figure 10 shows the setup at Continuum during the data collection phase.

The conversations that were chosen for exploration:

1. Had an appropriate conversation length, long enough to have generated enough charts for exploration by analysts.
2. Had voices that were easy to understand such that it could be easily transcribed by Google speech-to-text API.
3. Had complete tracking information in order to generate seated avatars throughout the conversation.

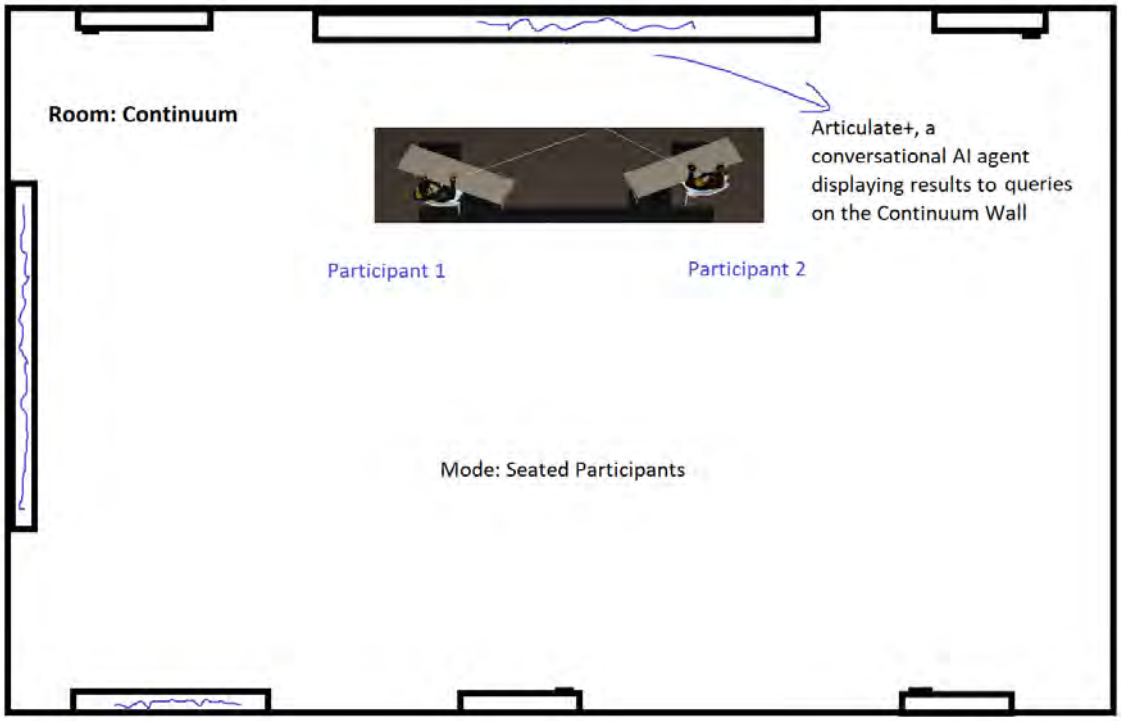


Figure 10: Shows the setup of the data collection for seated participants interacting with Articulate displaying results to queries on the Continuum Wall

3.1.2 Continuum Room Specifications

The continuum room is 20' x 40' space with 10' ceilings designed to be a sensor rich environment for data exploration. It is equipped with a state-of-the-art OptiTrack tracking system comprised of 24 Prime 13W cameras, capable of capturing motion with high precision. This tracking system was used for data capturing. The room also includes temperature and air quality sensors, along with integrated speech to text translation. It boasts a 6x3 tiled 2D 11,520x3240 resolution display wall and a 4x3 tiled 7680x3240 resolution 3D display wall. Activity on the 2D display wall was recorded using screen capture technology. Additionally, the room is outfitted with a 7.1 channel sound bar and two Shure MXA910w room microphone arrays that were used to record audio.

3.1.3 Implementation

We developed a prototype for comparing the experiences in Mixed Reality (MR) and Virtual Reality (VR) environments using the Microsoft HoloLens2 and Meta Quest2 devices, respectively. Figure 11 shows the MuSA implementation in VR. The prototype offers two modes of interaction to navigate to a different point in time: one with a slider and the other with a word cloud. The slider allows users to move to any point of interest in the conversation by dragging the button on either side. When an analyst stops the slider button at a particular point on the slider, the conversation, including video, audio, speech bubbles, and head positions, moves to that point in the conversation. On the other hand, when a word of interest is touched, the word lines for each of the avatars [74] get populated with capsule buttons representing all points in

Figure 11: Shows the application in VR simulated in the room where the conversation originally occurred. It has two participants engaged in a conversation. A line extends from between their eyes marking their approximate gaze location. Their conversation is shown with rising chat bubbles next to them (In the application, we also have the audio to match the rising speech text). On the display, we see all the visualizations generated by the AI agent Articulate+ based on requests it received by the participants. We also see an interactive word cloud consisting of the most highly occurring attributes in the conversation.

time where the words occurred in the conversation. Analysts can touch these capsule buttons to move the conversation to that occurrence of the word in the conversation.

It is worth noting that users interact with the application differently based on the device they are using. In the Mixed Reality part, we allow the users to use gesture-based interactions, such as touch gestures, to interact with menu buttons, words, and capsules on the word line. The main time slider is controlled using the pinch gesture. In contrast, users in the Quest2 part use controller-based interactions, such as using the controller to touch menu buttons, words, and capsules on the word lines, and grab and drag actions to move the slider button.

3.2 User Interface

3.2.1 Chat Bubbles

Through our literature review, we found that speech bubbles moving vertically upwards was one of the best ways to present live captioning of an ongoing conversation in an immersive environment cite. Hence we decided to implement rising speech bubbles next to the avatars representing people in the conversation so it is easier to identify who was talking at any point in time. These bubbles were also color-coded as an added visual cue to differentiate between individuals. The analysts could choose to turn off this feature if they found it distracting or blocking their field of view and were asked to use it only if they found it helpful.

Implementation: The average number of words per bubble was about 9 words. The transcriptions sometimes gave long chat snippets. In such cases we manually broke down the snippet into smaller pieces such that it would be readable by the analyst. We created an animation in Unity to give it an appear, rise, and fade effect. The animation ran for 900 frames and moved

2 meters in height. These numbers were determined through a process of trial and error. For chat snippets that had a time gap of fewer than 1.5 seconds, we set the gap at 2 seconds to make sure there was minimal overlapping between two chat snippets at any point in time. The size of the bubble also varied based on the size. Figure 16(e) shows examples of two chat snippets color-coded for two different people one with a short sentence and another with a longer sentence.

3.2.2 Avatar & gaze information

The participants in the data were represented by Avatars. Their gaze information was shown with a gaze visual (Figure 48). In the first user study we just used one avatar model to represent each of the participants. The avatar representation was used just as a means to show the humanoid representation of the participants and did not convey their gender information or any other physical attributes. In the second user study, we used two distinct avatars: one representing a male and the other a female in the conversation.

3.2.3 Menu

We designed a system menu (see Figure 16) that remains attached to the user, ensuring it's always visible in the analysts' field of view. Given its placement within the field of view, we had to position it close to the user. By toggling the menu button, users can access four functionalities: managing chat bubbles, navigating the word cloud, organizing tasks, and controlling the play and pause features of the application.

3.2.4 Wordcloud, word line, and timeslider

The application offers two modes of interaction to navigate to a different point in time: one with a slider and the other with a word cloud. The time-slider (Figure 16(b)) allows users to move to any point of interest in the conversation by dragging the button on either side. When an analyst stops the slider button at a particular point on the slider, the conversation, including video, audio, speech bubbles, and head positions, moves to that point in the conversation. On the other hand, when a word in a word cloud is touched, the word lines (Figure 16(a)) for each of the avatars get populated with capsule buttons representing all points in time where the words occurred in the conversation. Analysts can then touch these capsule buttons to move the conversation to that occurrence of the word in the conversation.

It is worth noting that analysts use two different interaction modes based on the device they use. In the Mixed Reality version, they use gesture-based interactions, such as touch gestures to interact with menu buttons, words, and capsules on the word line, and they use the pinch gesture to interact with the main time slider. In the Quest2 version, interactions are controller-based. Users interact with menu buttons, words, and capsules on word lines using controller touch, and employ grab and drag actions for the timeslider.

3.2.5 Representing Participant interactions on the display

In the first phase the display wall in the dataset was represented by a 3D model of the display that rendered the video of the screen share session of the AI conversation agent. This display was 1:1 scale with the actual display wall in the room. The screen share activity was synchronized with chat bubbles and tracking data. In the second phase since we only used the

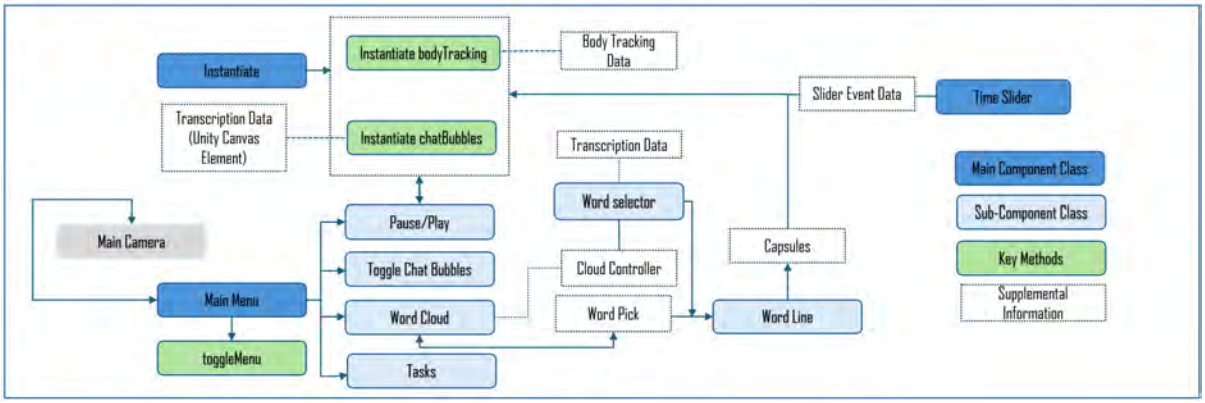


Figure 12: Flowchart detailing the initialization and user interface components of Phase I, with emphasis on data handling, visualization, and programming structure.

MR version of the immersiveness through a see-through device and since the contents of the display were constant we didn't need to replicate the model in the application.

Figure 12 displays a segment of a flowchart labeled "Phase I," which shows it's part of a multi-phase system. Here's a detailed description:

The process starts by setting up two key types of data: Body Tracking Data and Transcription Data. Body Tracking Data is crucial for creating avatars in the system, providing the necessary information to accurately represent movements and gestures. Transcription Data, on the other hand, is vital for creating chat bubbles, which are displayed using the Unity Canvas element, enhancing user interaction by visually representing spoken words.

The main menu is uniquely associated with the user, integrating directly with the camera object. This means that the menu options are visible within the user's viewpoint, ensuring a seamless interface experience directly from the user's perspective.

The Pause/Play feature is designed to control the playback of both the tracking and transcription data, as well as the video content. This allows users to easily manage their viewing and interaction experience within the system.

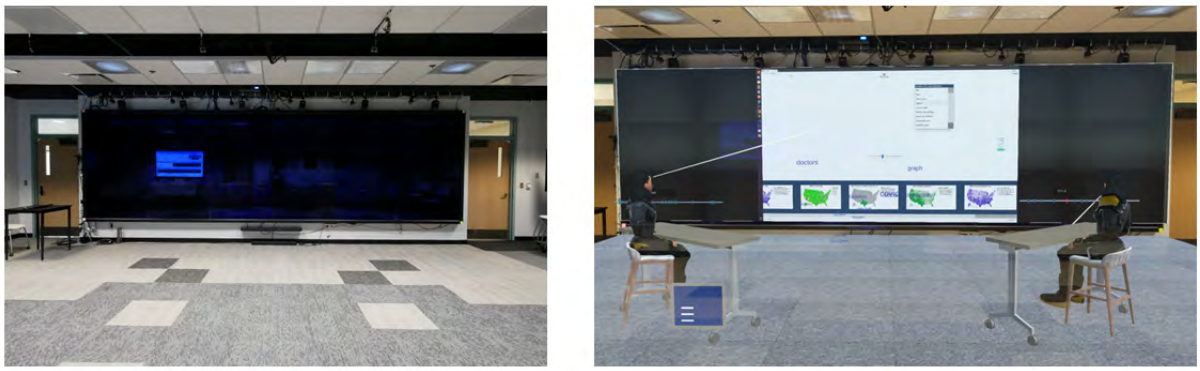
The WordCloud functionality is managed by an object called the cloud controller, which interacts with the word line component. This setup is further linked to the wordSelector, facilitating an interactive feature where selecting a word triggers the generation of capsules. These capsules are tied to both the instantiation component for visual representation and the video content for contextual relevance.

Similarly, the timeslider event data plays a critical role in synchronizing the control over tracking, transcription, and video playback. This feature ensures that users can navigate through the content timeline efficiently, enhancing the interaction with the system by allowing for precise control over the viewing experience.

This overview highlights the interconnectedness and functionality of the system's components, emphasizing the user-centric design and the integration of body tracking and transcription data to create an immersive and interactive experience.

3.3 User Study

We designed a within-subjects user study where the analysts used HoloLens2 in one part and Quest2 in the other part. The order of device usage was counterbalanced across participants. The analysts were given the ability to freely interact, explore and analyze a recorded conversation in immersive environments and understand the content of the conversation. The data used in the experiment is from a previous user study (#2022-0354 with the Articulate+ AI agent).



(a) (b)

Figure 13: (a) Room space where the User Study was conducted. (b) The Room with MuSA as seen through HoloLens2.

We provide each user the opportunity to participate in the study using two different immersive headsets - HoloLens2 which is a Mixed Reality headset and Meta Quest2 which is a Virtual Reality headset. Both devices used in the study are commercially available and widely used by experts and enthusiasts alike. The primary objective of the experiment is to gauge the efficacy of employing situated analytics to comprehend recorded conversations in their spatial context. The analyst would explore different but related datasets using each of the two headsets. The study was conducted in a room shown in Figure 13 located in the Engineering Research Facility at UIC. This room is equipped with large display screens, speech recognition, and motion capture systems. The dataset consists of conversation data in video, audio, and transcribed text formats. The application uses the dataset to instantiate avatars representing people in the conversation and provides the ability to navigate in a simulated environment where all objects in space are at the same location as they were in the original conversation. For the VR compo-

nent, we used a 3D model of a room that closely replicates the actual space, maintaining a 1:1 scale for all objects and dimensions. The model was created by Arthur Nishimoto, a graduate student in the lab at the time, for a separate project. He used Blender and Unity to develop this model [67].

The AR and VR headsets had OptiTrack markers attached so the analysts' head positions and gaze could be tracked. The analysts also wore lab coats with OptiTrack markers at the back of the coat for the entire duration of the study. The lab coat enabled body movement tracking in the classroom in order to record the space usage during the user study. The PI and a student volunteer were present in the classroom to conduct the study. The analysis was conducted in the same room where the data was captured, with the virtual AR and VR rooms precisely mapped and scaled to the physical room's dimensions. While tables and chairs were present in the classroom during data capture, they were removed from the analyst's exploration space to prevent any interruptions or obstacles that could hinder their smooth experience.

3.3.1 Participant Recruitment

We recruited 12 analysts from UIC's student population which consisted of a combination of students from graduate and undergraduate colleges. The pool consisted of 5 female and 7 male analysts between the ages of 20-35. No analyst reported uncorrected vision or motor impairment. The participants were recruited using internal university email lists. All except one analyst reported the Right hand as their dominant hand. Four analysts had never used a Mixed Reality or Virtual Reality Device. Six analysts had never used a Mixed Reality Device, and five analysts had never used a virtual reality device. During the study, we instructed the

participants to rely on the views provided in the application rather than their prior knowledge about COVID-19 data to complete the tasks. Additionally, this study was reviewed by the institutional review board at UIC and determined eligible for exempt research, as it poses minimal to no risks to the participants.

3.3.2 Procedure

Analysts were asked to fill out a pre-experiment survey to ensure that they have stereo vision. If the analysts meet the criteria for the experiment i.e. no visual/motor impairments or prone to motion sickness, they were accepted into the study. The PI then explained the purpose of the study and described the procedures to be carried out. The analyst was informed about their rights, and any questions they had were answered. After signing the consent forms the analysts became a part of the experimental population. Analysts used two immersive applications one on a Microsoft HoloLens2 and one on a Meta Quest2. The analysts start out with the application pre-loaded on their headset, at the beginning of each session. After being comfortable with the training tasks, the analyst would then perform the test tasks. The analysts were informed that they could take up to 15 minutes to complete the training tasks and up to 20 minutes for the test tasks. Head Movement, body movement, and total time taken to complete the assigned tasks were recorded. Additionally, application interaction logs along with a recording of the application session were captured. The analysts performed both training and test tasks using both devices. After completion of each part of the study, the analyst was asked to fill out an application-specific questionnaire and a subset of Witmer & Singer [94] presence questionnaire to evaluate their experience and sense of presence during the study. Figure 14 and Figure 15

show the side-by-side view of an analyst's view and analysts during the exploration in MR and VR, respectively. In the first image, we see the analyst wearing a HoloLens2 device; in the second image, we see an analyst wearing a Quest2 device.

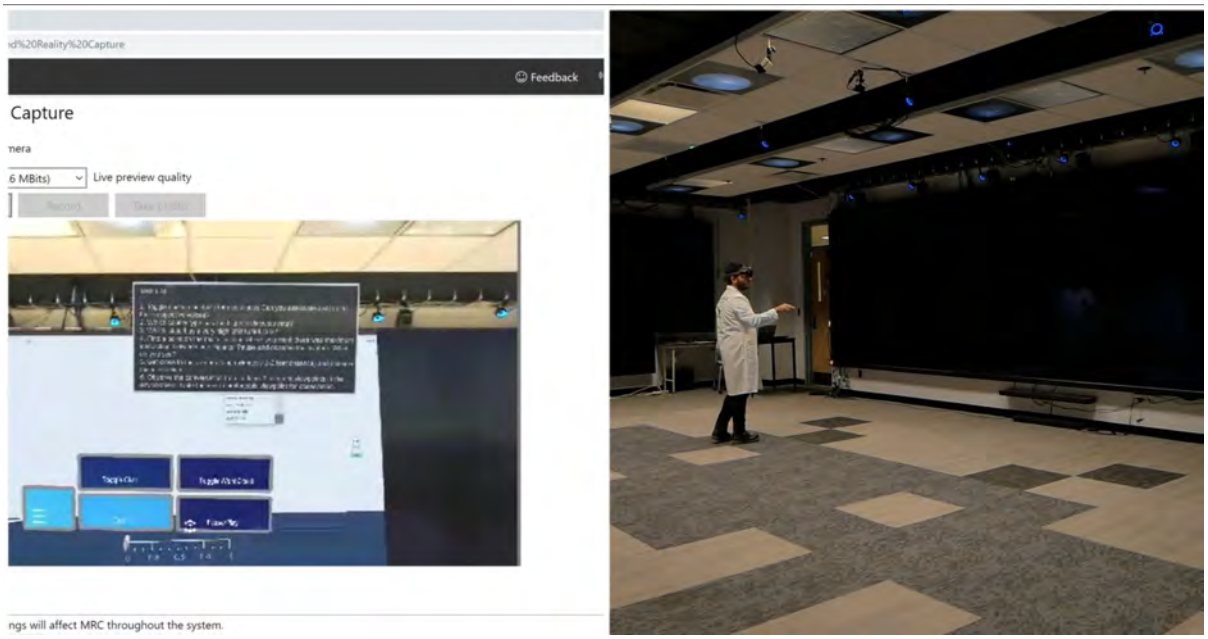


Figure 14: Shows a side-by-side view of the MuSA application as seen by the analyst recorded through Hololens2's live capture and the analyst engaged in the exploration.

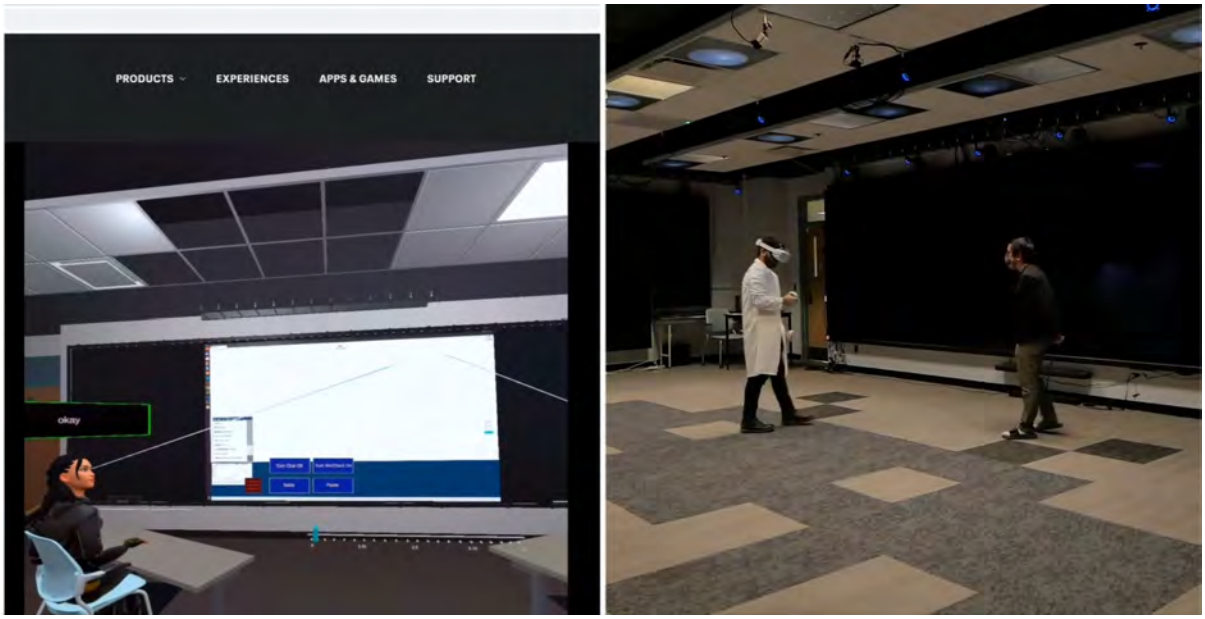


Figure 15: Shows a side-by-side view of the MuSA application as seen by the analyst recorded through Quest2's live capture and the analyst engaged in the exploration.

3.3.3 Tasks and Rationale

3.3.3.1 Training Tasks

1. Use the Time slider to update your current position in the conversation.

2. Toggle all the menu buttons and notice changes in the environment.

3. Which region in the US was affected the most by COVID?

3.3.3.2 Rationale for Training Tasks

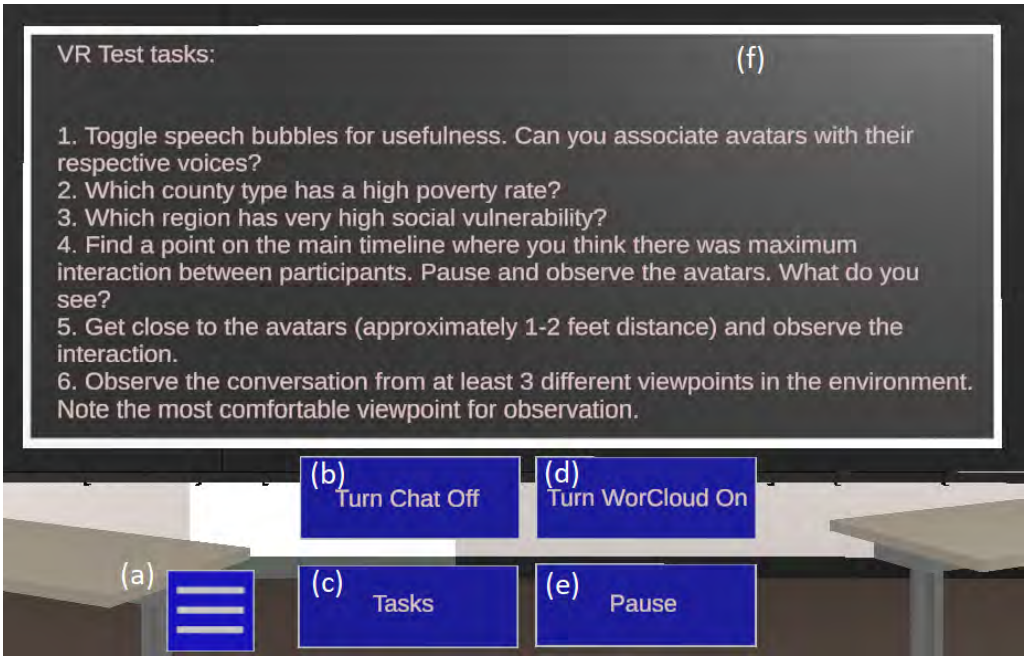
Training on each device consisted of three tasks and was mainly designed to -

1. Acclimatize the analyst with the environment.
2. Equip the analyst with the necessary skills to use the interface
3. Have the analyst use the interface such that they can arrive at an answer to a data-specific question.

3.3.3.3 VR Test Tasks

1. Toggle speech bubbles for usefulness. Can you associate avatars with their respective voices?
2. Which county type has a high poverty rate?
3. Which region has very high social vulnerability?
4. Find a point on the main timeline where you think there was maximum interaction between participants. Pause and observe the avatars. What do you see?
5. Get close to the avatars (approximately 1-2 feet distance) and observe the interaction.
6. Observe the conversation from at least 3 different viewpoints in the environment. Note the most comfortable viewpoint for observation.

Figure 16: Menu implementation in the VR environment. A similar menu was also created for the mixed reality part. (a) is the main menu button which is always available in the analysts' field of view. On toggling changes the menu buttons are invisible. By default, menu is minimized. On toggling it on 4 buttons become available. (b) Turn Chat Off button- is used to toggle the rising speech text. By default, rising speech text is on. (c) Tasks button - Used to toggle the tasks board. By default Tasks board is turned off. (d) Turn Word Cloud button - Used to toggle word cloud. By default Word Cloud is turned off. (e) Pause - Used to toggle between Play and Pause. Toggling it pauses the rising speech text, video, audio, and head movements of the avatars and does not stop the physics in the environment. By default, the application is in Play Mode. (f) Tasks Board - lists the tasks for the current session. By default, the tasks board is turned off.



3.3.3.4 MR Test Tasks

MR Tasks only differed in Tasks 2 and 3 in order to mitigate any potential bias stemming from prior experience. The two changes were:

2. Which county type has the highest diabetes rate? 3. Which state has a very high uninsured rate?

3.3.3.5 Rationale for Test Tasks

The Phase I of our evaluation mostly focused on understanding the usability of the MuSA prototype. In order to achieve this understanding we developed 6 different test tasks. The test tasks were designed to test the user's level of comfort with the application and the features that are accessible through the application. The first task was intended to gauge if the analyst is able to identify and tie the avatars to the audio and visuals in the environment. This is important because both avatars look the same and appear feminine, even though the dataset may contain voices from other genders. Also, the avatars' lips were stationary. Hence, the analysts would not able to use lip sync to identify who among the two participants was speaking. Tasks 2 and 3 were designed to understand if the application is easy enough to arrive at answers to data-specific questions in the environment. This was important to make sure the analysts could follow the conversation and reach at conclusions. Tasks 4, 5, and 6 were mainly present to understand if the analysts could maneuver the application such that they could understand how the participants were interacting with the AI Agent and with each other. Additionally, these tasks also shed light on space usage, and analysts' behavior when close to the avatars.

3.3.4 Survey Rationale

Pre-Study Survey: Firstly, we wanted to ensure that our user base did not have any visual or motor impairments or susceptibility to virtual reality sickness that would make it difficult for them to perform the experiment. Additionally, we aimed to investigate how prior experience

with mixed reality, virtual reality, or both technologies would influence their performance in the study. We also wanted to examine whether the dominance of one hand over the other (i.e., being right-handed or left-handed) had any impact on the users' ability to interact effectively with the virtual environments.

Post-Study Survey: We primarily sought to understand three key aspects of the users' experience. First, we examined whether users had a preference for one type of environment over the other—specifically, whether they favored mixed reality or virtual reality for performing exploration and analysis tasks. Second, we evaluated the usability of the application by determining how effectively users could navigate and utilize it, whether they could easily complete assigned tasks, and which types of tasks they found most engaging or challenging. Third, we looked into how the inherent features and capabilities of the device influenced users' exploration patterns and behaviors. Lastly, we assessed the overall effectiveness of the application in providing a sense of immersion and enabling mobility, ensuring that users felt fully engaged and free to move within the virtual or mixed reality environments.

3.4 Results

We present the results and detailed discussion of the performance evaluation for individual modules of the *Multimodal Situated Analytics* pipeline. Overall, all participants reported they were able to learn the gestures and actions required to interact with the applications in both environments during the training parts of the experiment.

3.4.1 Device Comfort

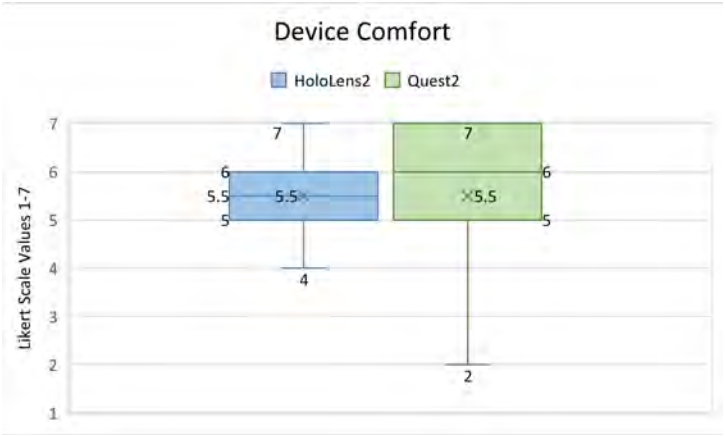


Figure 17: Shows device comfort levels experienced by analysts HoloLens 2 and Quest2 devices on a Likert scale of 1 (very uncomfortable) -7 (very comfortable).

All analysts reported experiencing little to no discomfort on both Quest2 and HoloLens2 devices. Each analyst was able to successfully finish all tasks during both the training and testing stages while utilizing both devices. They were asked to rate their comfort level on a Likert scale of 1 (very uncomfortable) -7 (very comfortable). The distribution of the Likert scale values for both devices can be seen in Figure 17.

3.4.2 Task Outcomes - Strategizing and Sensemaking

All analysts (except one in the VR environment) were able to associate the avatars with their voices in both MR and VR environments (Task #1). Two different strategies were used

to achieve this answer. First, the analysts used rising speech bubbles next to the avatars to associate the voice to the avatars as instructed in the task. However, some analysts used head movements to associate the voice with the avatars. This behavior goes to show that sometimes analysts may go beyond the laid-out rules and instructions to perform a task. Next, all users were able to arrive at near-accurate answers for both data-specific questions (Task #2 and Task #3) in both environments. One scheme that users employed to find the answer was to touch the most relevant attribute in the word cloud, use a point on the world lines associated with both avatars, and observe the charts on the screen. They would have to repeat this several times to find a point in data where an appropriate chart would occur. However, some analysts used several other relevant words that they thought might lead them to the answer. Some answers were purely based on the charts that appeared on the screen while others used both voice and charts to arrive at an answer. One task required the analysts to identify the state with a high uninsured rate (Task #3 in MR). When unable to state the name of the State the analysts were able to precisely point at the region on the map where a high uninsured rate existed. This was in line with the actual answer to the question and was recorded through live capture of the HoloLens2 device. Figure 18 shows an example of an analyst pointing to the state of Texas when unable to mention the name. Figure 19 and Figure 20 show the usefulness and accessibility of data attributes in MR and VR respectively.

For task #4 where the analysts had to find a point of maximum interaction between the participants, some simply used the main time slider to stop at different points in the conversation and check where most charts were visible on the screen, while others used both screen outcomes

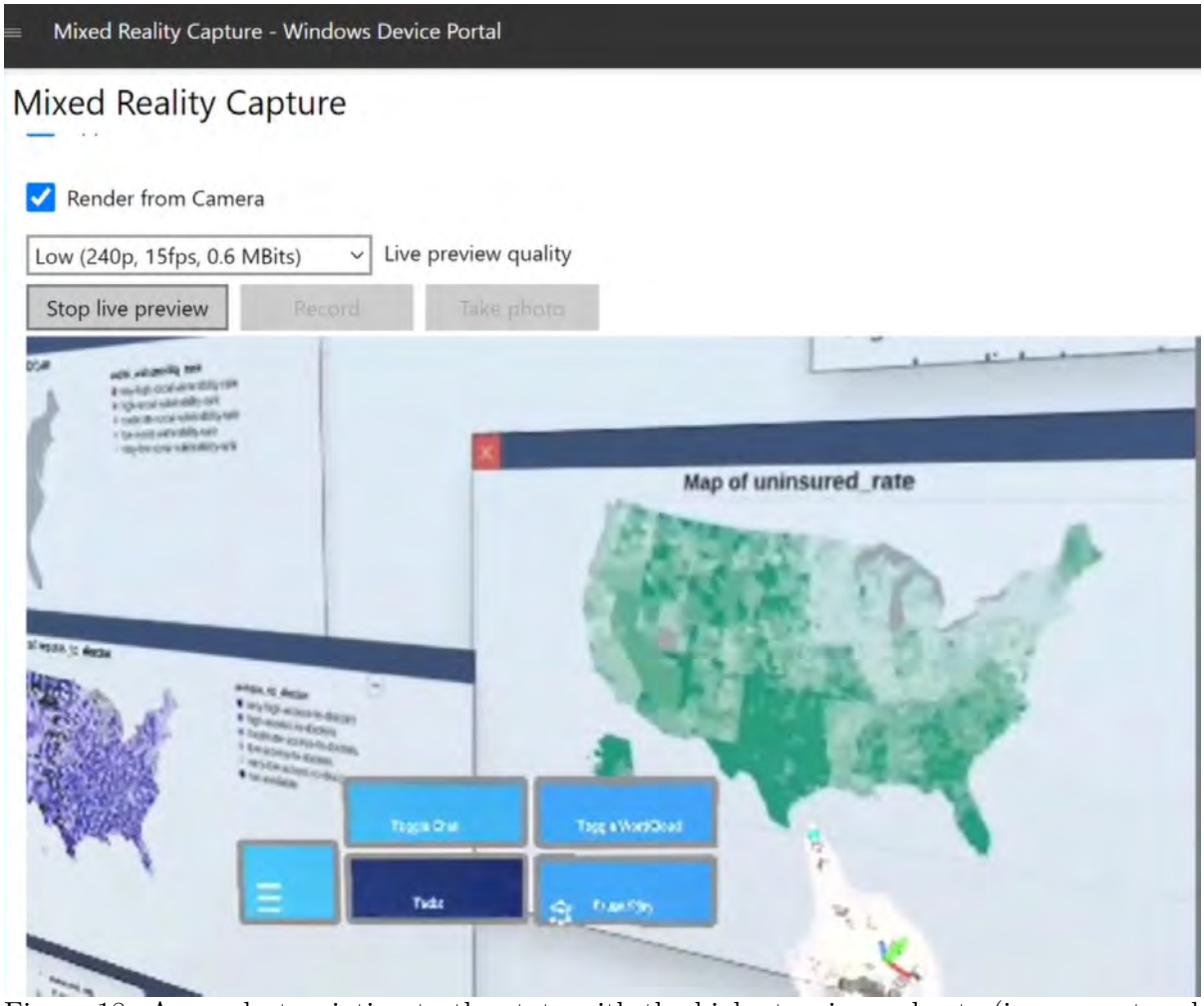


Figure 18: An analyst pointing to the state with the highest uninsured rate (image captured in low resolution, thereby reducing chances of overheating of HoloLens2 during analysis))

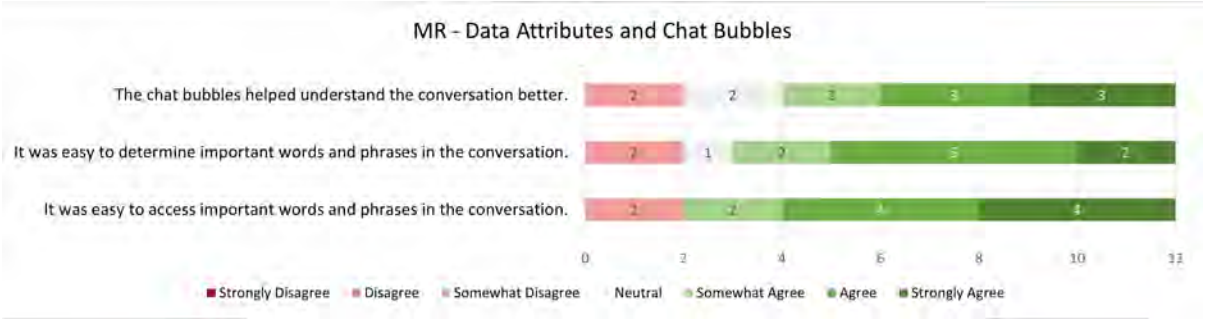


Figure 19: The survey responses of 12 analysts on the usefulness of chat bubbles and access to data attributes in MR, using a Likert scale ranging from 1 to 7, where 1 represents strongly disagree, and 7 represents strongly agree.

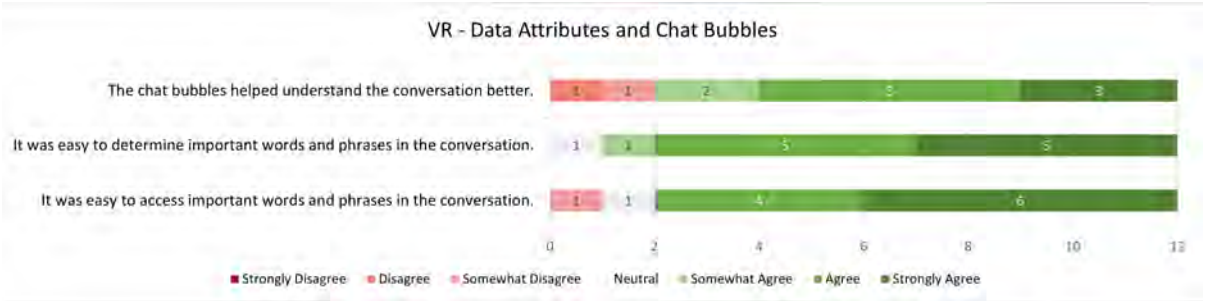


Figure 20: The survey responses of 12 analysts on the usefulness of chat bubbles and access to data attributes in VR, using a Likert scale ranging from 1 to 7, where 1 represents strongly disagree, and 7 represents strongly agree.

and participants' voice to arrive at the answer. Some analysts said they believed that most interaction occurred at the beginning of the conversation, where the participants were talking to each other and the AI agent to understand how the agent works to get the desired outcomes. Some analysts said the maximum interaction occurred in the middle and at the end where they had generated a lot of charts and were trying to summarize their results. Both answers were acceptable for our purposes as there were at least certain portions at the beginning, middle, and end of the conversation where a considerable amount of interactions occurred between the participants and the AI agent. For task #5 all participants were able to get close to the avatars and observe interactions between participants and the AI agent. Through a line drawn from the avatar's head, analysts could tell whether the participants were looking at each other, the desk, or the screen at any point during the conversation. Answers for task #6 are discussed in section 6.7.

3.4.3 Task Completion Times

Although the analysts were informed they could take up to 15 minutes for training tasks and 20 minutes for test tasks these time limits were not imposed. Some analysts voluntarily used the devices longer as they seemed engrossed in the experience/analysis and were focused on completing all the tasks. We saw that after the training task, all participants got comfortable with the device usage and interactions with the application. No tasks were skipped. Test task completion times had a mean of 17.00 minutes ($\pm$5.9 s.d.) in MR and had a mean of 16 minutes ($\pm$3.69 s.d.) in VR. A t-test (two-tailed, two samples with equal variance) showed no significance between the task completion times of the two groups (p-value = 0.59). Figure 21

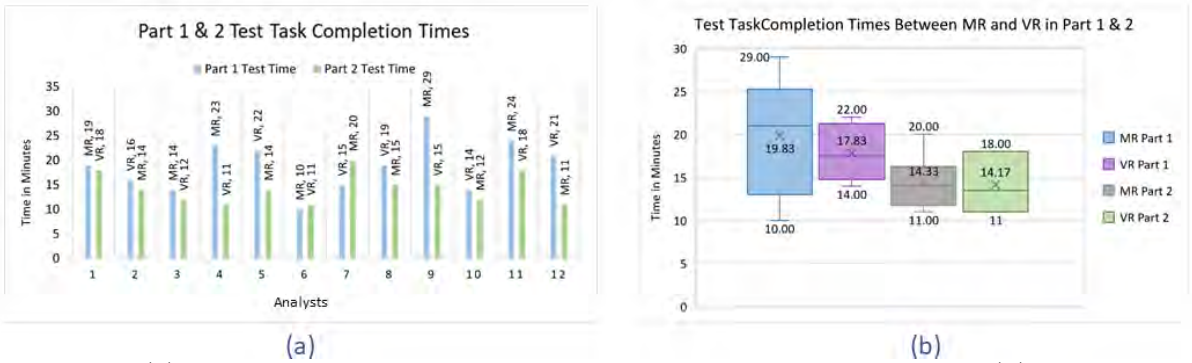


Figure 21: (a) Test Task Completion times for analysts in part 1 and part 2.(b) Distribution for test task completion times in MR and VR environments in part 1 and part 2. © 2023 IEEE

(a) shows the test task completion times for MR and VR experiences and Figure 21 (b) shows the distribution of test times between Parts 1 and 2.

3.4.4 Space Usage

We analyzed the space usage of 6 analysts in both MR and VR environments for the first 10 minutes. We observe that the analysts were more concentrated in the center of the exploration space during the first 3 minutes of their exploration in both MR and VR environments as seen in Figure 22 (a) and (b). As time progresses, we do observe that most analysts start exploring the environment and start spreading out. Between 3-6 minutes there was more activity in the rest of the space and it's even more at the next 4-minute interval as seen in Figure 23 (a) and (b).

Additionally, we analyzed the space usage of individual analysts and observed that irrespective of the environment they show similar patterns of space usage. Figure 24 shows the exploration patterns of Analysts 6,7,10 and 11 in Mixed Reality and Figure 25 shows the explo-

ration patterns of Analysts 6,7,10 and 11 in Virtual Reality. These results shows that the space usage patterns are independent of the device and potentially dependent on sensemaking capabilities afforded by the application for an individual and the choices they make to accomplish a task at hand.

To gain insight and observe any variations in exploration patterns between six analysts for both mixed reality (MR) and virtual reality (VR) environments, we generated heatmaps. We used the Unity's physics raycast system to generate the heatmaps by drawing a rays from the head mounted device captured at 17 frames per second. This dataset gives us 1020 points per analyst leading up to a total of 6120 points/minute for 6 analysts. These heatmaps measure activity in units of one minute, allowing us to gain clarity and identify any significant differences. Our analysis indicates that there is a concentration of activity around the center, with analysts gradually spreading out over time. As previously noted, the exploration patterns appear to be quite similar across both MR and VR environments. The usage of space in MR and VR environments by six analysts during their exploration is shown in Figures Figure 26 through Figure 35.

3.4.5 Distance Traveled

We computed the distance covered by six analysts during the first ten minutes of their analysis session in both MR and VR environments. The plot in Figure 36 illustrates the distance covered by the analysts in meters during the first ten minutes of the analysis session in both environments. We observe that in MR environment, the distance gradually increases until the 5th minute and then gradually decreases towards the end. However, in VR environment, we

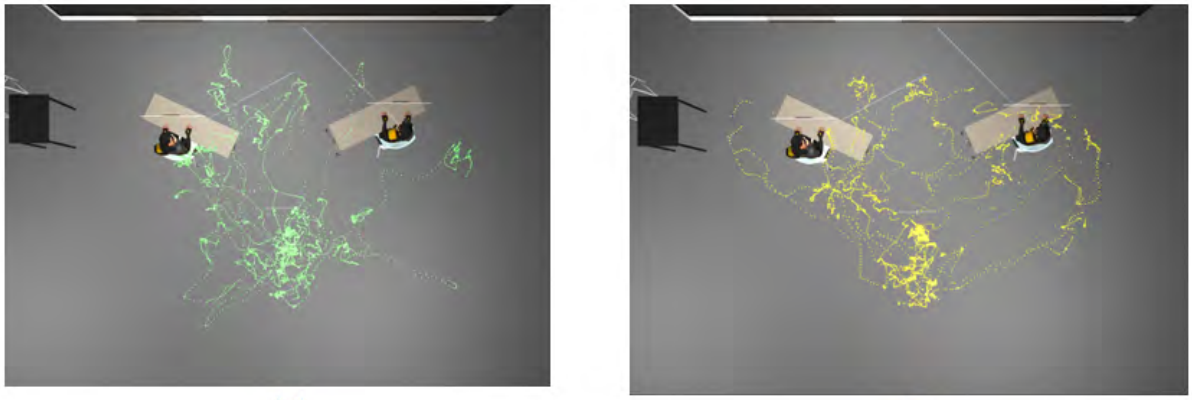


Figure 22: (a) and (b) Space usage of 6 analysts for the first 3 minutes in MR environment and VR environments, respectively.

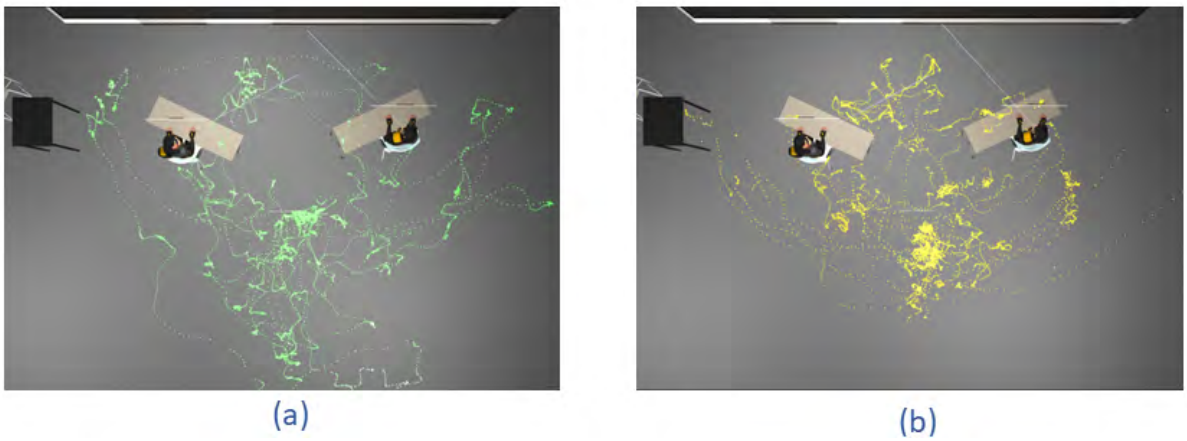


Figure 23: (a) and (b) Space usage of the same analysts from 7-10 minutes in MR environment and VR environments, respectively.

Figure 24: shows the space usage of 4 analysts in MR environment for the first 10 minutes

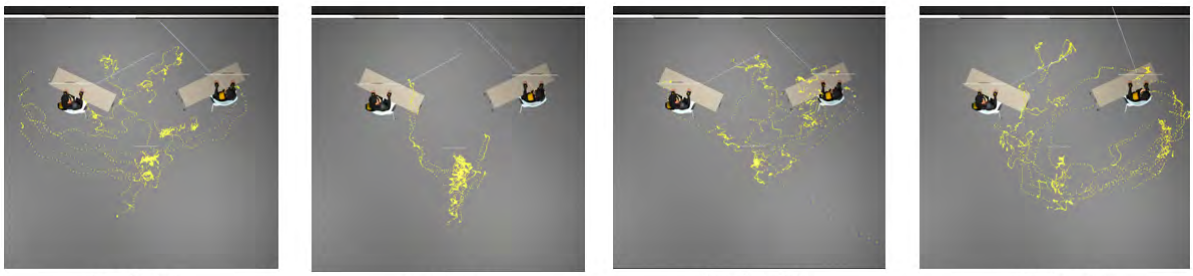


Figure 25: shows the space usage of the same analysts in VR environment for the first 10 minutes.

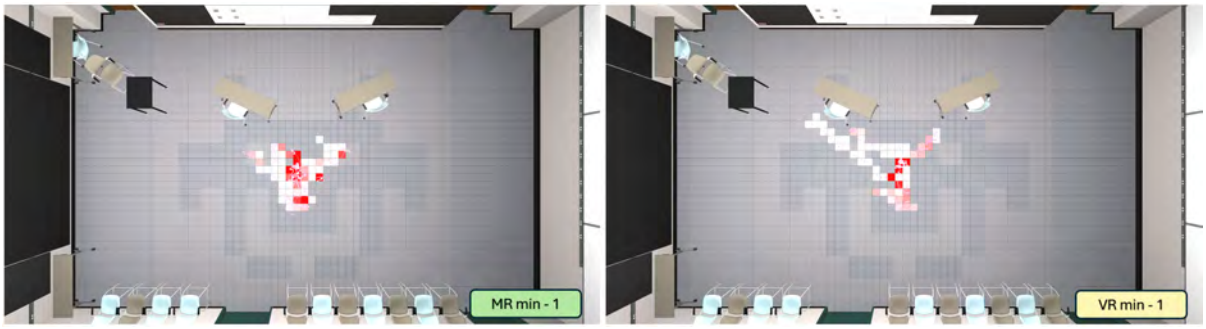


Figure 26: shows heatmaps for space usage of 6 analysts in MR and VR environments for the 1st minute

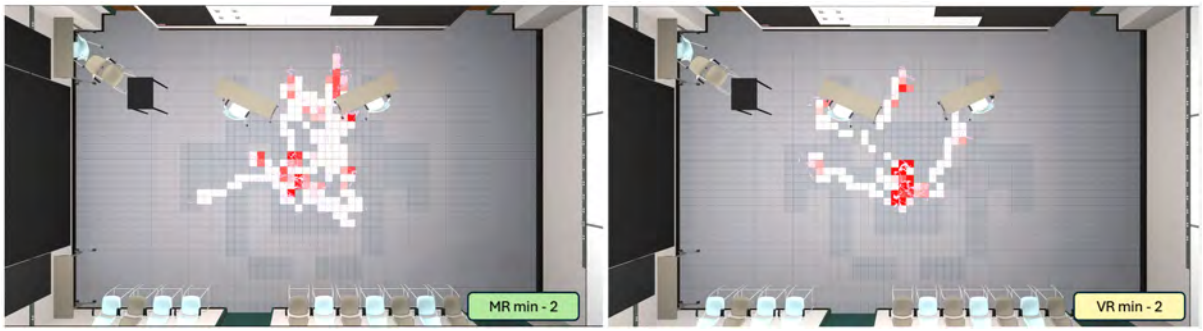


Figure 27: shows heatmaps for space usage of 6 analysts in MR and VR environments for the 2nd minute

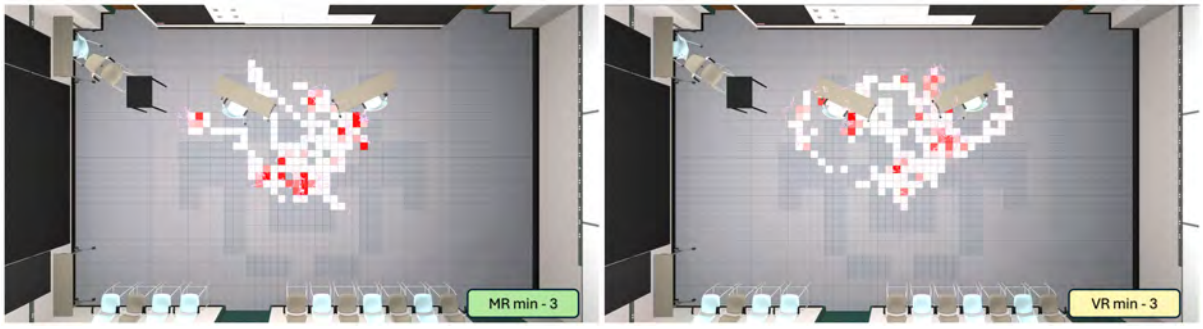


Figure 28: shows heatmaps for space usage of 6 analysts in MR and VR environments for the 3rd minute

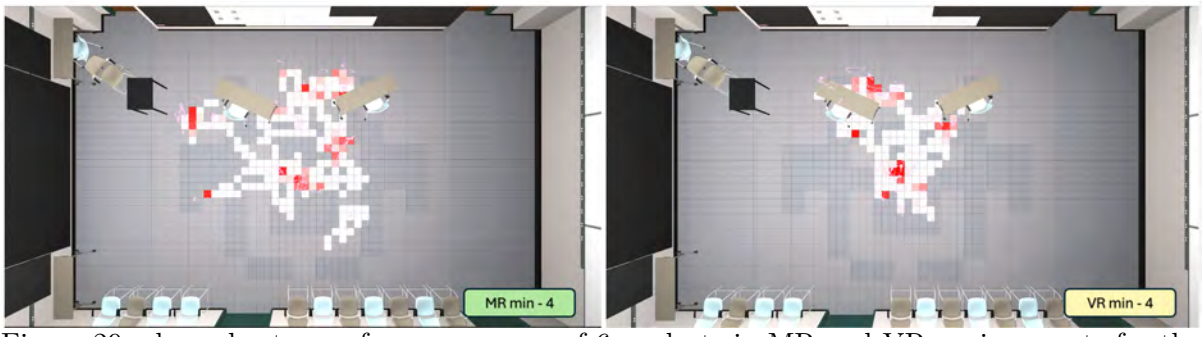


Figure 29: shows heatmaps for space usage of 6 analysts in MR and VR environments for the 4th minute

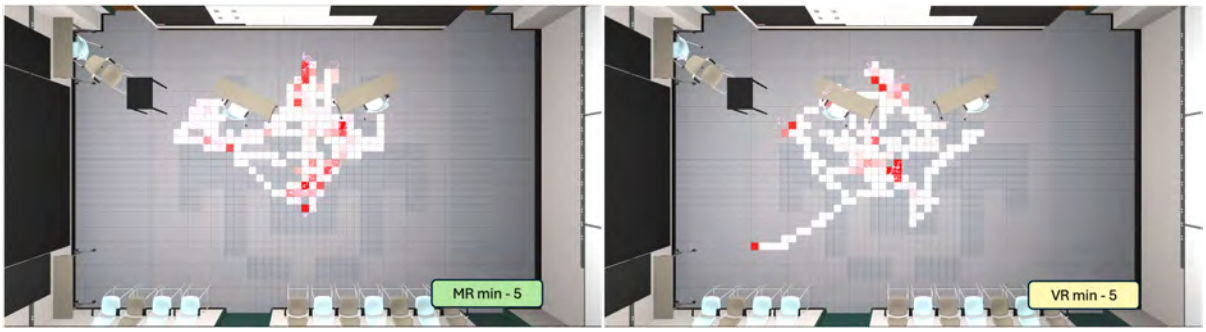


Figure 30: shows heatmaps for space usage of 6 analysts in MR and VR environments for the 5th minute

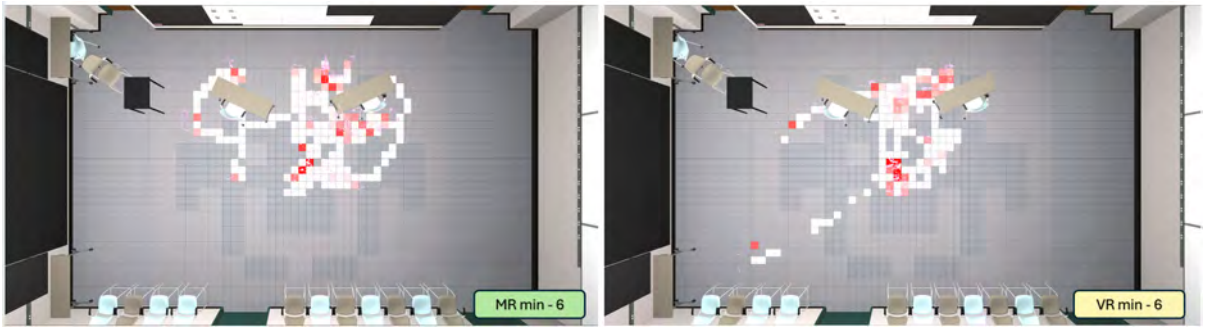


Figure 31: shows heatmaps for space usage of 6 analysts in MR and VR environments for the 6th minute

Figure 32: shows heatmaps for space usage of 6 analysts in MR and VR environments for the 7th minute

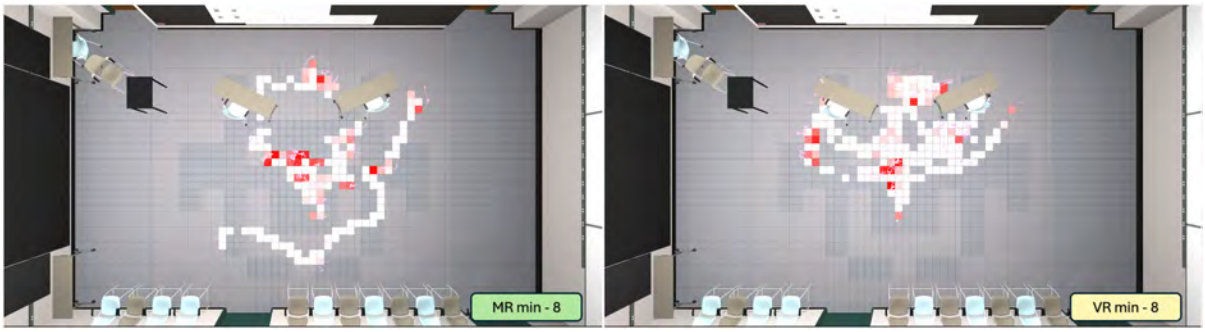


Figure 33: shows heatmaps for space usage of 6 analysts in MR and VR environments for the 8th minute

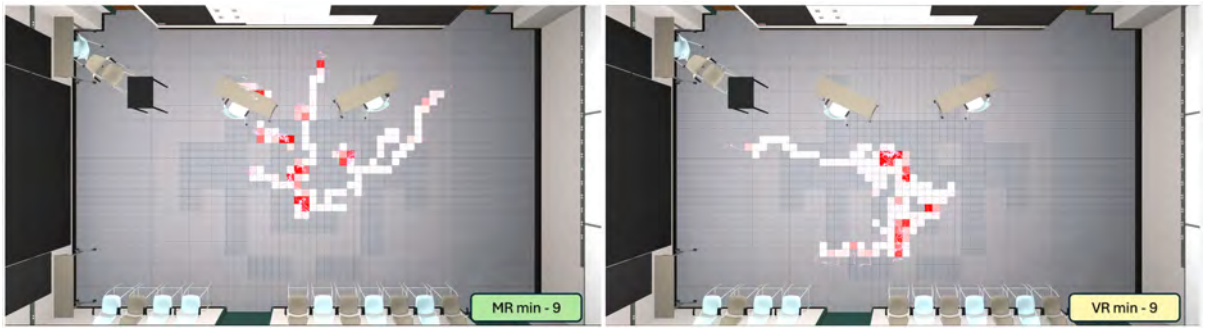


Figure 34: shows heatmaps for space usage of 6 analysts in MR and VR environments for the 9th minute

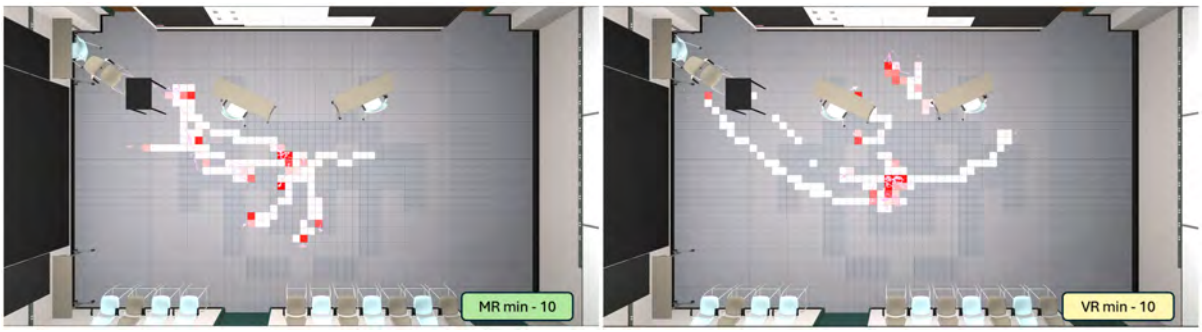


Figure 35: shows heatmaps for space usage of 6 analysts in MR and VR environments for the 10th minute

Figure 36: Distribution of distance traveled by 6 analysts across minutes 1 through 10 (a) in MR (b) in VR

observe two comparable peaks at the 3rd and 8th minute and a gradual decrease towards the extremities of the graph. The mean distance covered by six analysts in the MR environment is 105.82 meters, whereas, in the VR environment, it is 108.34 meters.

The rate of change in position is defined as:

$$R = (\sum s / \sum t) \tag{3.1}$$

where s = distance traveled by an analyst per minute and t is the total time in seconds.

To maintain uniformity across analysts we calculate the rate of displacement in position for 10 minutes. Hence t is constant, i.e. t = 10 * 60 = 600 seconds. Figure Figure 37 shows the distribution of rates of change in position for 6 analysts in MR and VR environments.

Figure 37: Distribution of rate of change in position for 6 analysts in MR and VR

Figure 38: Distribution of means of factors contributing to the Possibility to Act (a) and Possibility to Examine (b) in MR and VR environments rated on a Likert scale of 1 (not at all/not responsive)-7(completely/completely responsive). © 2023 IEEE

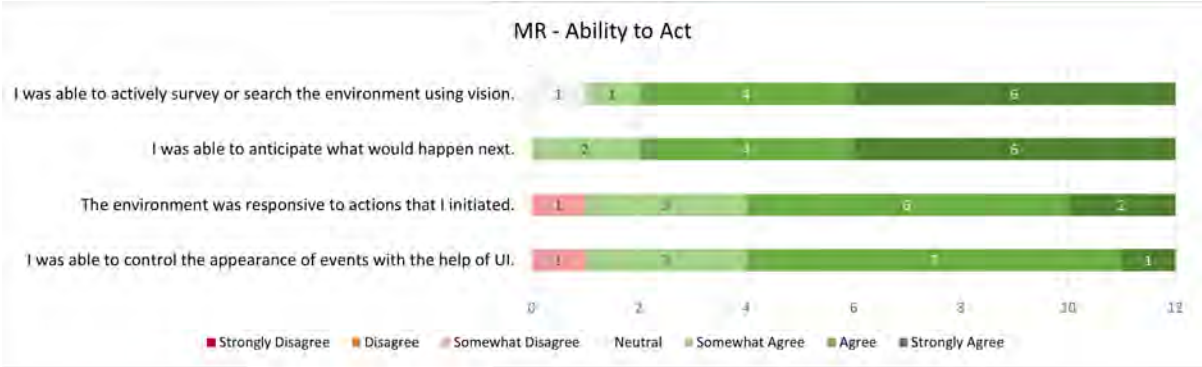


Figure 39: The survey responses of 12 analysts on the possibility to act in MR, using a Likert scale ranging from 1 to 7, where 1 represents strongly disagree, and 7 represents strongly agree.

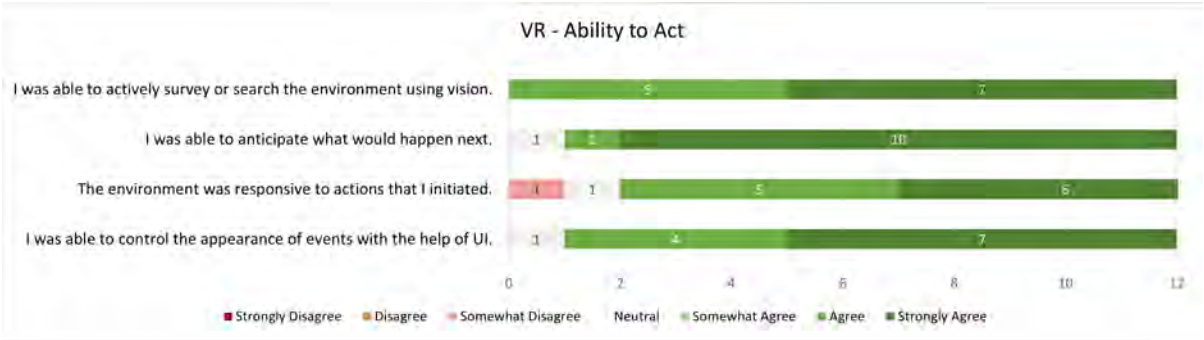


Figure 40: The survey responses of 12 analysts on the possibility to act in VR, using a Likert scale ranging from 1 to 7, where 1 represents strongly disagree and 7 represents strongly agree.

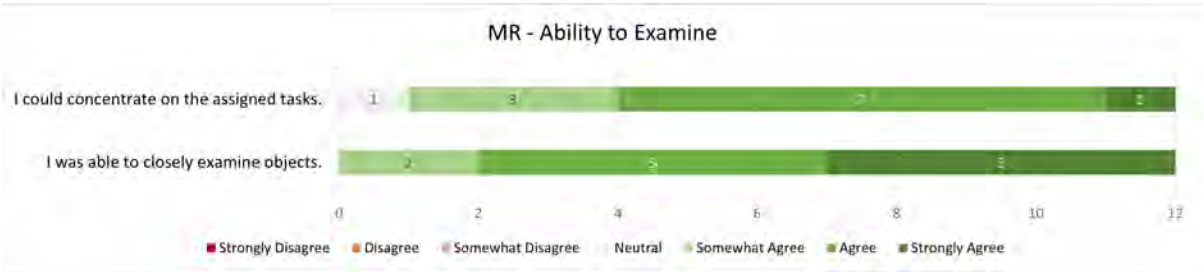


Figure 41: The survey responses of 12 analysts on the possibility to examine in MR, using a Likert scale ranging from 1 to 7, where 1 represents strongly disagree and 7 represents strongly agree.

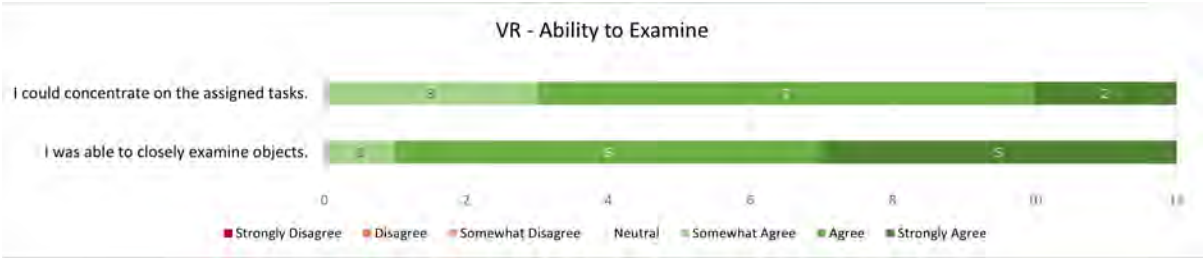


Figure 42: The survey responses of 12 analysts on the possibility to examine in VR, using a Likert scale ranging from 1 to 7, where 1 represents strongly disagree, and 7 represents strongly agree.

3.4.6 Possibility to Act and Examine

In order to evaluate the usability of the application, we used a subset of questions from the Witmer Singer presence questionnaire [94]. These questions were answered on a Likert scale of 1 (Not at all/Not responsive) to 7 (Completely/Completely Responsive). 4 questions were used to evaluate the possibility to act in the environment i.e. to understand analysts' ability to control the events, act, anticipate, and survey the environment. The average values of all 4 answers for each analyst were recorded. Figure 38 (a) shows the distribution of the Likert scale values of all analysts for the MR and VR environments. Additionally, 3 questions were asked to evaluate the possibility to examine the environment i.e. to understand the analysts' ability to closely inspect objects, concentrate on tasks and change viewpoints at convenience. The average values of all 3 answers for each analyst were recorded. Figure 38 (b) shows the distribution of the Likert scale values of all analysts for the MR and VR environments. Figure 39 & Figure 40 show the Likert scale distributions for the ability to act in MR and VR, respectively. Similarly, Figure 42 and Figure 41 show Likert scale distributions of the ability to examine in MR and VR, respectively.

3.4.7 Best Viewpoint

One of the tasks involved users looking at the space from different viewpoints in both MR and VR environments. The task was intended to understand two things: 1) the most comfortable location in space to access the information needed for analysis in both MR and VR environments. 2) How the field of view affected their exploration in both MR and VR environments. Based on the survey responses, it was found that most users in the MR environment

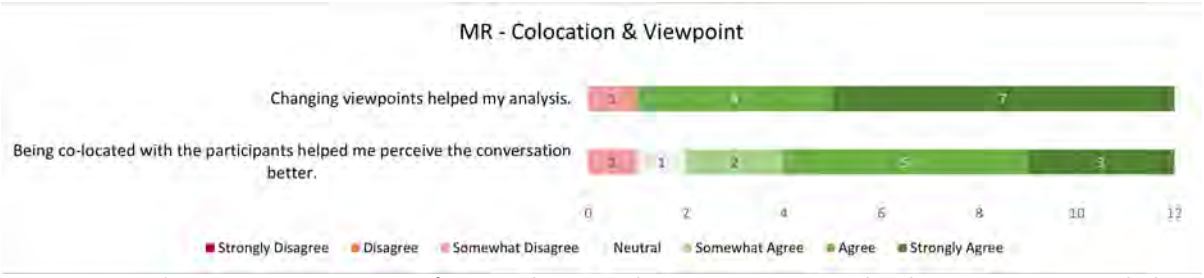


Figure 43: The survey responses of 12 analysts on how viewpoint and colocation impacted their analysis in MR, using a Likert scale ranging from 1 to 7, where 1 represents strongly disagree, and 7 represents strongly agree.

preferred to stand behind one of the participants, whereas, in the VR environment, most users preferred to stand at the center back of the room. The reasons behind these preferences were attributed to the difference in the field of view of the devices used in each environment. The HoloLens2 has a smaller field of view of 52 degrees which made it necessary for analysts to look around more to gather information from the environment. Therefore, standing behind an avatar at an angle gave them a better view of both participants and the screen. However, the Quest2 had a larger field of view of 89 degrees, which allowed analysts to comfortably stand at the center back of the room and still have a view of both participants, their respective speech bubbles, and the screen. This insight can be helpful in designing future MR and VR environments and selecting appropriate devices based on the intended use case. Figure 43 and Figure 44 show how helpful colocation and changing viewpoints are for their analysis.

3.4.8 Issues Encountered

Overheating of HoloLens2 Device- If the HoloLens2 device was actively being used to run the application for over 30 minutes (in cases where the users took more than the allotted 15

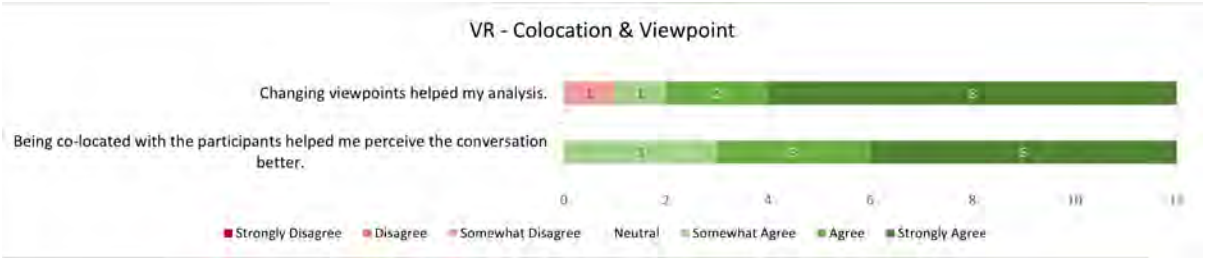


Figure 44: The survey responses of 12 analysts how viewpoint and colocation impacted their analysis in VR, using a Likert scale ranging from 1 to 7, where 1 represents strongly disagree, and 7 represents strongly agree.

minutes for training and then continued with the testing part), the device would display a warning message asking to shut down immediately due to overheating of the device. For a few of the studies we had to stop the study, shut down the device, and wait for about 5 minutes before we could restart the device and get back on track. Each time the device was stopped and restarted the times were noted and we only report the times used to perform the task.

Tracking data - We use only one human model to represent both participants in the conversation. It was hard to depict analysts of different heights with one model. Hence we do see some discrepancies between the avatars head positions which were observed by all analysts. We intend to fix this in further iterations of the application.

Google Speech-to-Text API - Since the transcription generated by Google speech-to-text API was not accurate, we had to manually listen to the audio and correct the transcriptions. Another issue was that some of the text snippets generated by the API were too long to be a part of a single text blob for the rising speech bubbles. Hence, the text and the timestamps

also had to be split accordingly. This manual editing of text and timestamps led to a lot of synchronization issues with the video and tracking data.

3.4.9 Lessons Learned

We utilize our experiences from the user study to create a roadmap for enhancing and perfecting future iterations of this project.

To ensure that motion-captured data produces natural-looking poses during simulations and enhances the user experience, it is important to either capture the entire body movement or limit the visualization to the head. Failure to do so may result in unnatural poses during simulations and potentially compromise the user's experience. Figure 45 shows an example of an unnatural pose in the VR environment.

To facilitate navigation and enable users to access different points of interaction in our application, we implemented a word cloud that included the top 20 most frequently occurring words. However, during the user study, we observed that users primarily selected dataset attributes based on the task at hand, indicating that the word cloud may not have been the most effective approach. While this observation provided valuable insight into users' exploration patterns, we believe that alternative methods may be more effective in achieving the same results. For example, one user suggested implementing a search or lookup function to enable users to easily find words of interest from the conversation, which could potentially achieve the same result in a more efficient manner.

Figure 45: Unnatural pose due to capturing only head movements.

3.4.10 Other Observations

While using the Virtual Reality environment one user said they almost wanted to sit on the desk that was close to the display since it seemed like an appropriate distance to view the screen where the visualizations were displayed. Although they quickly became aware of the fact that a physical table may be non-existent, they did ask the PI if a table existed at the location and if they could use it. We do want to clarify here that the user was well aware of the physical objects (or lack thereof) in the exploration environment before wearing the device. This observation is an example of immersiveness afforded by the application to the extent that

it can potentially become dangerous. However, we did have a student volunteer close to the analyst at all times to make sure such accidents were prevented.

Another observation was that even though the analysts knew that the avatars were only simulations of the participants and they were nonexistent in reality, all analysts went around the avatars and never through them to interact with the interactive worldline that was intentionally placed in front of the avatars.

3.4.11 Potential Applications

In their survey, the analysts reported that the proposed system could have numerous practical applications, ranging from crime scene investigations and training on educational platforms such as YouTube to medical classrooms and sports analysis. They thought that such a system could also be utilized in situations where individuals wish to learn more about inaccessible spaces, such as outer space or the deep sea, allowing for a safer and more immersive educational experience. It could also provide opportunities for historical education, allowing individuals to virtually witness speeches from historical figures and quickly detect important words by moving through time. Additionally, the system could facilitate interactive educational experiences with artists, as well as help those with learning disabilities to explore information at their own pace.

In this chapter, we introduced the first version of the MuSA system and provided a detailed explanation of its implementation in both Mixed Reality (MR) and Virtual Reality (VR) modes. We also described how we conducted the user study, including the rationale behind the tasks chosen for participants to perform. This discussion helps to understand the specific objectives and expected outcomes of the study. We explored how analysts utilize space in both MR and

VR environments, investigating whether their spatial usage differed significantly. Our findings indicate that there are no evident differences; instead, space usage tends to be based on the individual preferences of the user.

In the next chapter, we will detail how we conducted an expert evaluation session with experts from fields such as linguistics and communication. This session aimed to collect insights on the usability of the MuSA system for domain experts, helping us assess how well the system meets the specific needs of experts in these areas. The feedback gathered from this evaluation is invaluable, providing essential input that will guide the further development and refinement of the MuSA system.

CHAPTER 4

EXPERT EVALUATION AND DATA COLLECTION PHASE 2

In this chapter, we delve into the execution of an expert evaluation user study designed to assess the usability of the prototype within the fields of linguistics and communication. We begin by discussing the recruitment process, which differed significantly from our standard approach, highlighting the unique strategies we employed to engage relevant experts. Following this, we outline the specific protocol that guided the study, ensuring a structured and effective assessment. We then present the key findings and results, which were interpreted using Thematic Analysis to draw meaningful conclusions from the data gathered. Finally, we address the unmet expectations associated with the MuSA system, discussing areas where the prototype did not fully meet the anticipated outcomes or user needs.

4.1 Expert Evaluation

To re-evaluate MuSA, we conducted an expert evaluation user study using the contextual inquiry method. We chose to employ this approach to create a conducive setting, allowing the expert to carry out analysis and offer critical feedback. This feedback would drive our next phase of experiments. Contextual inquiry is a qualitative research method developed by Hugh Beyer and Karen Holtzblatt [9]. It is applied in human-centered design and user experience (UX) research where the main goal is to understand the context in which users engage with a product or system. It emphasizes the user-centric approach to product development. It helps

to gain a comprehensive understanding of the environment, activities, and conditions in which users engage with a product or system.

We conducted an expert evaluation with individuals with self-reported expertise in conversation analysis and related areas such as but not limited to Content Analysis, Discourse Analysis, Interaction Analysis, and Multimodal Analysis. We used the contextual inquiry approach by giving the participants a conducive environment over Zoom to conduct their analysis and observe and record their work and methods. We presented the participants with multiple versions of an interaction video, conducted interviews and surveys, and recorded their analysis, thoughts, and observations. The data used in the experiment is (a) from a different user study with protocol # 2022-0354 (video, audio, and head and body movements captured through an optitrack system) and (b) from a video recording of a mock/simulated user interaction session at ERF 2068. The original video of the interaction and a video of a computer-generated version of the interaction of MuSA in Unity was used. The study was conducted online over Zoom.

Through an email, the experts were sent a link to the Informed Consent form administered via Google Forms. Additionally, this email consisted of the details of the online Zoom meeting. This email was sent at least 24 hours before the scheduled meeting. If they consented to participate, i.e., after reading the contents of the form if they clicked on "Agree" followed by clicking on "Submit", in the Google form, we followed the rest of the procedure the following day. The Zoom session was recorded.

On the day/time of the scheduled session, the experts were asked to connect to an online Zoom session with the facilitator. They were asked to fill in an online pre-study survey Google

form. The facilitator then explained the purpose of the study and described the procedures to be carried out. They were informed about your rights, and any questions they had, were answered. The expert was asked to observe a 5–7-minute video consisting of 2 seated participants using vlc/windows media player and carry out the analysis. They could choose to view the video more than once based on their analysis needs. They were asked to make relevant notes and observations on the shared Google document if any. The time taken to complete this part was recorded. Next, they were asked to observe the video of an interactive session of the same dataset/video recorded in Unity's Game mode and carry out the analysis. They could again choose to view the video more than once based on their analysis needs. The time taken to complete this part was recorded. The prototype was then shown in Unity - a cross-platform game engine, for any additional questions they had about the prototype. The time taken to complete this part was recorded. They were asked to answer a series of interview questions about their experience. Next, they were asked to make relevant notes and observations on the shared Google document if any. This step was followed by observing a third video with standing and moving participants. They were then asked to answer a series of interview questions about their experience. The last part of the experiment consisted of filling out the post-study survey. The time taken to complete this part was recorded. This study was approved by the IRB (STUDY2023-1074) under the exempt category.

4.1.1 Eligibility Criteria and Number of Participants

Healthy adult subjects who have no self-reported visual or motor impairments and self-report as not prone to motion sickness participated in the study. Any such individual with

self-reported academic or professional experience with Conversation Analysis or related areas such as but not limited to Content Analysis, Discourse Analysis, and Multimodal Analysis who is at least 18 years old could participate in the study. Up to 25 participants could be involved in this research conducted at University of Illinois Chicago (UIC).

4.1.2 Recruitment

For this study, we recruited individuals who are at least 18 years old with self-reported academic or professional experience with Conversation Analysis or related areas such as but not limited to Content Analysis, Discourse Analysis, and Multimodal Analysis. We reached out to points of contact for departments/organizations that may have relevant experience in the Conversation Analysis area such as the departments of communication, colleges of Liberal Arts and Science, and Human-Computer Interaction research groups to help identify potential candidates for the study. Once the points of contact responded with a list of potential candidate/s we then reached out to the potential candidates to understand their interest/ availability for the study via email.

4.1.3 Expert Evaluation Sessions

We recruited 6 experts who were either trained in Conversation Analysis or had conducted research using related methods throughout their careers. Their expertise also lies in various other related domains such as discourse analysis, Multimodal analysis, thematic analysis, content analysis, textual and critical textual analysis, and film analysis. The sessions lasted from 90 minutes to 145 minutes.

4.1.4 Expert Evaluation Key Highlights

Our dataset comprised various forms of data, including survey responses, interviews, field notes, and audio/video recordings. For transcribing our audio recordings, we utilized OpenAI's Whisper model. Although the transcription quality was an improvement over other services we have used in the past, such as Google Speech to Text, it still had issues. The transcriptions from the Whisper model occasionally included errors such as repeating words and phrases multiple times due to the model's tendency to hallucinate. Additionally, there were instances where certain segments of data were not transcribed, resulting in missing information.

These transcription errors required manual corrections. To manage and analyze the corrected transcripts along with the interviews and notes, we used the software Atlas.ti. This tool facilitated the coding process, allowing us to systematically categorize and assess the data collected from our experts. This approach helped us to organize the extensive qualitative data efficiently and supported a more structured analysis.

Key Findings from our analysis:

1. The interface was intuitive and easy to understand. The users did not have any trouble understanding the features of the application or adapting to it to help their analysis process. The prototype enabled a seamless viewing experience.
2. The immersiveness of the prototype helps in understanding body language.
3. The Line of sight can help understand the focus of the participant or their lack thereof in any part of the conversation. This can be beneficial in highlighting the aspects or content of the conversation that the users were interested in or vice-versa. It also helps

in identifying the fixation of users on particular topics or media in the conversation. The inaccuracies can be distracting at times.

4. Prototype helps conduct an in-depth analysis of the content as it offers the viewer various perspectives of the same dataset.
5. Helpful in understanding the aspects of conversation dynamics like backchanneling and turn-taking.
6. Immersiveness creates a more visceral/engaging experience and helps more retention of the conversation.
7. Some areas where MuSA would be applicable/useful other than Conversation and Multimodal Analysis are market research, criminal justice, ethnography, media research, observational research, and aspects related to psychology.

4.2 Thematic Analysis

Through our interviews and surveys we were able to gather a lot of qualitative feedback that helped in understanding how MuSA could be useful for Multimodal Analysis.

4.2.1 Mobility & Positionality

The experts believed that analyzing how individuals position themselves in a conversation could reveal insights into the underlying power dynamics at play. This concept is further elaborated on through the thoughts of Expert 4, who mentioned the significance of positionality in understanding these dynamics, as well as its relevance to Communication Accommodation Theory. This theory examines how individuals adjust their communication styles to either

converge with or diverge from their conversation partners. According to Expert 4, the decision to align or distance oneself from the conversational style of others is deeply influenced by the power relationships between the participants. These insights suggest that observing positional cues and adjustments in a conversational style can offer valuable clues about who holds influence within the interaction, how it is exercised, and how it shapes the communication flow and interpersonal dynamics.

> *"..so the power dynamics and communication accommodation theory too. So those instances of moving towards the conversational style of the conversation partner versus against or diverging from a lot of that has to do with the positionality or power dynamics of the players themselves."* (e4)

The experts highlighted a distinct advantage of the immersive environment over traditional video analysis: the capacity to explore multiple perspectives and viewpoints at will, rather than being confined to a single or limited set of perspectives typically available through video. This multifaceted view offered by the immersive prototype allows for a deeper and more nuanced analysis of non-verbal cues, such as hand gestures and body language, which might signal discomfort or other emotional states. Such details, as noted, are often more perceptible within the prototype's environment, where the observer can freely navigate and change viewpoints. In contrast, these subtle cues might be easily missed or obscured in a conventional video setup, where the observer is restricted to the angles and moments captured during recording. This flexibility to examine the interaction from various angles in the virtual space significantly en-

riches the understanding of social dynamics and enhances the detection of nuanced behaviors and expressions.

> *.. are the hands, the gestures signaling some kind of discomfort? I mean, again, those are things that you can get more from the prototype than, .. it's easily overlooked, or hidden if you don't have that prototype, right? You're only looking at one viewpoint, one perspective, but in a virtual space, you can kind of navigate around that.* (e4)

4.2.2 Communication Accomodation and Sensemaking

MuSA potentially helps in examining the nuances of gaze and perhaps more intuitive way to analyzing how people adjust their communication styles to match or differ from those around them. This approach, grounded in the observation of communication accommodation, leverages subtle cues available through gaze—such as shifts in attention, engagement levels, and the dynamics of convergence (where individuals adapt their communication style to become more similar) and divergence (where they accentuate differences). By focusing on eye contact, one can glean insights into underlying emotions, intentions, and the relational dynamics at play, including the ability to detect nuances like interruptions, tone of voice, and signs of emotional states such as frustration.

> *"So the more information that I can have about instances of communication convergence and divergence, and that includes things like talking over someone's vocal tone. Obviously, it goes into consideration. You can pinpoint evidence of frustra-*

tion and that sort of thing. I would probably be more likely to look at the second scenario or look at the tools of the second versus the first in that vein, the physical body language"(e2)

Drawing from the detailed observations shared, the statement emphasizes how the utilization of gaze in the design of a prototype can significantly enhance the user's ability to absorb and retain information without disrupting their engagement or flow of experience. Specifically, the ability to maintain a continuous interaction without the necessity to pause, backtrack, or disengage is highlighted as a key benefit. The firsthand experience of the expert, who notices a marked improvement in their own information retention and understanding while interacting with the prototype, underscores the effectiveness of gaze as a tool for bridging information gaps in a way that feels natural and uninterrupted.

"..in your prototype,.. the first thing I realize is I'm not having to stop and go back. Like even now that we're having a conversation a lot of times when I'm saying things and looking here and looking there and doing this.. But the fact that I'm able to see it (gaze) makes a difference in me even understanding and retaining that information." (e3)

The immersive environment provides a unique insight into the participants' engagement with data by allowing observers to discern where their visual attention is directed. This capability to track visual focus offers clues to how participants process and interpret the data they are exploring. Although it's acknowledged that this method doesn't lead to definitive conclusions

on its own, it nonetheless enables the drawing of correlations between the focus of attention and the cognitive processes of meaning-making. By observing where participants look, how long they gaze at certain data points or areas, and how their focus shifts over time, researchers can infer aspects of the participants' thought processes and how they go about understanding or making sense of the information before them. This approach underscores the value of the immersive environment in enhancing the depth of analysis regarding participants' interactions with data.

> *"So in the second scenario, you're able to kind of maybe understand a little bit about ..where they're visually focused. Now obviously you can't come to a lot of conclusions, but you can, you know, make some correlations about that focus and maybe what the meaning making process is."*(e2)

The experts appreciated the prototype's capability for creating an environment that facilitates deeper analysis by providing analysts with the advantage of synced time and a conducive space for reflection. This setting allows for an enhanced exploration of the participants' sense-making process. Specifically, the ability to pause and reflect within the virtual space offers analysts the chance to closely examine not just the overt actions of participants, such as where they are looking or what they are reading, but also the subtler aspects of their engagement, like how they pause to absorb information. While direct observation of certain nuances, like exact gaze points or facial expressions, may not always be possible, the tool compensates by affording analysts the opportunity to slow down, observe closely, and reflect on the behavior and thought processes of participants. This pause for thought is instrumental in achieving a more nuanced

understanding of how participants interact with and make sense of the data presented to them, thus enhancing the overall analytical process.

> *"the prototype allows, ..the person who's analyzing to have ..more time.. and more space for thought, right? So, in that virtual space, you can kind of zoom in, you may not be able to see what is.. real in terms of how the person is looking exactly, what they're looking at, ..their facial expressions. However, it gives you a moment to just kind of, .. stop and pause.. I did that a couple of times, just stop and pause to think about where they were looking, what they were reading, and how they were pausing to get information... So, again, I think that the tool enhances just in terms of being able to have, .. a different sense.., or enhancing another sense.."* (e4)

4.2.3 Bridging the Distance in Conversation

The ability to closely approach participant avatars within the prototype was seen as a significant benefit, particularly in terms of gaining insights into the specific information participants were engaging with. This feature of the prototype enables an observer to zoom in on and directly observe the content being viewed by participants. Such an approach contrasts sharply with the limitations encountered in traditional video analysis, where the observer is often left guessing about the details of what participants are looking at or whether their verbal responses accurately reflect the visual data presented on the screen. The enhanced observational capability provided by the prototype not only allows for a more detailed examination of participant interaction with content but also aids in verifying the congruence between what participants

say and what they are actually viewing, thereby offering a more comprehensive understanding of their engagement and responses.

> *"I think also in the prototype, you also gain an understanding of being able to zoom in and look at that information that they're looking at, whereas in the first video, I could not zoom in and I was wondering, you know, what was on the screen and if what they were saying also matched what was on the screen as well."* (e4)

4.2.4 Deciphering Conversation through Body Language, Intonation & Diction

One expert expressed confidence in the prototype's ability to offer insights into the demonstration of trust within conversations, particularly in interactions involving artificial intelligence (Arti from the first Dataset). By analyzing the second video, observers can discern where participants' attention is directed—whether it is towards the accuracy of the information presented or specific keywords. This level of observation provides a deeper understanding of the human behavior associated with accepting information from artificial intelligence without skepticism. The expert believes that observing what participants choose to focus on, as well as what they disregard, can reveal much about their inclination to trust the information being conveyed by AI. This approach underscores the prototype's utility in exploring not just the content of conversations but the underlying psychological dynamics, such as trust, that influence how information is received and processed.

> *"And in the second video, you're able to kind of see where their focus is and if they're focused on things like the accuracy of the information or if there's focus on*

the keywords, right? And so I think the second one would give me personally more information on just the human behavior behind kind of trusting what the artificial intelligence is telling you without questioning it, right? And that could be gleaned from where they're looking what they're not looking at, right?" (e2)

They also mentioned that being able to closely observe characters could aid in understanding the dynamics of turn-taking in conversation. By zooming in, viewers can see the subtle head movements and shifts in gaze between speakers, which indicate when individuals take turns speaking. This enhanced perspective could provide a clearer understanding of the nonverbal cues that govern conversational exchanges.

"Because you can zoom in, you can see how faces turn to look at each other when they speak, right? They're doing turn-taking." (e4)

Lack of focus: Researchers often face challenges while capturing the full spectrum of communication during studies and interviews, particularly the nuances of body language and non-verbal cues, which are often as telling as the spoken word. Given the difficulty of noting every detail in real-time, a tool like MuSA presents a valuable solution, enabling researchers to revisit and reconstruct these interactions. This retrospective analysis can uncover aspects of the conversation that may have been overlooked or lost, allowing for a deeper dive into the layers of non-verbal communication. The following quote underscores the added dimension that MuSA offers, not only in recapturing the verbal exchange but in enriching the data with insights into the participants' unspoken thoughts and feelings. This might indicate distraction, engagement,

or a connection with certain individuals or objects in the environment, researchers can gain a fuller understanding of the subjects' attitudes and emotions towards the topics discussed.

> *"And so to be able to go back to it have it recreated and then have these extra non-verbal cues that you can hone it on can I think add to the richness of the information rate that that has been presented. So I mean not just the conversation but maybe what the person maybe thinking at that point of time, I mean the gaze somewhere could be suggesting that maybe was there a lack of focus.. or they were connecting with some people in the room or some aspects in the room and not anywhere else so could that have something to do with you know how they also feel about that issue which they're not able to verbalize or put into conversation."* (e3)

4.2.5 Proxemics & Physicality

MuSA could potentially provide insightful observations on Proxemics, and the effects that population density has on behavior, communication, and social interaction. Specifically, it suggests that by analyzing how participants interact with one another within a given space, a wealth of information can be uncovered about their social and spatial relationships. This includes insights into how individuals position themselves relative to each other, the distances they maintain, and how these factors influence their interactions and communication.

> *"So I can glean a lot about Proxemics. The utilization of space they are interacting with each other. You can glean a lot of information just from each other and haptic relationship"* (e2)

The essence of MuSA's design is to allow an observer to enter the conversational space virtually—giving them the ability to get close to the interaction, to observe and analyze participants closely, almost as if they were an invisible third party. This capability is particularly valuable in settings where understanding the nuances of communication and interaction is crucial, such as in usability testing, psychological research, or immersive experiences where maintaining the natural flow of conversation is essential for data integrity or user experience.

The prototype seems to have been successful in fostering a sense of presence within the conversation for the observer, without the negative connotation of being intrusive or altering the dynamic of the conversation being observed. It suggests a balance was struck, where the observer could "intrude upon a conversation" in the sense of becoming closely involved and making detailed observations, yet do so in a way that is not perceived as interfering or interruptive by the participants. This delicate balance enhances the ability to gather insights and understand interactions in a more natural and unaltered state, thereby providing richer, more authentic data or experiences.

> *"Um, the prototype, I mean, it was really cool. I think just, .., you feel like a third person in there that, that where you can intrude upon a conversation, um, without, you know, really being intrusive, um, and to make observations, right, you can go in that space, you can look at that person."* (e4)

> *"walking around is great because that gives you a sense of like you are physically with them."* (e5)

One expert in media and film analysis expressed a desire for the ability to immerse themselves within a film scene to gain a comprehensive understanding of a character's body language from multiple viewpoints. The expert highlighted that observing a scene as if physically present could uncover nuances typically missed in conventional 2D viewing, such as gestures or expressions visible only from certain angles within the room. This capability, they noted, would significantly enhance their ability to analyze characters' behaviors and interactions, providing deeper insights that are not evident when viewing scenes from a fixed perspective.

> *"I do a lot of like media analysis, like, ..if it was a scene from a film .. it would be so cool to be able to step in and pick up on body language, or pick up on a different perspective from the room, .. that if I was watching this, .. in 2D, ..I just would never see because I wasn't on the other side of the room. So, like, if it allowed me things like that, ..if I had, like, participants who I was studying, um, that would be really cool."* (e6)

4.2.6 Unmet Expectations of the MuSA

In typical face-to-face conversations, expressions can signal confusion, concern, or the anticipation of a response, thereby naturally guiding the flow of dialogue and indicating when it's appropriate for someone else to speak or respond. These non-verbal cues, as highlighted in the quote below, are essential for smooth and intuitive communication, suggesting when an individual expects a reply or is seeking further clarification.

In the context of the prototype's evaluation, while it was noted that valuable data could be collected through monitoring the direction of head movements and where a person's gaze

was focused, there was a significant shortfall in capturing the full spectrum of non-verbal communication. The absence of this capability meant that the nuanced dynamics of conversation, such as recognizing when someone is puzzled or wishes to continue the discussion, could not be adequately detected or interpreted. This limitation hindered the prototype's effectiveness in facilitating natural interactions, as it lacked the ability to replicate the rich, facially driven cues that play a critical role in human communication.

> *" Whereas in the prototype, you have to really listen for that. Um, so you can see, for example, in the face facial expressions in the first video, um, when they are looking a little concerned or confused and they want to continue with the conversation or they are expecting the other person to pick up the conversation. You can't really get that from the prototype, um, just the facial expressions kind of stuff, um, that signals when it is the next person's turn to speak or, or they're seeking answers from the next person or the other person"* (e4)

The analysts appreciated having access to gaze information and the line of sight, indicating they found these features to be quite intriguing. They valued the ability to understand where the subject was looking at any given time. However, despite this interest, they also experienced moments when these visual elements became overwhelming or distracting. This distraction arose from the continuous movement of the lines indicating the gaze direction or when it wasn't immediately clear why the subject's gaze was focused on particular areas, such as the floor. The analysts expressed a desire for more control over this feature, suggesting that it would be beneficial to have the option to toggle the gaze information and line of sight on and off as

needed. This capability would allow them to minimize distractions when necessary, making the overall experience more user-friendly and tailored to their preferences at any given moment.

> *"..the line of sight was super interesting. And maybe you can like click on or off to see that. But sometimes a little distracting too, just seeing like the little lines moving all the time, or I'd be like, why is it staring at the floor? .. So I just find that a little, distracting. "* (e6)

Table II provides an overview of the themes that emerged during our analysis of the evaluations.

In this chapter, we covered the contextual inquiry method and conducted interviews with experts in linguistics and communication to assess the usability and potential adoption of the MuSA system in these fields. The primary findings reveal that experts particularly value MuSA's ability to capture and analyze non-verbal cues and body language. They also appreciated the system's immersiveness, which allows users to immerse, engage, and become a part of the conversation.

In the next chapter, we will explore how we integrated feedback from the initial user study and the expert evaluation to enhance the MuSA prototype. We will then detail the subsequent user study, presenting the methods used and the results obtained.

Mobility & Positionality	**Communication Accommodation And Sensemaking**	**Bridging Distance in Conversation**	**Non-verbal Cues**	**Proxemics and Physicality**	**Limitations**
"positionality or **power dynamics** of the players themselves"; "in a virtual space, you can kind of navigate around that" (e4)	"You can pinpoint evidence of frustration"; "you can, **make some correlations about that focus** and maybe what the meaning making process is." (e2)	"also gain an understanding of being able to **zoom** in and look at that information that they're looking at," (e4)	"you're able to kind of see where their focus is and if they're focused on things like the accuracy of the information or if there's focus on the keywords, right?" (e2)	"you feel like a third person in there" (e4)	"for example, in the face **facial expressions** in the first video, um, when they are looking a little concerned or confused .. You can't really get that from the prototype" (e4)
	"But the fact that I'm able to see it (gaze) makes a difference in me even understanding and retaining that information." (e3)		"you can zoom in, you can see how faces turn to look at each other when they speak, right? They're doing **turn-taking**." (e4)	"walking around is great because that gives you a sense of like you are **physically with them**." (e5)	
	"think that the tool enhances just in terms of being able to have,.. a different sense.., or enhancing another sense" (e4)			"it would be so cool to be able to step in and pick up on body language, or pick up on a different perspective from the room" (e6)	"the line of sight was super interesting. And maybe you can like click on or off to see that. But sometimes **a little distracting** too, just seeing like the little lines moving all the time" (e6)

TABLE II: Evaluation Quotes Categorized by Theme

CHAPTER 5

USER EVALUATION PHASE II

This chapter explores the methods and adjustments implemented to tackle the primary challenges identified in our initial user study. We begin by revisiting and refining our research questions to better align with the insights gained from our expert evaluation. Following this, we discuss the strategies employed for participant recruitment and describe the data collection process carried out in phase II of our study. The chapter further delves into the two distinct parts of the study: the conventional approach and the exploration through the MuSA system. Finally, we detail the enhancements made to the prototype based on the feedback and findings from these studies.

Our initial evaluation indicated that an application such as MuSA shows considerable potential for analysis of multimodal meetings. We observed different strategies used by the analysts to arrive at answers, noting they managed to complete most tasks and also saw potential uses for the technology. Some analysts appreciated the see-through capability of MR, though they were concerned about its limited Field of View (FOV). Conversely, others favored VR's wider FOV and immersive experience but felt less in control and less confident while navigating the environment. Given the mixed preferences among analysts and the ongoing advancements in MR technology, we decided to focus solely on the MR environment for the second phase.

We conducted a within-subjects study in which analysts used a traditional mode for the first scenario and an immersive mode for the second scenario to analyze multimodal meetings. It took about 30 minutes to complete tasks in both scenarios.

5.1 Research Questions

Through the feedback and lessons learned from our first user study, we further refined our questions to understand more complex nuances of multimodal analysis.

1. RQ1: Can Multimodal Situated Analytics in XR support analysis of conversations?
   (a) Does it help in understanding interest and engagement in conversations?
   (b) Does mobility, positionality, and physicality help in understanding and sensemaking of the data?
2. RQ2: Does immersiveness support exploring non-verbal cues such as gaze, silence, pause, and head movements in conversations?
3. RQ3: What strategies does the analyst employ to analyze data and user behavior?
4. RQ4: Do moving participants force the analyst to move or choose the best viewpoint for analysis?

5.2 Participants (Analysts)

We recruited 13 students from the university, all with backgrounds in Information Technology or Computer Science, including four females and nine males. Five of these users had previous experience with the HoloLens2 device. The age distribution was six subjects between

18-25 years, six between 26-35 years, and one aged between 36 to 45 years. The user study sessions took place in the same room used for data collection.

5.3 Data

The dataset for the second phase of our study comprised observations from participants, both standing and moving, as they interacted with a 3D asset accessed via a QR code in the Continuum room. This phase included 59 participants from another study (#STUDY2023-0509-MOD001). However, each session lasted only between 5 and 10 minutes, which was insufficient to gather enough interactions and data points for our prototype's analysis needs. Consequently, we organized a simulated data collection session. In this session, we recorded two participants—both standing and moving—as they interacted with the 3D asset through their smartphones. We captured their audio, video, and head and body movements to enrich our dataset.

5.3.1 Content Details

All the content for this study came from another protocol, #STUDY2023-0509-MOD001 (PI - Tanja Aitamurto). As a part of the protocol, the articles for the Allensworth buildings were written by journalist and educator Lakshmi Sarah and the 3D models were generated by creative technologist and consultant Ben Kreimer. The images for the 3D models were captured using a DJI Mavic 2 Pro drone and the DJI Ground Station Pro mobile app. In addition to the drone photographs, a Lumix G7 camera was also used to capture images of the structures. The images captured for this project, and all photogrammetry work, have at least 70% overlap with one another, which enables photogrammetry software to analyze and process the images

into 3D models. The images were processed using Capturing Reality's RealityCapture software. The largest model, of the park as a whole, was processed from 2678 images. The School model was processed from 1603 images. The Allensworth House was processed from 1584 images. The Library was produced from 1314 images. The 3D content is presented to the viewer through an AR content creation platform called ZapWorks[45].

5.3.2 Data Collection

During our data collection phase, two participants were recorded as they explored three historical buildings from the 20th century. These structures were located within the Colonel Allensworth State Historic Park in Tulare County, California. The participants' task was to engage in a detailed comparison of these buildings, using a variety of visual aids. They were provided with a series of images showing each building from five distinct perspectives: the front, right, left, back, and top views. These multi-angled views aimed to offer a comprehensive visual understanding and to facilitate a comprehensive comparison.

To enrich their exploration, the participants were also given articles that delved into the history and significance of the Colonel Allensworth State Historic Park and its buildings. This textual information served to complement the visual imagery, providing a deeper context for the participants' discussions. Additionally, they had access to a map that pinpointed the park's location within the United States.

As part of the study, a document containing six QR codes was included, each corresponding to one of the three buildings. When a participant scanned a QR code using their smartphone

camera, it launched an application link. This link prepared the application for the subsequent step, which involved the generation of a 3D model.

An anchor image had been placed on a table in the center of the room for the AR experience. Scanning this anchor image with their smartphone cameras prompted the generation of an AR 3D model of the building in question, effectively anchoring it in the physical space of the room. This intuitive approach allowed the participants to walk around and interact with the virtual models as if they were tangible objects.

For each building, the participants examined two different AR model types. The first type was a textured mesh model that rendered the structures with detailed textures, providing a realistic representation of the building surfaces. The second type was a point-cloud model, which depicted the buildings with 1 million points, offering a different, more abstract interpretation of their forms.

Participants took about 28 minutes to complete their exploration of the three buildings. This duration reflects the level of engagement with the models as they thoroughly explored and interacted with the mesh and point-cloud versions of each building. Through this exercise, the participants were able not only to see but also to physically maneuver around the models, gaining a tangible sense of the architecture and the space it occupies. Figure 46 shows two non-seated participants engaged in exploration.

5.4 Scenario 1 - Conventional Approach (S1)

In Scenario 1, the analysts were provided with a video of a conversation on one laptop, the transcription of the conversation on another laptop, and a few questions about the video on a

Figure 46: Two participants engaged in the exploration of 3D building models through their phone during the data collection of Phase II

spreadsheet. The contents on the display wall remained the same as seen during the data collection phase. They were asked to find instances of user engagement, agreement, disagreement, and intonation and report their timestamps on the spreadsheet. They were also asked to find a building the users were least interested in. While the analysts successfully completed most tasks, they found it challenging and hard to interpret the data. They took about 30 minutes to complete the activity. A primary difficulty was their limited ability to only view 2D images of the buildings, rather than the building models the participants were exploring. Figure 47 shows the study setup for S1.

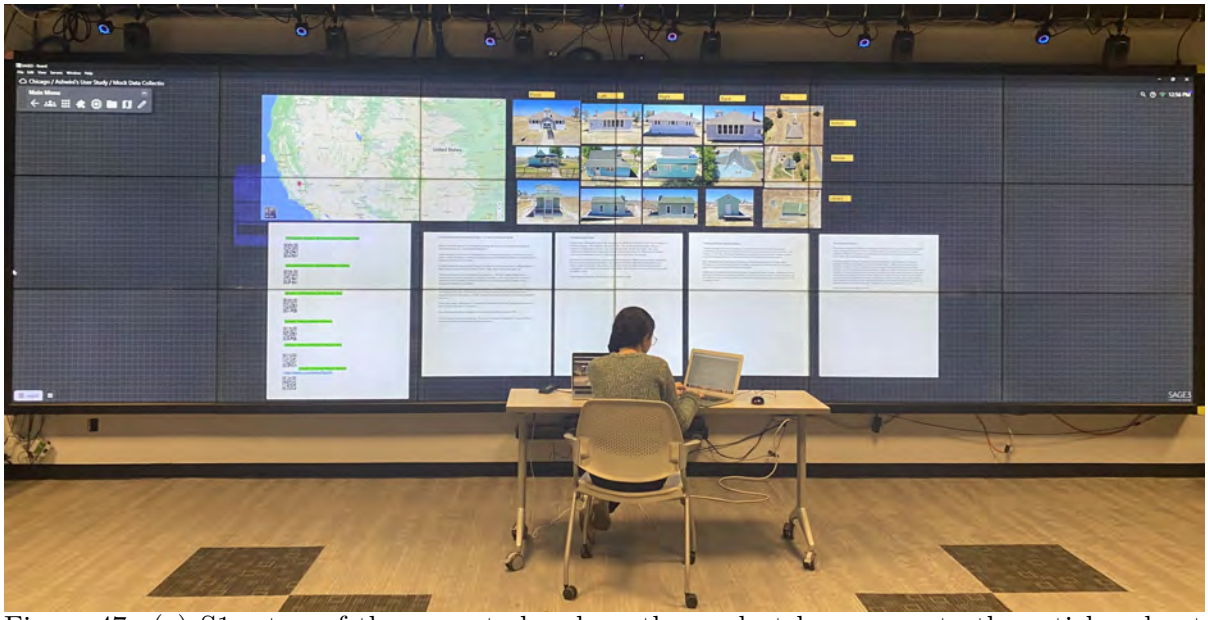


Figure 47: (a) S1 setup of the user study where the analyst has access to the articles about the buildings, building images, a map showing where the buildings are located, video, and transcription of the conversation

5.5 Scenario 2 - MuSA (S2)

The analysts were explained about the setup and interactions they would be using in the HoloLens2 application.

They were given 15 minutes to train on the device and the application. We used an identical version of the application in training and testing sessions except for the task list in each of them. The questions in the training session were mainly focused on getting the users comfortable with the device and the application.

Similar to Scenario 1, the test session contained questions about logging instances of engagement, agreement, disagreement, and changes in intonation and gaze that helped in analysis

of the conversation. However, they were provided with an in-application menu to log their answers conveniently. They were also asked to observe and report which building the participants in the conversation enjoyed exploring the most. Analysts were initially tasked with selecting three different spots in the room to observe the conversation from and then choosing the one they felt most comfortable in. This step was prioritized to identify the optimal viewing angle for a seamless analysis experience, considering the limited field of view (FOV) provided by the HoloLens2.

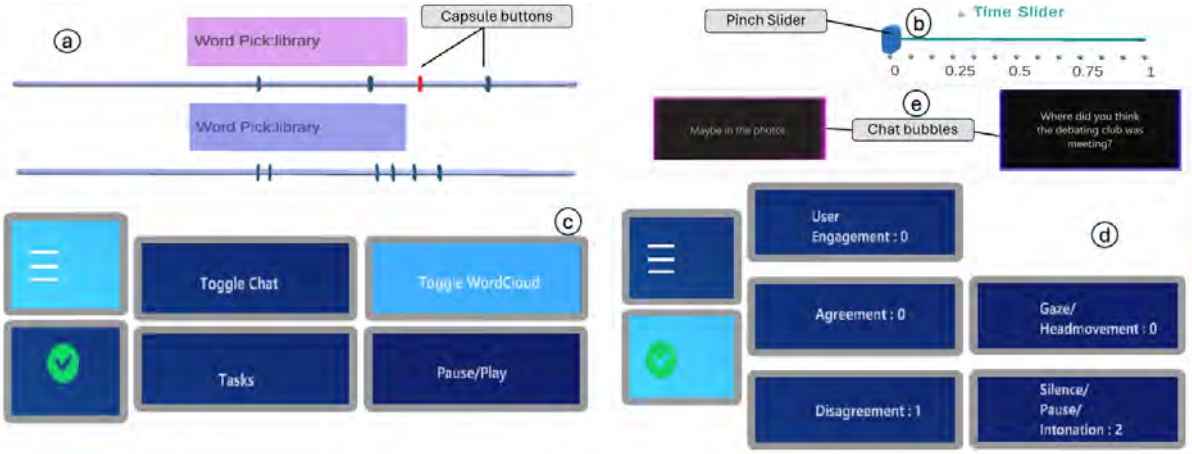


Figure 48: Updated controls of the MuSA interface is composed of several interactive components designed to enhance meeting analysis: (a) 'Wordline,' a selection bar for conversation keywords, which is populated from a word cloud and uses color-coding to represent different participants. (b) 'TimeSlider' is a dynamic control for moving through the conversation timeline. (c) A 'Main Menu' offering a variety of general options for customization and control. (d) An 'Answer Menu' where occurrences of verbal and non-verbal cues during the meeting are meticulously logged. (e) Samples of chat bubbles of varying lengths also employing the color-coding system to denote different speakers.

5.6 System Enhancements

In evolving our prototype to better suit the needs of participants interacting with 3D building models through augmented reality (AR), we introduced several modifications. The changes were driven by our experiences from Phase I, particularly the feedback on the awkwardness caused by full-body avatars [74]. Thus, we simplified the avatars to include just the head and neck, utilizing body tracker positions, with the aid of a lab coat for body movement and orientation and head tracking for head orientation. We first chose two different avatars one each to represent the male and female participants. To simplify the avatars, we had to hide/disable the meshes for the following sub-assets - Wolf3D_Body, Wolf3D_Outfit_Top, Wolf3D_Outfit_Bottom, and Wolf3D_Outfit_Footwear. We then added a pink cylinder to the neck of the female avatar and a blue cylinder to the neck of the male avatar to enhance clarity and aid in differentiating between the two avatars during exploration.

We also integrated six new building models for participants to explore within the AR environment, ensuring these were in sync with existing data sets. This addition necessitated adjustments to the user interface layout; we reorganized elements like the word line and the time slider to make space for the building models' display. The wordline is also color-coded to match with avatars.

Furthermore, to capture the nuances of user interaction, we designed an answer logging menu, which allowed for detailed recording of user engagement, agreement, disagreement, silence, pauses, and variations in intonation—all critical elements that contribute to group discussions.

Figure 49: WordCloud in Phase II

Lastly, to enhance the collaborative experience in AR, we implemented translucent, color-coded phone models. These provided users with a 'window' to see what others were viewing on their phones, thereby creating a shared visual context and facilitating a better understanding of each participant's perspective. (Figure 48).

In Phase I of implementing the word cloud, we used Unity's TextMeshPro without incorporating a background for each word. User feedback from the initial study indicated difficulties in distinguishing words that blended into the background. Consequently, we introduced a translucent background for each word in the cloud to improve visibility. The improvements are illustrated in Figure 49.

In Figure 50, we present a comprehensive diagram that outlines the architecture of a software system, detailing its components and their interconnections, with enhancements introduced in Phase II emphasized.

Data Sources: An additional data type has been integrated into the Instantiate class at the diagram's upper section. Specifically, the integration of Phone tracking is responsible for initializing the phone's presence within the system and managing its tracking functionalities. Figure 51 shows an analyst's view of a participant in MuSA exploring a building model through their phone.

Asset Management: We also implemented the 'Asset Changer' module, which has methods such as 'getAssetName' and 'changeARAsset' to manage the changing assets in the environment. These functionalities suggest the application's capability to dynamically modify the virtual content, such as altering the virtual models of buildings displayed to the user, indicating interactivity and personalization within the software. Figure 52 shows an analyst's view of a participant exploring a building model in textured mesh mode.

Answer Menu: The addition of a new component dedicated to managing answer logging is also significant. This module is designed to capture and log various forms of user engagement, including instances of agreement, disagreement, and even subtle non-verbal interactions like 'Gaze/Head Movement,' along with 'Silence/Pause/Intonation'. This suggests a nuanced approach to capturing user responses, providing a rich dataset for understanding user interactions and preferences.

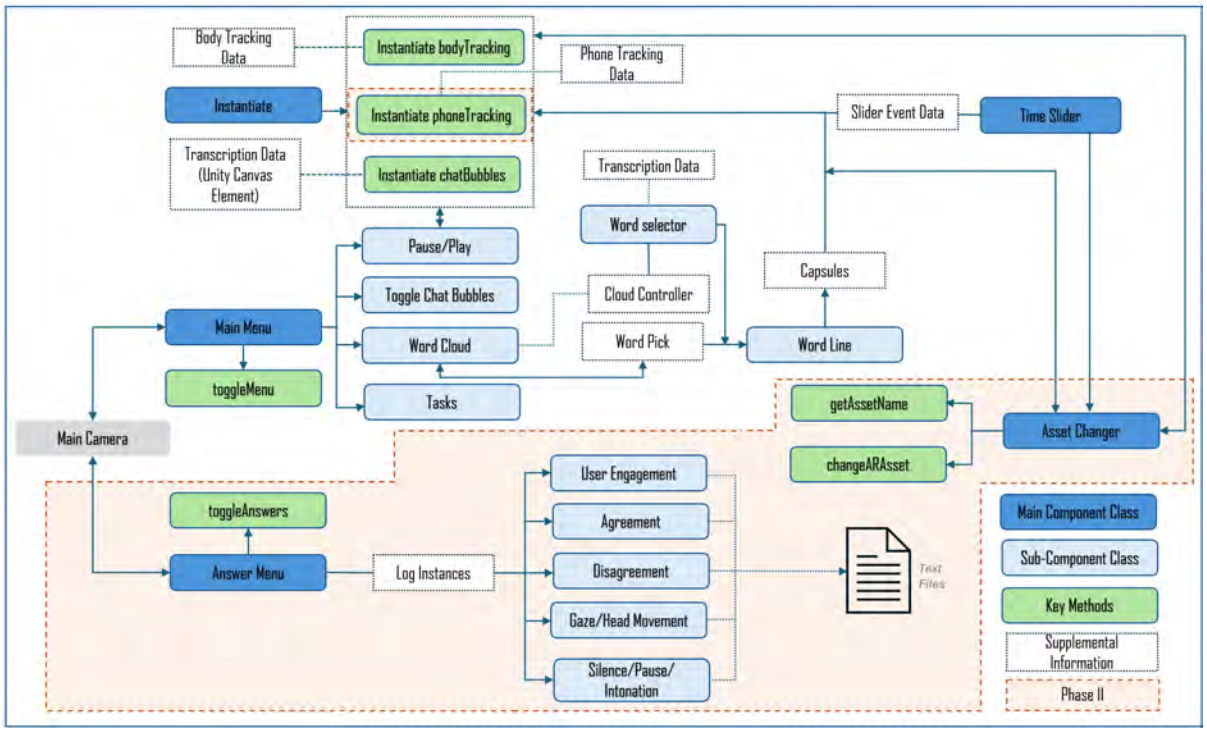


Figure 50: Flowchart of MuSA highlighting its data processing, user interface, and engagement tracking components, including enhancements for Phase II

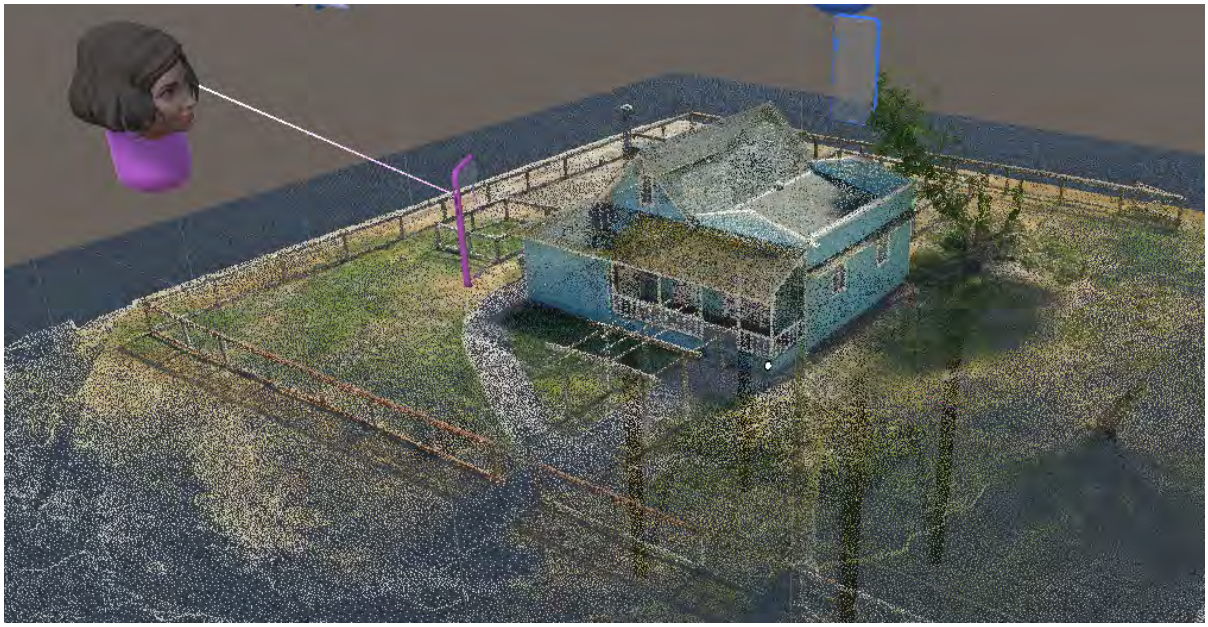


Figure 51: Analyst's view of a participant watching a 1 million point cloud AR model of Allensworth's house through their smartphone

5.7 Results

5.7.1 Chat Bubbles Usefulness

Figure 53(a) shows the usefulness of the chat bubble. The analysts found the bubbles useful for several reasons: they made it easier to tell the two people in the conversation apart, allowed analysts to catch up if they got distracted since the words lingered longer than they were spoken, aided in understanding unfamiliar accents, and helped when words were mumbled or not fully pronounced.

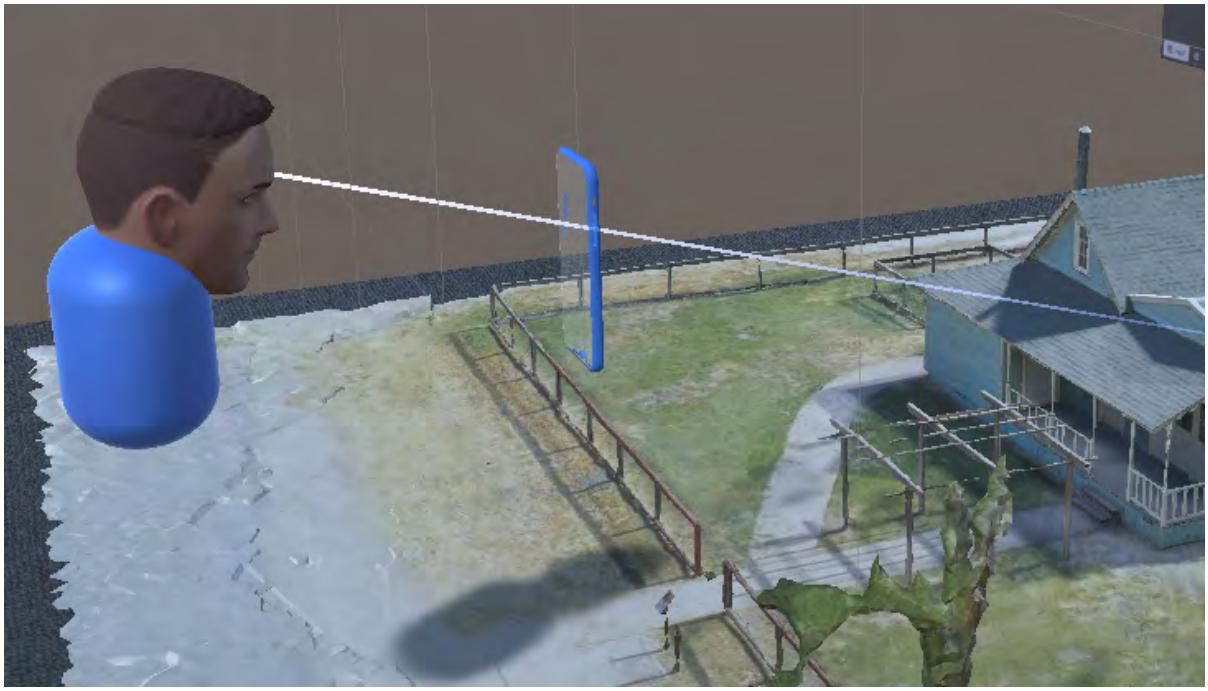


Figure 52: Analyst's view of a participant watching a textured mesh AR model of Allensworth's house through their smartphone

This chart is a visual representation of responses to a survey assessing the usefulness of chat bubbles and the ease of accessing and determining important words and phrases within them. Let's break down the information presented:

For the statement "The chat bubbles were helpful," 3 respondents strongly agreed, 2 agreed, 3 somewhat agreed, 2 were neutral, and 1 somewhat disagreed and two disagreed. Regarding whether "It was easy to access important words and phrases," 7 respondents strongly agreed, 5 agreed, 1 somewhat agreed. When asked if "It was easy to determine important words and phrases," 9 respondents strongly agreed, 1 agreed, and 3 somewhat agreed. From the

distribution of responses, we can infer that the majority of respondents found the chat bubbles to be a helpful tool, particularly when it comes to determining important words and phrases. The strongest positive responses were in relation to determining important words and phrases, which suggests that this aspect of the chat bubbles was most effective.

However, the responses were more mixed when it came to the ease of accessing important words and phrases, with one respondent feeling very negative about it and a noticeable number being neutral. This might indicate that while the chat bubbles were generally well-received, there could be room for improvement in making important content more accessible or prominent.

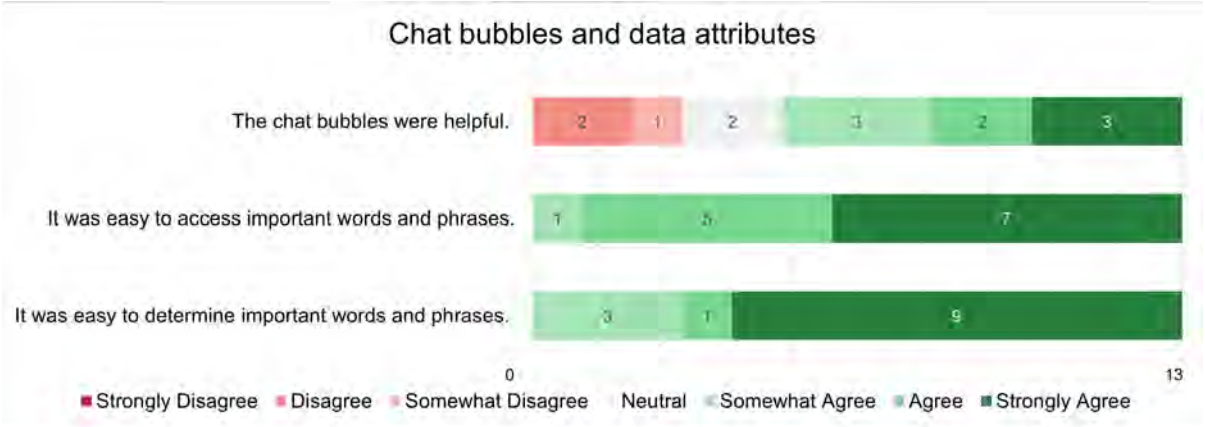


Figure 53: The survey responses of 13 analysts on features provided by the MuSA application, using a Likert scale ranging from 1 to 7, where 1 represents *strongly disagree* and 7 represents *strongly agree*. The aspects evaluated included (a) the accessibility of text, (b) preferences for movement during exploration, and (c) experiences of immersion and colocation.

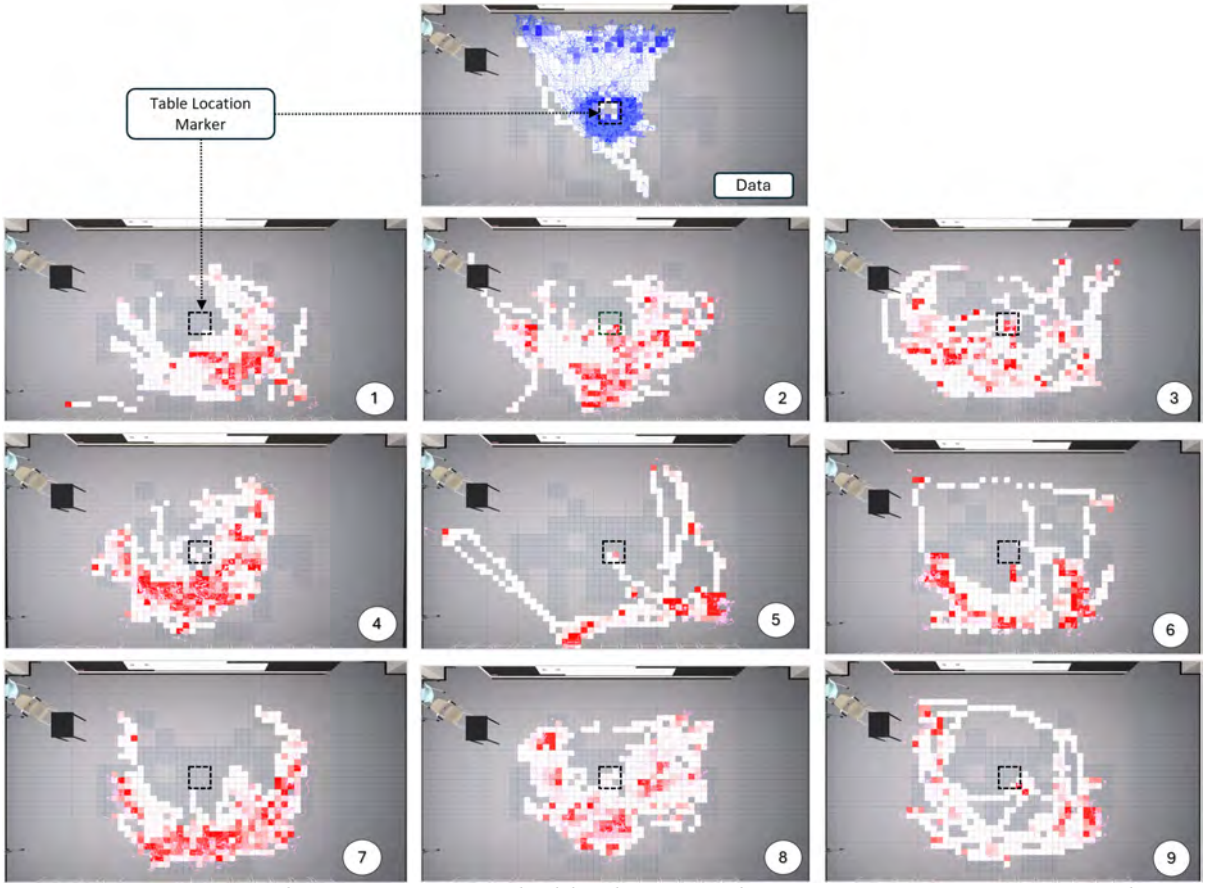


Figure 54: Heatmaps show space usage: the blue heatmap depicts participant movements during data collection, and the nine red heatmaps show spatial movements of 9 analysts.

5.7.2 Interest and Engagement (RQ1 (a))

Our findings illustrate that analysts were able to understand key aspects of engagement, such as interest and whether participants agreed or disagreed, much more effectively than with the traditional method used in Study 1 (S1), where manual input of findings was required. This enhancement is largely attributed to the direct input method enabled by the application,

where participants could use touch gestures to input their responses. This method proved to be highly efficient as it allowed for a smooth integration into the analysts' existing workflow, facilitating a more fluid and uninterrupted analysis of conversations.

Additionally, we assessed the usability of MuSA, our application, by calculating the System Usability Scale (SUS) score [12], which came out to be 78.41%. This score is significant as it reflects a high degree of user satisfaction and usability. It's worth noting that we adapted the standard SUS questionnaire to better suit our needs; instead of evaluating the system itself, we focused on assessing the workflow associated with using the application for multimodal analysis. The adaptability of our approach is evident in the positive response captured by the SUS score.

To visually represent our findings, we included a heatmap of the scores provided by the 13 analysts in our report, illustrated in Figure 55 and Figure 56. This heatmap provides a clear, at-a-glance understanding of how the application performed across various usability metrics, further underscoring the effectiveness of the touch gesture input method and the overall workflow in enhancing the analysis process.

Q. No.		1	2	3	4	5	6	7	8	9	10	11	12	13
1	I think that I would like to use this workflow to analyze conversations frequently.	4	5	4	5	5	1	5	5	5	3	4	4	4
3	I thought the workflow was easy to use.	4	4	4	4	5	1	5	5	5	3	4	5	5
5	I found the various functions in this workflow were well integrated.	5	3	4	3	5	4	5	5	5	4	3	3	4
7	I would imagine that most people would learn to use this workflow very quickly.	4	3	5	5	5	3	5	5	5	5	4	3	5
9	I felt very confident using the workflow.	4	4	5	5	5	1	4	5	5	3	4	5	4

1 Strongly Disagree 2 Disagree 3 Neutral 4 Agree 5 Strongly Agree

Figure 55: Survey responses heatmap for odd-numbered System Usability Scale (SUS) questions from 13 analysts in S2. These responses are rated on a Likert scale from 1 to 5, where 1 represents *Strongly Disagree* and 5 represents *Strongly Agree*.

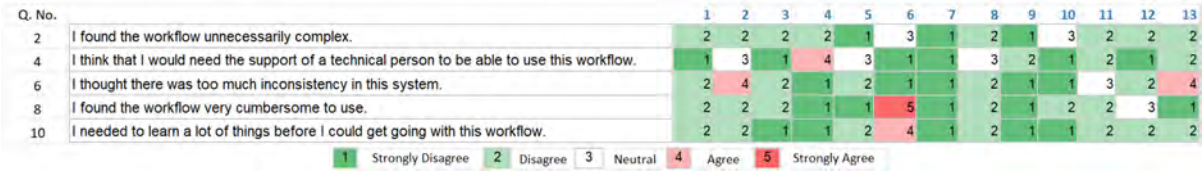


Figure 56: Survey responses heatmap for even-numbered System Usability Scale (SUS) questions from 13 analysts in S2. These responses are rated on a Likert scale from 1 to 5, where 1 represents *Strongly Disagree* and 5 represents *Strongly Agree*.

5.7.3 Gaze (RQ1 (b))

Integrating gaze direction into the discussion about architectural models and images significantly improved how users understood the conversations. It allowed them to see where participants were looking, which enriched the discussion by providing a clearer context of what was being talked about.

One user emphasized the value of this gaze information in revealing the dynamics of interaction between participants. They remarked:

> *"Gaze information made me understand the complementary information shared between the users. For example, one user was providing information from the wall while the other was confirming or challenging the remarks from the 3d scene"* (a3).

This insight highlights how gaze direction can illuminate the process of collaboration, showing how participants share and react to information from different sources, enriching the conversation and understanding of the subject matter.

Another user pointed out the utility of gaze direction in discerning the participants' focus and engagement. They observed:

> *"I could tell when they were getting really close to the model to do the analysis, they would also look at the wall and images so that really helped me with what they were looking at and their point of view"* (a2).

This comment underlines the importance of gaze direction in providing insights into how participants engage with the material, enabling observers to follow the exploratory process and understand the perspectives being presented.

Through gaze direction, users could navigate the complex dynamics of collaborative discussions more effectively, gaining insights into both the content being discussed and the collaborative context. This enhanced not only the comprehension of the architectural subjects but also enriched the interactive and collective experience of the participants.

5.7.4 Mobility & Positionality (RQ1 (b))

Moving around in space helped the analysts get a better visual understanding of the space. They could move around and explore the buildings from different perspectives. They could also get a better understanding of where the participants were by moving in space and staying engaged for a longer period of time.

> *"I can see where the participants are seeing clearly by moving around in space"* (a13).

"being able to move reset my own perception of the scene and conversation, allowing me to stay engaged for a longer period of time" (a1).

These observations help us understand how mobility can help enhance multimodal analysis. Figure 57 (b) illustrates the analysts' preferences for mobility, indicating whether they favored standing in one spot, moving around, or employing a combination of both methods during analysis.

The survey focused on the role of immersion and co-location in collaborative environments. The responses suggest that both elements were generally perceived as positive, enhancing participants' understanding and engagement in the task at hand.

Understanding Silences and Pauses: A majority found that being co-located helped interpret silences and pauses during discussions, with seven strongly agreeing. However, it's noted that there were some mixed feelings, with one respondent disagreeing and one strongly disagreeing.

Gaining Various Perspectives: Proximity to other participants was highly valued, with eight strongly agreeing that it helped them gain different perspectives on the data. This suggests that physical presence and the ability to move closer to others during discussions can enrich understanding.

Aiding the Thinking Process: Immersion was also seen as beneficial for the thinking process, with seven strongly agreeing. The agreement suggests that being physically present in a shared space can positively affect cognitive processes.

Providing Room for Exploration: Similarly, eight strongly agreed that immersion gave them more room for exploration, underscoring the value of an engaging physical space for exploration and discovery.

While the trend in responses was positively skewed towards agreement, indicating a successful integration of these features, the presence of neutral and negative responses points to areas that could be further refined to enhance the experience for all users.

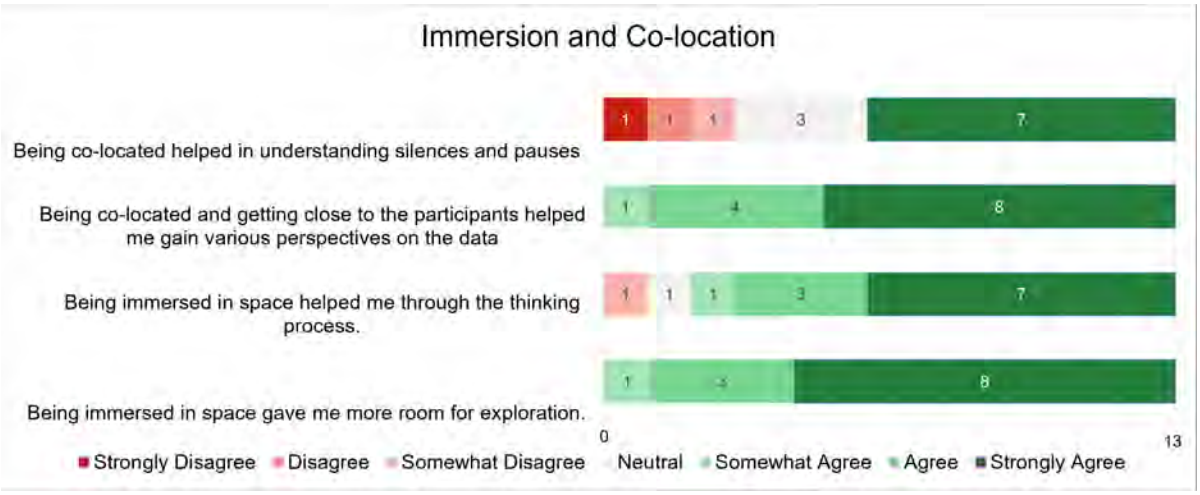


Figure 57: The survey responses of 13 analysts on features provided by the MuSA application, using a Likert scale ranging from 1 to 7, where 1 represents *strongly disagree* and 7 represents *strongly agree*. The aspects evaluated included (a) the accessibility of text, (b) preferences for movement during exploration, and (c) experiences of immersion and colocation.

5.7.5 Immersiveness & Colocation (RQ2)

The analysts reported that being colocated in space helped them to get the participants' point of view and also examine the data at the same time.

> *"Because it gave me a better view point to see what the users were exploring and at the same time I could observe or investigate the same thing"* (a4).

> *Being colocated with the users I was able to notice exactly how they were observing the scene using their phones and what they might have been experiencing from their point of view* (a3).

Being colocated also helped them make sense of the reasons for pause and silence in the conversation.

> *It made it easier for me to assume that they were making these pauses in their conversation because they were critically thinking about what they were looking at. These pauses in conversation were when they were looking at the mesh or looking at the projector presentation*(a9).

Figure 58 (c) shows analysts' feedback about immersion and colocation in the application.

The survey captured participants' preferences regarding their physical approach to conducting analysis, revealing varied inclinations toward mobility, stillness, or a combination of both.

Preference for Moving Around to Conduct Analysis: A total of 6 participants strongly preferred to move around while conducting their analysis, indicating a favor towards

a mobile approach. However, this preference wasn't universal, as 2 participants disagreed, and 1 somewhat disagreed, suggesting some variance in the preferred styles of working.

Preference for Standing in a Particular Position for Analysis: Standing still was strongly preferred by only 2 participants, while 4 agreed but with less intensity. The preference for staying stationary was notably less popular, with 3 participants feeling neutral and 3 strongly disagreeing, highlighting a tendency against remaining fixed in one spot during analysis.

Preference for Using a Mix of Both Approaches: Combining both movement and stillness emerged as the most popular approach, with 7 participants strongly agreeing with this method. This indicates a significant appreciation for the flexibility that a mixed approach provides, allowing analysts to adapt to different tasks and situations during their analysis. Only 1 participant strongly disagreed with a mixed method, reflecting a rare opposition to this flexibility.

In examining preferences for movement, the majority of participants appreciate the freedom to move around during their analysis, but there's also a noticeable preference for a stationary position. The most favored approach, however, is a blend of mobility and stillness, suggesting that flexibility is key in analytical environments. This balance affords participants the adaptability to engage with their work in ways that suit diverse tasks and personal preferences, while a small minority express a clear preference for either extreme—strict mobility or complete stillness.

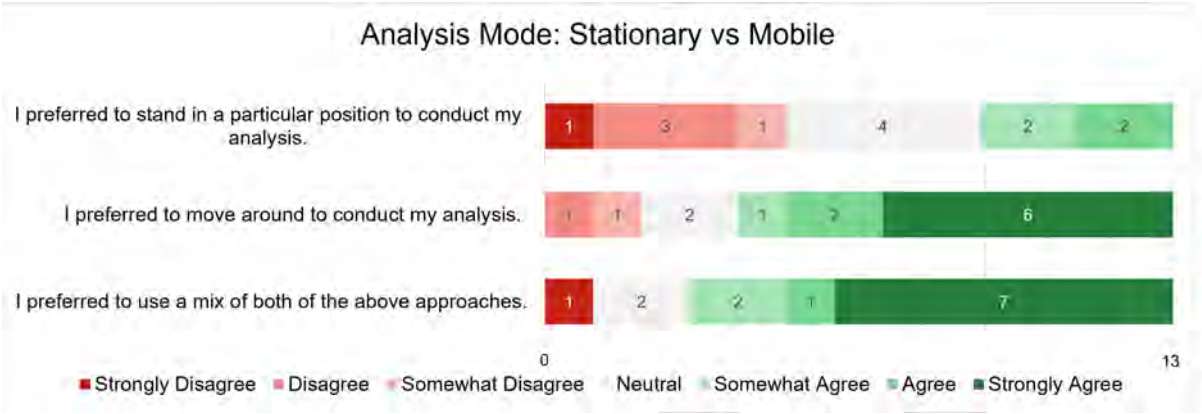


Figure 58: The survey responses of 13 analysts on features provided by the MuSA application, using a Likert scale ranging from 1 to 7, where 1 represents *strongly disagree* and 7 represents *strongly agree*. The aspects evaluated included (a) the accessibility of text, (b) preferences for movement during exploration, and (c) experiences of immersion and colocation.

5.7.6 Workflows Used for task completion (RQ3)

Analysts employed varied methodologies to derive answers to questions, particularly when determining which building was most enjoyed by participants during their explorations. The approaches to answering this question varied significantly among analysts. A few analysts pinpointed the most popular building by analyzing word clouds, where the building's name that appeared in the largest font was deemed the favorite. Another group inferred user preference based on the amount of time spent exploring a particular building, considering longer durations as indicators of higher interest. Meanwhile, a separate group of analysts turned to voice intonation analysis, interpreting variations in tone and pitch as markers of enjoyment, thereby identifying the most appreciated building.

When it came to identifying instances of verbal and non-verbal cues within conversations, analysts again split into two distinct approaches. One group opted to observe the conversations from a distance, allowing them a broad overview from which they could comfortably log answers. This method enabled them to capture verbal cues without intruding on the conversation. Conversely, the second group chose a more immersive approach, positioning themselves close to the avatars and building models. This proximity allowed them to closely observe and log details related to gaze direction and body movement, providing insights into non-verbal communication cues. Through these diverse strategies, analysts were able to gather comprehensive data on user preferences and interactions within the virtual environment. Figure 59 shows the time taken to complete tasks between the conventional part S1 and MUSA S2.

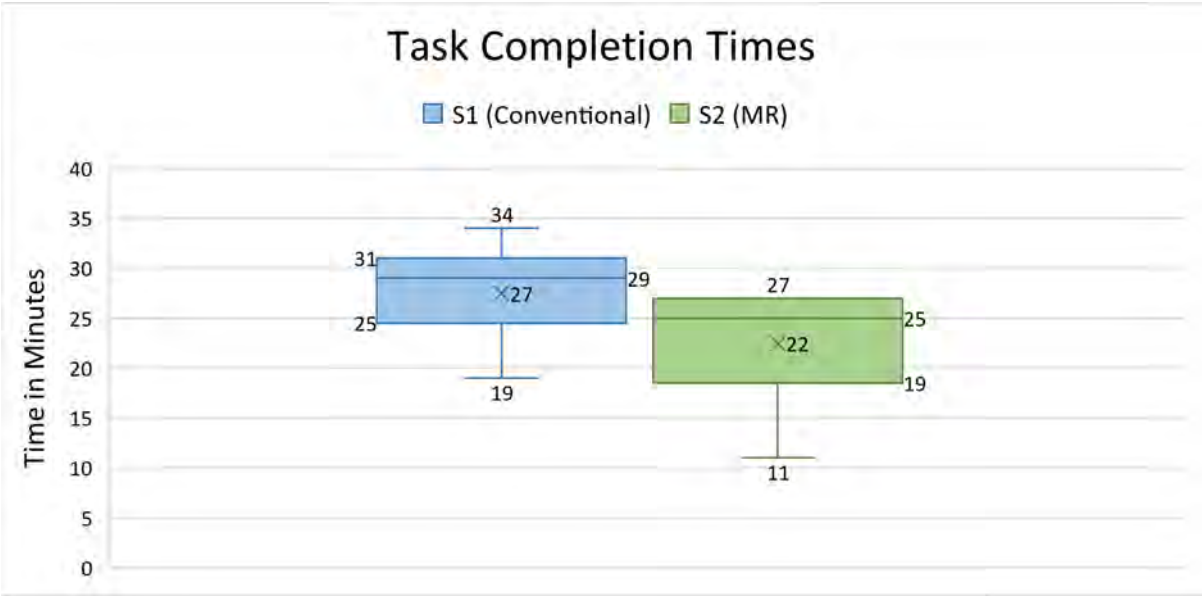


Figure 59: Task Completion Times between S1 and S2

5.7.7 Best ViewPoint (RQ4)

The analysts were tasked with identifying and reporting the optimal view point for conducting their exploration. Every analyst consistently selected either the back of the room or the back corners as their preferred viewpoints for analysis. This consensus arose because the device used for the exploration offered a limited field of view. To overcome this limitation and achieve a more complete perspective of the scene—which included building models, avatars, the display wall, and various user interface (UI) elements—the analysts found it necessary to position themselves at a distance. This strategic positioning allowed them to encompass the majority of these elements within their field of vision, thereby facilitating a more thorough analysis.

5.7.8 Space Usage (RQ4)

In our study, depicted in Figure 54, we present a detailed analysis of space utilization by the analysts using heatmaps to depict their movement patterns. The data reveals that analysts covered a distance ranging from a minimum of 55 meters to a maximum of 154 meters, averaging at 102 meters traveled by each individual. The duration of their exploration sessions also varied, with the shortest session lasting 11.7 minutes and the longest 27.6 minutes, with an average session time of 23.2 minutes per analyst.

The heatmaps serve a dual purpose: The blue heatmap illustrates the areas frequented by participants during the data gathering phase. It is evident that participant movement was primarily concentrated in the central area of the room and near the main wall. This

pattern suggests that participants were actively engaged in comparing three-dimensional models positioned in the room's center with two-dimensional images displayed on the wall.

Furthermore, we have nine red heatmaps that represent the space utilization by each of the nine analysts individually. These heatmaps show that the analysts' exploration was mainly focused around the three-dimensional building models located centrally in the room. The heatmaps indicate heightened activity at the rear and the back right corner of the room. This pattern of movement is primarily attributed to the analysts' efforts to closely examine the building models and to interact with the participant avatars.

Moreover, the activity observed in the back corners of the room can be attributed to the analysts' attempts to simultaneously view multiple assets within the application. This behavior is likely influenced by the limited Field of View (FOV) offered by the HoloLens, which necessitates closer inspection and specific positioning to view the assets effectively. These insights into the analysts' spatial behavior provide valuable context for understanding how they interact with the virtual environment and engage with the visual data during the analysis process.

The movement seen in heatmaps talks about how people attempted to understand the reasons for the users movement in space.

> *..moving around the scene so not specifically staying in one place but when moving around I was able to better understand the movements of the users as well* (a3)

Their movement was also guided by their ability to understand the conversation and the movement of the users while exploring the content.

..being able to understand several aspects of the conversation helps you to follow it, because I could see their position I could see where they were looking at and also subtle movements. (a1)

5.7.9 Statistical Analysis through T-Test (RO1)

Although we had a small sample of 13 participants, we conducted statistical tests to understand the differences in analysts' experiences between the two scenarios, S1 and S2.

We conducted two-tailed paired t-tests for two different measures. Firstly, we examined the time taken to complete tasks. The average time taken to complete S1 was 27 minutes, while for S2 it was 22 minutes, with a p-value of 0.025 (standard deviation ± 5.71). We also conducted t-tests for Likert scale values regarding the ease of determining and accessing important words and phrases. For the question on whether it was easy to determine important words and phrases, we obtained a p-value of 0.019 (standard deviation 1.34). For the question on whether it was easy to access important words and phrases, we obtained a p-value of 0.011 (standard deviation ± 1.51).

5.8 Thematic Analysis

5.8.1 MuSA's Potential for Enhancing Multimodal Analysis

When asked about their preference for using the two different workflows S1 (traditional mode) and S2 (MuSA), 10 analysts preferred using the S2 over S1, two analysts preferred using both S1 & S2, and one preferred S1 over S2. Being able to see the AR models in the application gave analysts much more context to the analysis than S1.

> *I mean I found most of the context to be more relevant when using the HoloLens than with the actual video.. I am really seeing the potential of where this is going and also.. I could clearly see the difference between the interaction with the laptop and the HoloLens was much more efficient. ..I couldn't figure out the gaze and head movements there (S1), but it was really very clear with the HoloLens (S2)* (a8).

It also helped the analysts be more detail-oriented, helping them dive deeper into the analysis.

> *..I was focusing more on trying to figure out what parts of the thing (building model) they were looking at, and then their conversations.. whereas with the video, I was .. just trying to figure out what they're actually doing without being able to see what they're looking at* (a10).

> *I think all the flaws, all the problems that I had with the previous one (S1) with context switching and all of that, .. was pretty much solved in the second one (S2)* (a1).

5.8.2 Seamless Navigation

The analysts found it easier to follow the conversation in S2 when compared to S1. Being embedded in the application gave them clarity as to why the participants were moving vs. being stationary. It was easier to navigate to different parts of the conversation using the word cloud. It also helped them get an understanding of the purpose of the experiment and seemed more natural.

> *I think it makes a lot more sense (S2).. I felt like that activity (S1) was a lot more contrived and this one (S2) felt more natural. Okay, conversation analysis, that's what I'm doing. It makes sense. I was actually looking at a conversation and I think I had all the tools* (a5).

Their experience was also enhanced with the ability to be able to view from multiple viewpoints at will vs. its counterpart where they had a pre-determined viewpoint of the camera.

> *"I would say it's much better than the previous one because, ..there are many points of view that you can look into.. and see how the conversation is going on"* (a4).

5.8.3 Other valuable insights and feedback

Analysts appreciated the experiences MuSA provided, noting it allowed them to feel both close to and distant from their study subjects simultaneously. An analyst described feeling like an outsider looking in, yet also connected. In scenario (S2), the addition of analytical tools like word clouds created a sense of separation while engaging with the data. This dual perspective of being close and distant offered a distinct viewpoint that improved their understanding and interaction with the data.

> *So I was almost like an outsider but inside the view .. this contrast between being close and being far away.. in this one (S2), I think that was much stronger. I was close, but I was far away because I had all these metrics of word cloud and stuff.. so it was two distances at the same time. So that helped a lot"*(a1).

The ability to freely move and choose optimal positions within the environment for viewing 3D objects enhanced the understanding of content. This approach offered perspectives unachievable with 2D images, thus facilitating more effective sensemaking.

> *But it really, really helps that I can see what they're seeing because .. when I was like seeing the video I had no clue.. I can make sense of it but it took me a lot of effort just to make sense.. So they said bell tower (part of the school model).. And also like the trees right they talk about the one in the back (part of the house model) which I can't see in the picture .. that is like very hard to do when I was doing it in the video (S1) but here (S2) it's very easy.*(a7).

An analyst expressed that experiencing the meeting through MuSA provided a better and more effective learning experience about the conversation compared to S1, and even more so than attending the meeting in person.

> *I could see their faces, where they were looking at, and figure out where they were pointing to, and go back and check the buildings in different perspectives, which was impossible through the first approach. Even if I was in person, I could not do this since I would mess their experience* (a7).

The analysts appreciated the ability to use a see-through device as they felt confident to move around in space.

> *The see-through headset was vital because it gave me the confidence to move around the space* (a6).

They said it helped maintain a sense of where they were at all times.

> *I didn't lose sense of where I was as it could happen in a fully VR experience. It also helped to be in the same physical environment where the conversation happened.* (a1)

It also helped them get the best of both worlds.

> *I can see both the big screen and the participants at the same time, as well as the model in the middle.*(a13).

5.8.4 Navigating Challenges

We faced multiple challenges at various stages of the pipeline, and have detailed two specific issues below. Additionally, we continued to experience HoloLens2 overheating problems during this phase.

Transcription and sync - To replicate the environment accurately, we needed to synchronize the voice, chat bubble movements, head and body motions, and asset changes. We used whisper transcription for the voice recordings to generate chat bubbles, aiming to accurately timestamp different spoken segments. However, the whisper transcription was not precise. Moreover, the model's inaccuracies led to timestamp errors, requiring manual adjustments to generate near-accurate timestamps. These complications resulted in synchronization issues between the voice recordings and their transcriptions, occasionally making the process more challenging for analysts.

Low-Lighting Conditions - In this scenario, we had to keep the display wall on so the analysts had access to the same content as the participants in the data collection step along with its spatial context. This setup offered more precise alignment when compared to its representation using 3D models instead. To facilitate this setup, we needed to dim the lights in the room. This adjustment allowed analysts to smoothly explore the conversation by clearly viewing both the application content and real-world objects. However, this made real objects less visible in the environment and also complicated the process of recording the analysts' activities during the study.

In this chapter, we examined the system enhancements implemented for the Phase II user study, including the involved components, their workflow, and changes made to the user interface (UI). We also briefly outlined the protocol adopted for the user study. Subsequently, we presented our results in detail, explaining how they relate to our research questions. Additionally, we analyzed the usage of space by analysts in Phase II and its correlation with the field of view (FOV) afforded by the device. We further addressed the limitations of the system. In the next chapter, we will discuss the results from all three evaluations, identify areas for improvement, and consider potential avenues for future work.

CHAPTER 6

CONCLUSION, DISCUSSION, FUTURE WORK

Parts of this chapter have been published in the proceedings of ISMAR 2023 [90] and UIST 2023 [91].

6.1 Conclusion

In this work, we initially established a comprehensive pipeline for conducting multimodal analysis within immersive environments. This pipeline served as the foundation for developing interactive applications tailored for Mixed Reality (MR) and Virtual Reality (VR) using the Unity software platform. Our development process encompassed several key stages, including tracking, data capturing, data cleaning, prototype development, and ultimately, deploying the final product to the end-user's hardware. Our investigation focused on understanding the impact of embodied cognition and situated analytics on multimodal conversations. This was achieved by immersing analysts within a simulated conversational environment through two distinct user evaluations, aiming to assess how such immersive applications could aid analysts in strategizing and sensemaking while navigating complex data sets within these virtual spaces.

The first user evaluation involved 12 participants and utilized HoloLens2 and Quest2 devices to examine how seated participants interact within the immersive environment. This step was followed by an expert evaluation involving specialists with expertise in communication and linguistics using the contextual inquiry method. Through this process, we developed an

understanding of the MuSA's usability and identified areas for improvement, along with its potential use cases and applications. We then refined MuSA for the next phase of user studies. The second user study expanded the research scope to 13 participants, contrasting traditional interaction methods with immersive experiences facilitated by the HoloLens2 device, focusing on the dynamics of non-seated, moving participants.

A significant finding from our studies is that the ability for analysts to change their viewpoints and physically move around within the space significantly enhances their engagement with multimodal conversations. This dynamic interaction provides deeper insights into the unfolding of conversations across various contexts.

This research contributes to the broader exploration of immersive technologies beyond their traditional applications in recreation, entertainment, and training. It offers new directions for future studies in the design and evaluation of these technologies, especially in the context of multimodal analysis. By leveraging MuSA to visualize and analyze multimodal conversations, this work suggests potential improvements to the efficiency of data analysis pipelines and enriches our understanding of complex, richly multimodal conversations.

6.2 Discussion, Future Work

Accurate tracking information is crucial during the data collection phase to ensure the prototype conveys reliable data. Inadequate tracking can lead to confusion among analysts, compromising their experience and the integrity of the analysis results. Key practices include ensuring that Optitrack markers are always visible and unobstructed. For example, it has been observed that participants' hair can obscure the markers on the back, and their fingers may

cover the markers on the phone while they explore it. Such obstructions can cause significant discrepancies in the data, necessitating the repetition of the data collection process to maintain accuracy.

In the initial version of our prototype, the gaze lines were white, which some experts found distracting during the evaluation session. To address this, we introduced color-coded gaze lines in the second version, reducing their visual intensity and making them less intrusive, which enhanced the analysts' experience. Feedback also suggested adding the ability to toggle the gaze lines on or off, a feature we plan to consider for future updates.

Additionally, there were concerns that the avatars did not accurately represent the gender of the participants. To resolve this, we introduced gender-specific avatars with color-coding in the subsequent phase and limited the avatar visualization to the neck to avoid unnatural poses caused by incomplete body tracking data.

Another issue identified was with the word cloud feature; the words lacked a distinct background, often blending into the overall interface background. In response, we added a translucent background to each word in the second version, which improved visibility and facilitated more precise word selection, enhancing the user experience.

During the expert evaluation, it was noted that the interface lacked a tagging feature for logging non-verbal behaviors. We addressed this by introducing a new menu in the interface that includes examples to assist in logging such behaviors, further improving the user experience.

It has been noted that the field of view (FOV) of the devices used significantly influenced the behavior of the analysts. During the first phase of the experiment, analysts had the choice

between two devices: the HoloLens2, which has a smaller FOV, and the Quest2, which offers a larger FOV. It was observed that the analysts showed a preference for the Quest2 due to its wider FOV. A smaller FOV in the HoloLens2 led to increased physical movements as analysts needed to compensate for the limited visibility by frequently adjusting their head position to gather more information. This necessity for additional head movements and positional adjustments was less pronounced with the Quest2. Additionally, spatial preferences were noted: analysts using the Quest2 tended to prefer the back side of the room, while those using the HoloLens2 gravitated towards the back corner. In the second phase, the trend continued, with the FOV playing a decisive role in how analysts explored and utilized space. The preferred positions at the back of the room or the back corner were likely chosen to allow analysts an optimal view of both the architectural models and the display in the room.

Our work presents a framework for immersing users in a recorded conversation using two different immersive environments: one utilizing the HoloLens 2 device for mixed reality, and the other using the Quest2 device for virtual reality. To evaluate the effectiveness of our prototype, we conducted a user study with 12 analysts. The recorded data from a previous user study involved two seated participants, limiting the movement of the models to head movements. However, analyzing the analysts' movements in space would provide valuable insights into their level of engagement and comfort during the study, and therefore, it may be worth exploring in future research. Furthermore, we could understand how users would interact with any physical object in the environment such as placing a chair or desk in the exploration space that the analysts could use when they experience fatigue while performing the tasks.

During Phase I of the study, the analysts were required to use two different interfaces: the HoloLens 2 for mixed reality (MR) and the Quest2 for virtual reality (VR). This arrangement imposed an additional challenge as the analysts needed to learn and adapt to two distinct input systems. It would be beneficial to have consistency in the input systems across all the devices used in the study. To achieve this we could either move to a gesture-based system on Quest2 (HoloLens 2 already uses a gesture-based input system) or use a voice-enabled input system. Implementing such a system would potentially decrease the training and test times of the experiment. We implemented a menu-based system for interaction (Figure 16). Since we wanted to have the menu easily accessible and available at all times we had to place it in the analysts' field of view and within the users' arms reach which seemed to have an undesirable effect on the analysts. A few analysts suggested that they would have preferred it as a distance. Hence implementing a pointer-based system might have been useful in such a case. The task instructions seemed to be a comfortable place for most users. Hence placing the menu at such a distance might be helpful where people could point to the menu and access it with a tap or pinch gesture.

To recreate the environment in Virtual Reality, we used a 3D model of a classroom in our lab that was previously developed using Blender and Unity. It is important to note that this method may not be easily replicable for other settings. Thus, we suggest exploring other options, such as using structure from motion or 360 videos in the VR environment instead, which may be more feasible for certain scenarios. Using Neural Radiance Field (NeRF) may be another way to consider for generating assets and recreating physical space required for situated analytics in

VR. Our current setting is limited to our lab space, where we can record and analyze the data in situ. However, with the advancements in immersive technologies and ubiquitous computing, it may become possible to replicate the behavior outdoors as well.

With the advent of better MR headsets such as the Vision Pro and Quest3 and their automatic adaptability to different lighting conditions, the need to limit the experiences to low-lighting conditions may be eliminated. With advancements in AI and ML, using lifelike avatars and lip sync may also become possible. Better and more accurate AI models for transcription, along with built-in speaker diarization, could potentially eliminate the need for manual transcription and editing. This would help avoid sync issues with other media in the meetings, thereby leading to better exploration experiences. All these advancements can potentially improve the end-user experience and significantly reduce the pipeline's development and deployment time.

Another use case for the system is to conduct multi-level analysis and exploration. In this work we have 12 analysts (say analysts-set1) explore the recorded data of two participants. However, since the analysts were tracked and the experiment was conducted where the conversation originally occurred we could have the next set of analysts(say analysts-set2) explore the analysis sessions of both participants and the analysts-set1. This can further be extended to any number of levels if needed while having a way to visually differentiate analysts from various sets.

CITED LITERATURE

Bibliography

[1] Contemporary issues in conversation analysis: Embodiment and materiality, multimodality and multisensoriality in social interaction. *Journal of Pragmatics*, 145:47–62, 2019. ISSN 0378-2166. doi: https://doi.org/10.1016/j.pragma.2019.01.016. Quo Vadis, Pragmatics?

[2] Keith Anderson, Elisabeth André, Tobias Baur, Sara Bernardini, Mathieu Chollet, Evi Chryssafidou, Ionut Damian, Cathy Ennis, Arjan Egges, Patrick Gebhard, et al. The tardis framework: intelligent virtual agents for social coaching in job interviews. In *Advances in Computer Entertainment: 10th International Conference, ACE 2013, Boekelo, The Netherlands, November 12-15, 2013. Proceedings 10*, pages 476–491. Springer, 2013.

[3] Christopher Andrews, Alex Endert, Beth Yost, and Chris North. Information visualization on large, high-resolution displays: Issues, challenges, and opportunities. *Information Visualization*, 10(4):341–355, 2011.

[4] Ágnes Karolina Bakk. Magic and immersion in vr. In *Interactive Storytelling: 13th International Conference on Interactive Digital Storytelling, ICIDS 2020, Bournemouth, UK, November 3–6, 2020, Proceedings 13*, pages 327–331. Springer, 2020.

[5] Michal Balazia, Philipp Müller, Ákos Levente Tánczos, August von Liechtenstein, and François Brémond. Bodily behaviors in social interaction: Novel annotations and state-of-the-art evaluation. New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450392037. doi: 10.1145/3503161.3548363.

[6] P Bandyopadhyay, L Lisle, C North, and DA Bowman. Immersive space to think: The role of 3d space for sensemaking. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. ACM*, 2020.

[7] Tobias Baur, Ionut Damian, Florian Lingenfelser, Johannes Wagner, and Elisabeth André. Nova: Automated analysis of nonverbal signals in social interactions. In *Human Behavior Understanding: 4th International Workshop, HBU 2013, Barcelona, Spain, October 22, 2013. Proceedings 4*, pages 160–171. Springer, 2013.

[8] Roman Bednarik, Shahram Eivazi, and Michal Hradis. Gaze and conversational engagement in multiparty video conversation: an annotation scheme and classification of high and low levels of engagement. In *Proceedings of the 4th workshop on eye gaze in intelligent human machine interaction*, pages 1–6, 2012.

[9] Hugh Beyer and Karen Holtzblatt. *Contextual Design: Defining Customer-Centered Systems*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1997. ISBN 9780080503042.

[10] Jonathan Bollen, Julie Holledge, and Joanne Tompkins. Putting virtual theatre models to work:'virtual praxis' for performance research in theatre history. *Theatre and Performance Design*, 7(1-2):6–23, 2021.

[11] Doug A Bowman and Ryan P McMahan. Virtual reality: how much immersion is enough? *Computer*, 40(7):36–43, 2007.

[12] John Brooke. *SUS – a quick and dirty usability scale*, pages 189–194. 01 1996.

[13] Barry Brown, Moira McGregor, and Donald McMillan. Searchable objects: Search in everyday conversation. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 508–517, 2015.

[14] Wolfgang Büschel, Annett Mitschick, and Raimund Dachselt. Demonstrating reality-based information retrieval. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–4, 2018.

[15] Giuseppe Caggianese, Valerio Colonnese, and Luigi Gallo. Situated visualization in augmented reality: Exploring information seeking strategies. In *2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pages 390–395. IEEE, 2019.

[16] Tom Chandler, Maxime Cordeil, Tobias Czauderna, Tim Dwyer, Jaroslaw Glowacki, Cagatay Goncu, Matthias Klapperstueck, Karsten Klein, Kim Marriott, Falk Schreiber, et al. Immersive analytics. In *2015 Big Data Visual Analytics (BDVA)*, pages 1–8. IEEE, 2015.

[17] Debaleena Chattopadhyay, Kenton O'Hara, Sean Rintel, and Roman Rädle. Office social: Presentation interactivity for nearby devices. CHI '16, New York, NY, USA, 2016. Association for Computing Machinery.

[18] Xin Chen, Jessica Zeitz Self, Leanna House, John Wenskovitch, Maoyuan Sun, Nathan Wycoff, Jane Robertson Evia, and Chris North. Be the data: Embodied visual analytics. *IEEE Transactions on Learning Technologies*, 11(1):81–95, 2017.

[19] Zhutian Chen, Yijia Su, Yifang Wang, Qianwen Wang, Huamin Qu, and Yingcai Wu. Marvist: Authoring glyph-based visualization in mobile augmented reality. *IEEE Transactions on Visualization and computer graphics*, 26(8):2645–2658, 2019.

[20] Yi Wei Chew. Performing presence with the teaching-body via videoconferencing: A postdigital study of the teacher's face and voice. *Postdigital Science and Education*, 4(2): 394–421, 2022.

[21] Haeyong Chung, Sharon Lynn Chu, and Chris North. A comparison of two display models for collaborative sensemaking. In *Proceedings of the 2nd ACM International Symposium on Pervasive Displays*, pages 37–42, 2013.

[22] Maxime Cordeil, Andrew Cunningham, Tim Dwyer, Bruce H Thomas, and Kim Marriott. Imaxes: Immersive axes as embodied affordances for interactive multivariate data visualisation. In *Proceedings of the 30th annual ACM symposium on user interface software and technology*, pages 71–83, 2017.

[23] Carolina Cruz-Neira, Daniel J Sandin, and 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*, pages 135–142, 1993.

[24] Kylie Davidson, Lee Lisle, Kirsten Whitley, Doug A Bowman, and Chris North. Exploring the evolution of sensemaking strategies in immersive space to think. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[25] Ciro Donalek, S George Djorgovski, Alex Cioc, Anwell Wang, Jerry Zhang, Elizabeth Lawler, Stacy Yeh, Ashish Mahabal, Matthew Graham, Andrew Drake, et al. Immersive and collaborative data visualization using virtual reality platforms. In *2014 IEEE International Conference on Big Data (Big Data)*, pages 609–614. IEEE, 2014.

[26] Vanessa Echeverria, Roberto Martinez-Maldonado, Lixiang Yan, Linxuan Zhao, Gloria Fernandez-Nieto, Dragan Gašević, and Simon Buckingham Shum. Huceta: A framework for human-centered embodied teamwork analytics. *IEEE Pervasive Computing*, 2022.

[27] Neven ElSayed, Bruce Thomas, Kim Marriott, Julia Piantadosi, and Ross Smith. Situated analytics. In *2015 Big Data Visual Analytics (BDVA)*, pages 1–8. IEEE, 2015.

[28] Neven AM ElSayed, Ross T Smith, Kim Marriott, and Bruce H Thomas. Blended ui controls for situated analytics. In *2016 Big Data Visual Analytics (BDVA)*, pages 1–8. IEEE, 2016.

[29] Neven AM ElSayed, Bruce H Thomas, Kim Marriott, Julia Piantadosi, and Ross T Smith. Situated analytics: Demonstrating immersive analytical tools with augmented reality. *Journal of Visual Languages & Computing*, 36:13–23, 2016.

[30] Ulrich Engelke, Casey Rogers, Jens Klump, and Ian Lau. Hypar: Situated mineralogy exploration in augmented reality. In *The 17th International Conference on Virtual-Reality Continuum and Its Applications in Industry*, pages 1–5, 2019.

[31] Negin Esmaeili and Jahanyar Bamdad Soofi. Expounding the knowledge conversion processes within the occupational safety and health management system (osh-ms) using concept mapping. *International Journal of Occupational Safety and Ergonomics*, 28(2):1000–1015, 2022.

[32] Alessandro Febretti, Arthur Nishimoto, Terrance Thigpen, Jonas Talandis, Lance Long, JD Pirtle, Tom Peterka, Alan Verlo, Maxine Brown, Dana Plepys, et al. Cave2: a hybrid reality environment for immersive simulation and information analysis. In *The Engineering Reality of Virtual Reality 2013*, volume 8649, pages 9–20. SPIE, 2013.

[33] Siska Fitrianie, Zhenke Yang, Dragoş Datcu, Alin G. ChiŢu, and Léon J. M. Rothkrantz. *Context-Aware Multimodal Human-Computer Interaction*. Springer Berlin Heidelberg, 2010. ISBN 978-3-642-11688-9. doi: 10.1007/978-3-642-11688-9_9.

[34] Sabiha Ghellal, Einav Katan-Schmid, Ramona Mosse, Christian Stein, Nitsan Margaliot, and Lisanne Goodhue. From immersion to interference: Sites of collaboration in playing with virtual realities. *Global Performance Studies*, 4(2), 2021.

[35] Edward Twitchell Hall. In *The Silent Language*, 1959. URL `https://api.semanticscholar.org/CorpusID:143072138`.

[36] Alexander Heimerl, Tobias Baur, Florian Lingenfelser, Johannes Wagner, and Elisabeth André. Nova-a tool for explainable cooperative machine learning. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 109–115. IEEE, 2019.

[37] Y Hong, Benjamin Watson, Kenneth Thompson, and Paul Davis. Talk2hand: Knowledge board interaction in augmented reality easing analysis with machine learning assistants. In *EuroVis Workshop on Visual Analytics (EuroVA), K. Vrotsou and J. Bernard, Eds. The Eurographics Association*, 2021.

[38] Sebastian Hubenschmid, Jonathan Wieland, Daniel Immanuel Fink, Andrea Batch, Johannes Zagermann, Niklas Elmqvist, and Harald Reiterer. Relive: bridging in-situ and ex-situ visual analytics for analyzing mixed reality user studies. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–20, 2022.

[39] Elizabeth Keating and Maria Egbert. Conversation as a cultural activity. *A companion to linguistic anthropology*, pages 167–196, 2005.

[40] Sarah Ketchell, Winyu Chinthammit, and Ulrich Engelke. Situated storytelling with slam enabled augmented reality. In *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*, pages 1–9, 2019.

[41] Karsten Klein, Michael Sedlmair, and Falk Schreiber. Immersive analytics: An overview. *it-Information Technology*, 2022.

[42] Dylan Kobayashi, Nurit Kirshenbaum, Roderick S Tabalba, Ryan Theriot, and Jason Leigh. Translating the benefits of wide-band display environments into an xr space. In *Proceedings of the 2021 ACM Symposium on Spatial User Interaction*, pages 1–11, 2021.

[43] Kathrin Koebel and Doris Agotai. Embodied interaction for the exploration of image collections in mixed reality (mr) for museums and other exhibition spaces. In *HCI International 2020-Posters: 22nd International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part III 22*, pages 291–299. Springer, 2020.

[44] Steffi Kohl, Kay Schröder, Mark Graus, Emir Efendic, and Jos G.A.M. Lemmink. Zooming in on the effect of sociometric signals on different stages of the design process. New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450393270. doi: 10.1145/3527927.3532810.

[45] Ben Kreimer. Allensworth-building-models, 2022. URL `https://sketchfab.com/benkreimer/collections/allensworth-building-models`.

[46] Jingya Li and Zheng Wang. An interactive augmented reality graph visualization for chinese painters. *Electronics*, 11(15):2367, 2022.

[47] Tica Lin, Yalong Yang, Johanna Beyer, and Hanspeter Pfister. Labeling out-of-view objects in immersive analytics to support situated visual searching. *IEEE Transactions on Visualization and Computer Graphics*, 2021.

[48] Lee Lisle, Xiaoyu Chen, JK Edward Gitre, Chris North, and Doug A Bowman. Evaluating the benefits of the immersive space to think. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 331–337. IEEE, 2020.

[49] Lee Lisle, Kylie Davidson, Edward JK Gitre, Chris North, and Doug A Bowman. Sensemaking strategies with immersive space to think. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pages 529–537. IEEE, 2021.

[50] Weizhou Luo, Anke Lehmann, Yushan Yang, and Raimund Dachselt. Investigating document layout and placement strategies for collaborative sensemaking in augmented reality. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–7, 2021.

[51] G Elisabeta Marai, Angus G Forbes, and Andrew Johnson. Interdisciplinary immersive analytics at the electronic visualization laboratory: Lessons learned and upcoming challenges. In *2016 Workshop on Immersive Analytics (IA)*, pages 54–59. IEEE, 2016.

[52] G Elisabeta Marai, Jason Leigh, and Andrew Johnson. Immersive analytics lessons from the electronic visualization laboratory: A 25-year perspective. *IEEE computer graphics and applications*, 39(3):54–66, 2019.

[53] Kim Marriott, Falk Schreiber, Tim Dwyer, Karsten Klein, Nathalie Henry Riche, Takayuki Itoh, Wolfgang Stuerzlinger, and Bruce H Thomas. *Immersive analytics*, volume 11190. Springer, 2018.

[54] Sven Mayer, Lars Lischke, Paweł W Woźniak, and Niels Henze. Evaluating the disruptiveness of mobile interactions: A mixed-method approach. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2018.

[55] Daniel McDuff, Kael Rowan, Piali Choudhury, Jessica Wolk, ThuVan Pham, and Mary Czerwinski. A multimodal emotion sensing platform for building emotion-aware applications. *arXiv preprint arXiv:1903.12133*, 2019.

[56] Ross Alan Mead. *Situated Proxemics and Multimodal Communication: Space, Speech, and Gesture in Human-Robot Interaction*. PhD thesis, University of Southern California, 2016.

[57] Brunelli P Miranda, Vinicius F Queiroz, Tiago DO Araújo, Carlos GR Santos, and Bianchi S Meiguins. A low-cost multi-user augmented reality application for data visualization. *Multimedia Tools and Applications*, 81(11):14773–14801, 2022.

[58] Lorenza Mondada. The local constitution of multimodal resources for social interaction. *Journal of Pragmatics*, 65:137–156, 2014.

[59] Lorenza Mondada. Challenges of multimodality: Language and the body in social interaction. *Journal of sociolinguistics*, 20(3):336–366, 2016.

[60] Lorenza Mondada. Multiple temporalities of language and body in interaction: Challenges for transcribing multimodality. *Research on Language and Social Interaction*, 51(1):85–106, 2018.

[61] Alessandro Montanari, Zhao Tian, Elena Francu, Benjamin Lucas, Brian Jones, Xia Zhou, and Cecilia Mascolo. Measuring interaction proxemics with wearable light tags. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(1):1–30, 2018.

[62] AJung Moon, Daniel M Troniak, Brian Gleeson, Matthew KXJ Pan, Minhua Zheng, Benjamin A Blumer, Karon MacLean, and Elizabeth A Croft. Meet me where i'm gazing: how shared attention gaze affects human-robot handover timing. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 334–341, 2014.

[63] Philipp Müller, Michael Dietz, Dominik Schiller, Dominike Thomas, Hali Lindsay, Patrick Gebhard, Elisabeth André, and Andreas Bulling. Multimediate'22: Backchannel detection and agreement estimation in group interactions. Association for Computing Machinery, 2022. ISBN 9781450392037. doi: 10.1145/3503161.3551589.

[64] A. G. Naik and A. E. Johnson. Psa: A cross-platform framework for situated analytics in mr and vr. In *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 92–96, Los Alamitos, CA, USA, oct 2023. IEEE Computer Society. doi: 10.1109/ISMAR-Adjunct60411.2023.00027. URL `https://doi.ieeecomputersociety.org/10.1109/ISMAR-Adjunct60411.2023.00027`. © 2023 IEEE.

[65] Ashwini G Naik and Andrew E Johnson. Using personal situated analytics (psa) to interpret recorded meetings. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, UIST '23 Adjunct, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400700965. doi: 10.1145/3586182.3616697. URL `https://doi.org/10.1145/3586182.3616697`.

[66] Huyen Nguyen, Sarah Ketchell, Ulrich Engelke, Bruce H Thomas, and Paulo de Souza. Augmented reality based bee drift analysis: A user study. In *2017 International Symposium on Big Data Visual Analytics (BDVA)*, pages 1–8. IEEE, 2017.

[67] Arthur Nishimoto. Virtualuic-evl, January 2019. URL `https://bitbucket.org/arthurnishimoto/virtualuic-evl`.

[68] Chris North, Tim Dwyer, Bongshin Lee, Danyel Fisher, Petra Isenberg, George Robertson, and Kori Inkpen. Understanding multi-touch manipulation for surface computing. In Tom Gross, Jan Gulliksen, Paula Kotzé, Lars Oestreicher, Philippe Palanque, Raquel Oliveira Prates, and Marco Winckler, editors, *Human-Computer Interaction – INTERACT 2009*, pages 236–249, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg. ISBN 978-3-642-03658-3.

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

[70] Anna Penzkofer, Philipp Müller, Felix Bühler, Sven Mayer, and Andreas Bulling. Conan: A usable tool for multimodal conversation analysis. In *Proceedings of the 2021 International Conference on Multimodal Interaction*, pages 341–351, 2021.

[71] Arnaud Prouzeau, Maxime Cordeil, Clement Robin, Barrett Ens, Bruce H Thomas, and Tim Dwyer. Scaptics and highlight-planes: Immersive interaction techniques for finding occluded features in 3d scatterplots. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.

[72] Arnaud Prouzeau, Yuchen Wang, Barrett Ens, Wesley Willett, and Tim Dwyer. Corsican twin: Authoring in situ augmented reality visualizations in virtual reality. In *Proceedings of the International Conference on Advanced Visual Interfaces*, pages 1–9, 2020.

[73] André LM Ramos, Thiago Korb, and Alexandra Okada. Immersive analytics through holosenai motor mixed reality app. In *Intelligent Computing: Proceedings of the 2019 Computing Conference, Volume 2*, pages 1259–1268. Springer, 2019.

[74] Ready Player Me. Human characters (free sample pack). URL `https://assetstore.unity.com/publishers/50744`.

[75] Kadek Ananta Satriadi, Barrett Ens, Maxime Cordeil, Tobias Czauderna, and Bernhard Jenny. Maps around me: 3d multiview layouts in immersive spaces. *Proceedings of the ACM on Human-Computer Interaction*, 4(ISS):1–20, 2020.

[76] Steve Sawyer, Joel Farber, and Robert Spillers. Supporting the social processes of software development. *Information Technology & People*, 10(1):46–62, 1997.

[77] Alexander Schafer, Tomoko Isomura, Gerd Reis, Katsumi Watanabe, and Didier Stricker. Mutualeyecontact: A conversation analysis tool with focus on eye contact. In *ACM Symposium on Eye Tracking Research and Applications*, pages 1–5, 2020.

[78] Richard Skarbez, Nicholas F Polys, J Todd Ogle, Chris North, and Doug A Bowman. Immersive analytics: Theory and research agenda. *Frontiers in Robotics and AI*, 6:82, 2019.

[79] Kalin Stefanov, Baiyu Huang, Zongjian Li, and Mohammad Soleymani. Opensense: A platform for multimodal data acquisition and behavior perception. In *Proceedings of the 2020 International Conference on Multimodal Interaction*, pages 660–664, 2020.

[80] Natalia Stewart Rosenfield, Kathleen Lamkin, Jennifer Re, Kendra Day, LouAnne Boyd, and Erik Linstead. A virtual reality system for practicing conversation skills for children with autism. *Multimodal Technologies and Interaction*, 3(2):28, 2019.

[81] Giota Stratou and Louis-Philippe Morency. Multisense—context-aware nonverbal behavior analysis framework: A psychological distress use case. *IEEE Transactions on Affective Computing*, 8(2):190–203, 2017.

[82] Wolfgang Stuerzlinger, Tim Dwyer, Steven Drucker, Carsten Görg, Chris North, and Gerik Scheuermann. Immersive human-centered computational analytics. *Immersive Analytics*, pages 139–163, 2018.

[83] Erin Sullivan. *Shakespeare and Digital Performance in Practice*. Springer Nature, 2022.

[84] Roderick Tabalba, Nurit Kirshenbaum, Jason Leigh, Abari Bhatacharya, Andrew Johnson, Veronica Grosso, Barbara Di Eugenio, and Moira Zellner. Articulate+: An always-listening natural language interface for creating data visualizations. In *Proceedings of the 4th Conference on Conversational User Interfaces*, pages 1–6, 2022.

[85] Roderick S Tabalba, Nurit Kirshenbaum, Jason Leigh, Abari Bhattacharya, Veronica Grosso, Barbara Di Eugenio, Andrew E Johnson, and Moira Zellner. An investigation into an always listening interface to support data exploration. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, pages 128–141, 2023.

[86] Stephanie Tan, David MJ Tax, and Hayley Hung. Multimodal joint head orientation estimation in interacting groups via proxemics and interaction dynamics. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(1):1–22, 2021.

[87] Stephanie Teasley, Lisa Covi, Mayuram S Krishnan, and Judith S Olson. How does radical collocation help a team succeed? In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pages 339–346, 2000.

[88] Lucia Terrenghi, Aaron Quigley, and Alan Dix. A taxonomy for and analysis of multi-person-display ecosystems. *Personal and Ubiquitous Computing*, 13:583–598, 2009.

[89] Bruce H Thomas, Gregory F Welch, Pierre Dragicevic, Niklas Elmqvist, Pourang Irani, Yvonne Jansen, Dieter Schmalstieg, Aurélien Tabard, Neven AM ElSayed, Ross T Smith, et al. Situated analytics. *Immersive analytics*, 11190:185–220, 2018.

[90] Lucy Thornett. The scenographic potential of immersive technologies: virtual and augmented reality at the prague quadrennial 2019. *Theatre and performance design*, 6(1-2): 102–116, 2020.

[91] Johannes Wagner, Florian Lingenfelser, Tobias Baur, Ionut Damian, Felix Kistler, and Elisabeth André. The social signal interpretation (ssi) framework: multimodal signal processing and recognition in real-time. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 831–834, 2013.

[92] Keziah Wallis and Miriam Ross. Fourth vr: Indigenous virtual reality practice. *Convergence*, 27(2):313–329, 2021.

[93] Zhen Wen, Wei Zeng, Luoxuan Weng, Yihan Liu, Mingliang Xu, and Wei Chen. Effects of view layout on situated analytics for multiple-view representations in immersive visualization. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):440–450, 2022.

[94] Bob G Witmer and Michael J Singer. Presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 1994.

[95] Stephen Wittek and David McInnis. *Shakespeare and Virtual Reality*. Cambridge University Press, 2021.

[96] Bingjie Xu, Shunan Guo, Eunyee Koh, Jane Hoffswell, Ryan Rossi, and Fan Du. Arshopping: In-store shopping decision support through augmented reality and immersive visualization. In *2022 IEEE Visualization and Visual Analytics (VIS)*, pages 120–124. IEEE, 2022.

[97] Yalong Yang, Tim Dwyer, Kim Marriott, Bernhard Jenny, and Sarah Goodwin. Tilt map: Interactive transitions between choropleth map, prism map and bar chart in immersive environments. *IEEE Transactions on Visualization and Computer Graphics*, 27(12):4507–4519, 2020.

[98] Syed Fawad M Zaidi, Niusha Shafiabady, and Justin Beilby. Persistent postural-perceptual dizziness interventions—an embodied insight on the use virtual reality for technologists. *Electronics*, 11(1):142, 2022.

[99] Mengya Zheng and Abraham G Campbell. Location-based augmented reality in-situ visualization applied for agricultural fieldwork navigation. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 93–97. IEEE, 2019.

APPENDIX

PERMISSIONS FOR REUSE OF PREVIOUSLY PUBLISHED MATERIAL

A.1 IEEE

Statement from IEEE RightsLink regarding previously published work included in this dissertation, specifically [64] in chapters 1, 2, 3, and 6.

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1. In the case of textual material (e.g., 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.

APPENDIX (Continued)

2. 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.
3. 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.

A.2 ACM

Statement from Author Rights & Responsibilities (acm.org) regarding previously published work included in this dissertation, specifically [65] in chapters 1, 2, 3, and 6.

Reuse

Authors can reuse any portion of their own work in a new work of their own (and no fee is expected) as long as a citation and DOI pointer to the Version of Record in the ACM Digital Library are included.

Contributing complete papers to any edited collection of reprints for which the author is not the editor, requires permission and usually a republication fee. Authors can include partial or complete papers of their own (and no fee is expected) in a dissertation as long as citations and DOI pointers to the Versions of Record in the ACM Digital Library are included. Authors can use any portion of their own work in presentations and in the classroom (and no fee is expected). Commercially produced course-packs that are sold to students require permission and possibly a fee.

APPENDIX

SUPPLEMENTARY FIGURES

B.1 Additional Box Plots and Trajectories

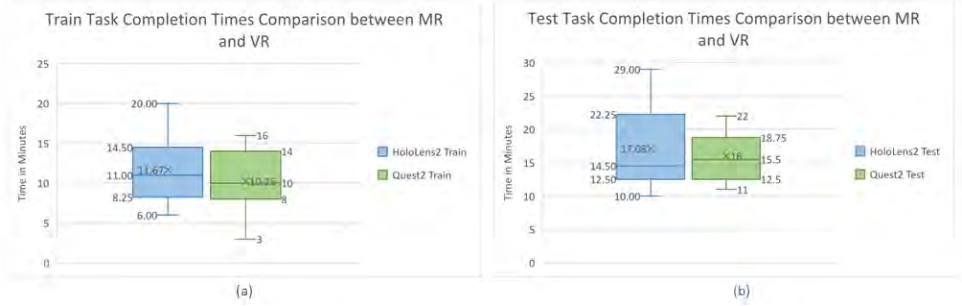


Figure 60: Distribution of Task Completion Times for MR and VR environments (a) Train Tasks (b) Test Tasks

APPENDIX (Continued)

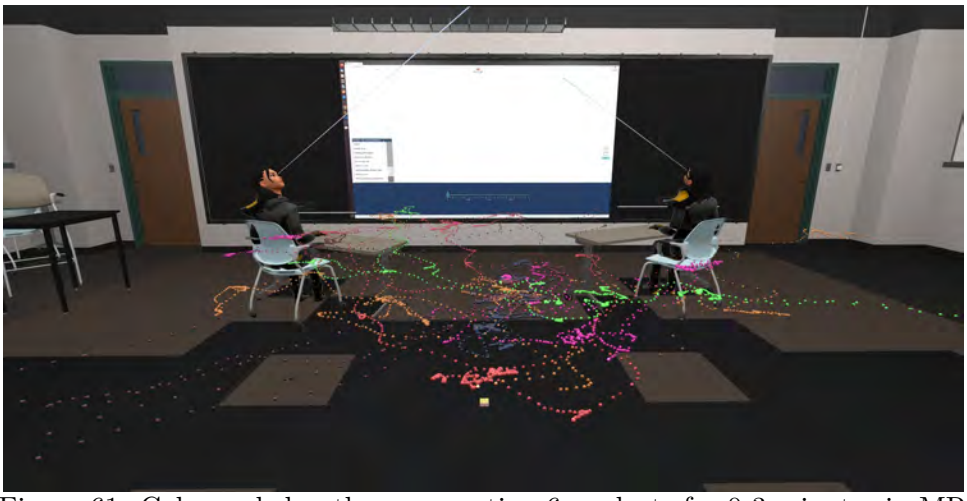


Figure 61: Color-coded paths representing 6 analysts for 0-3 minutes in MR

Figure 62: Color-coded paths representing 6 analysts for 0-3 minutes in VR

APPENDIX (Continued)

Figure 63: Color-coded paths representing 6 analysts for 3-6 minutes in MR

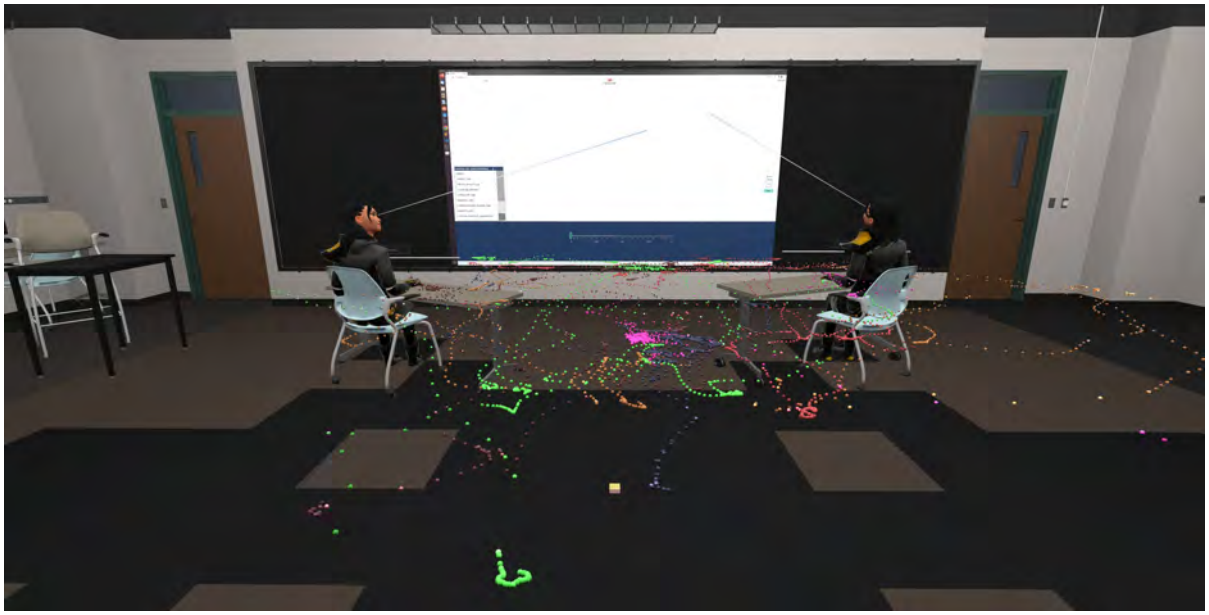


Figure 64: Color-coded paths representing 6 analysts for 3-6 minutes in VR

APPENDIX (Continued)

Figure 65: Color-coded paths representing 6 analysts for 6-9 minutes in MR

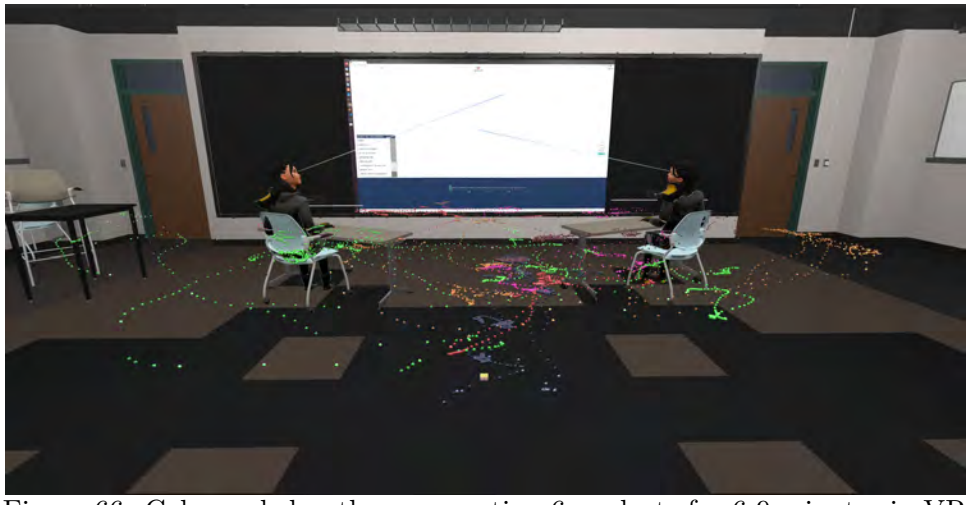


Figure 66: Color-coded paths representing 6 analysts for 6-9 minutes in VR

APPENDIX (Continued)

B.2 Additional Heatmaps from Phase I

The following images show heat maps of space usage for individual analysts from Phase I.

APPENDIX (Continued)

Figure 68: Heatmaps comparing analyst 7's activity in MR and VR for the first 5 minutes

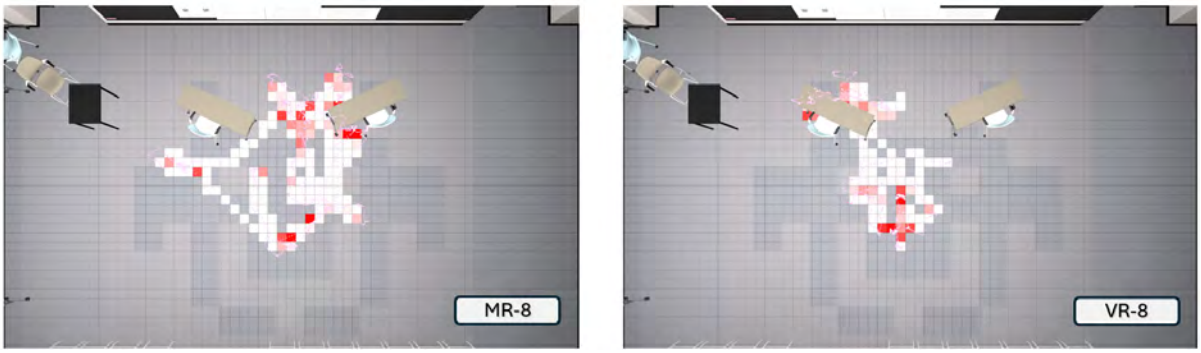


Figure 69: Heatmaps comparing analyst 8's activity in MR and VR for the first 5 minutes

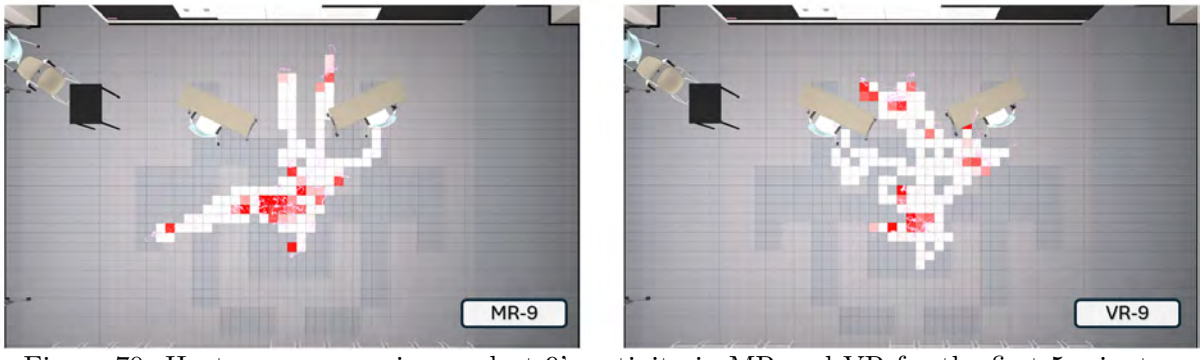


Figure 70: Heatmaps comparing analyst 9's activity in MR and VR for the first 5 minutes

APPENDIX (Continued)

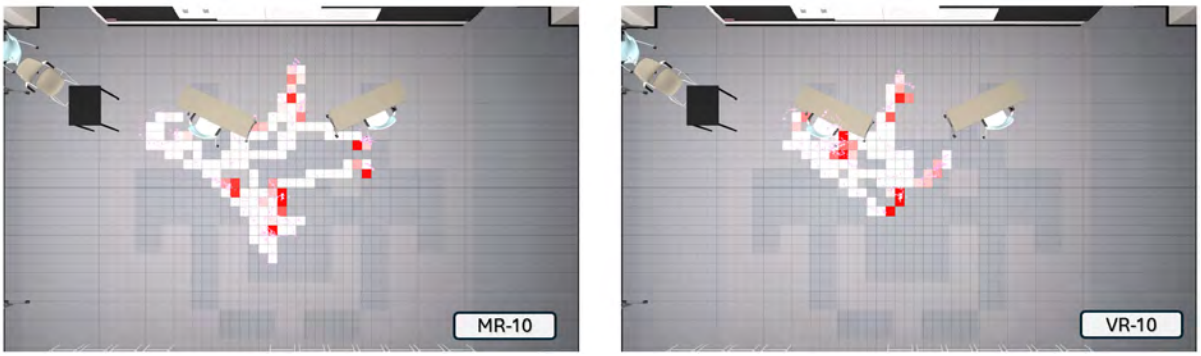


Figure 71: Heatmaps comparing analyst 10's activity in MR and VR for the first 5 minutes

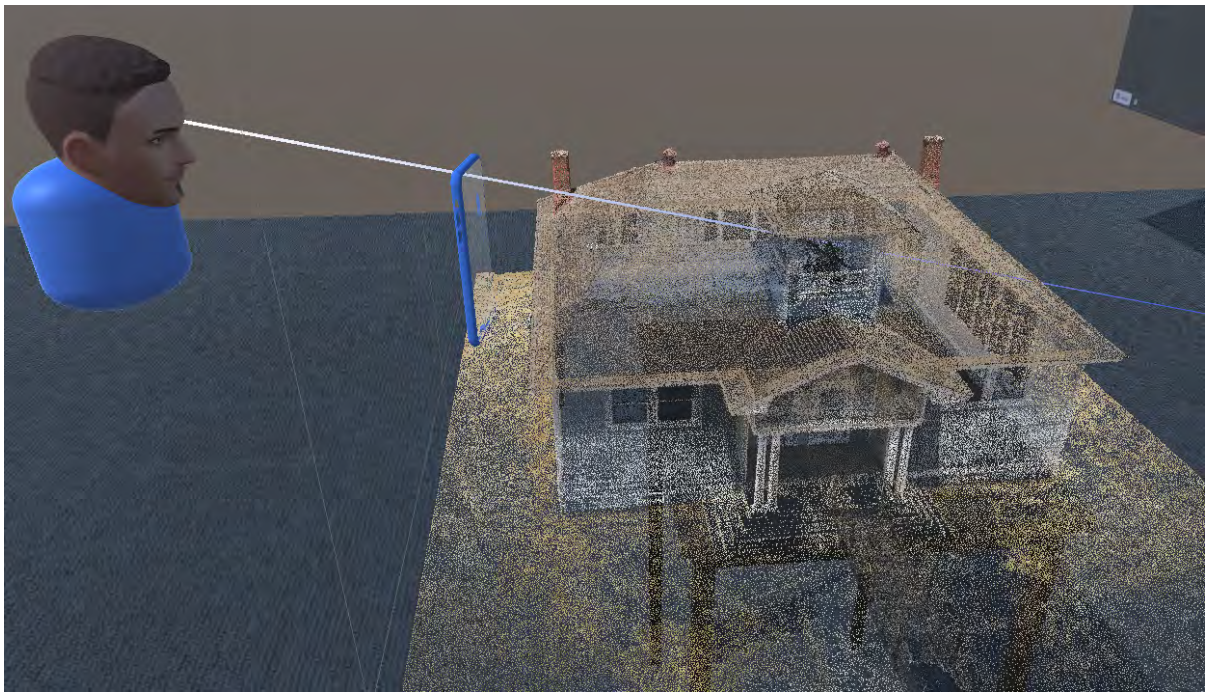


Figure 72: Analyst's view of a participant watching a 1 million point cloud model of Allensworth's school

APPENDIX (Continued)

Figure 73: Analyst's view of a participant watching a textured mesh model of Allensworth's school

Figure 74: Analyst's view of two participants watching a 1 million point cloud model of Allensworth's library

APPENDIX (Continued)

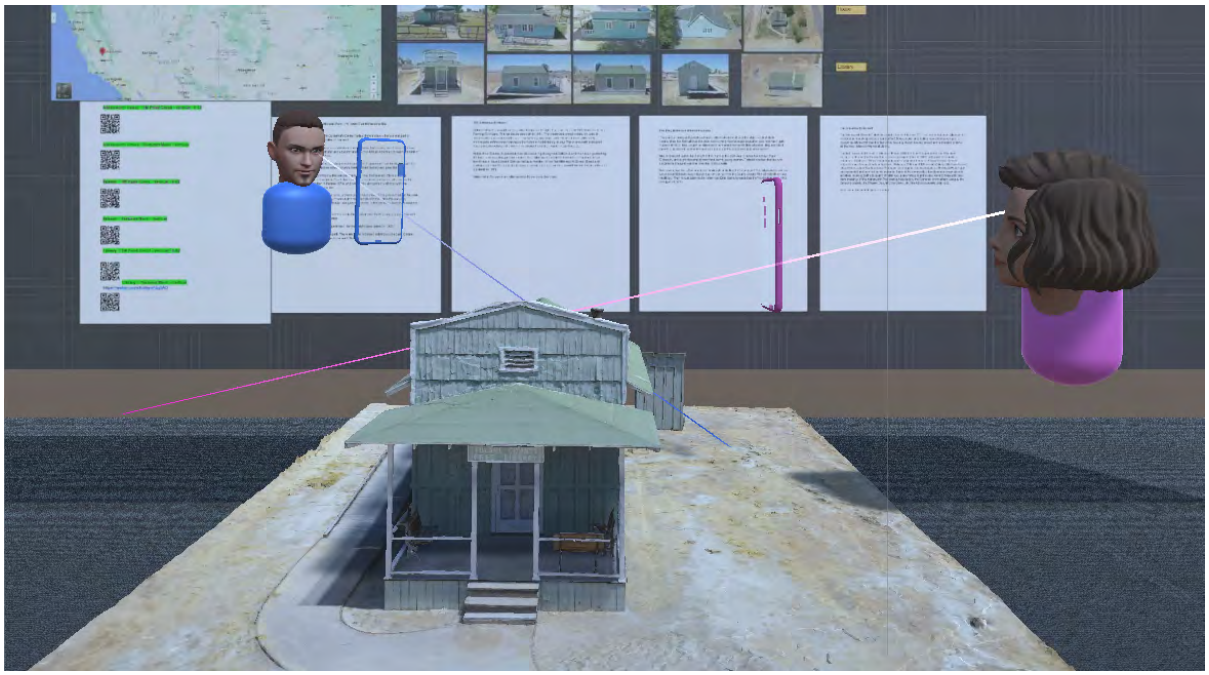


Figure 75: Analyst's view of a participant watching a textured mesh model of Allensworth's library

APPENDIX

IRB-APPROVALS

EXEMPT DETERMINATION

January 30, 2023

Ashwini Naik
anaik3@uic.edu

Dear Ashwini Naik:

On 1/30/2023, the OPRS reviewed the following Claim of Exemption application:

Type of Review:	Initial Study
Title:	Improving Conversation Analysis pipeline through Personal Situated Analytics in Extended Reality
Principal Investigator:	Ashwini Naik
IRB ID:	STUDY2023-0018
Funding:	None
Documents Reviewed:	• Informed Consent.pdf, Category: Consent Form; • Media Consent.pdf, Category: Consent Form; • MeetXR-PostStudyQuestionnaire.pdf, Category: Subject Survey; • MeetXR-Pre Study Questionnaire.pdf, Category: Subject Survey; • Protocol-PersonalSA.pdf, Category: IRB Protocol; • Recruitment.pdf, Category: Recruitment Materials;

The OPRS has determined that this protocol meets the criteria for exemption from IRB review on 1/30/2023.

To document consent, use the consent documents that were approved and stamped by the IRB. Go to the Documents tab to download them.

In conducting this protocol, you are required to follow the requirements listed in the Investigator Manual (HRP-103), which can be found by navigating to the IRB Library within the IRB system.

This determination applies only to the activities described in this submission and does not apply should any changes be made. If changes are made and there are questions about whether these activities impact the exempt determination, please submit a Modification to OPRS for review.

Sincerely,
Charles Hoehne
UIC Office for the Protection of Research Subjects

EXEMPT DETERMINATION

October 6, 2023

Andrew Johnson
3129963002
ajohnson@uic.edu

Dear Andrew Johnson:

On 10/6/2023, the IRB/OPRS reviewed the following Claim of Exemption application:

Type of Review:	Initial Study
Study Title:	Evaluating PSA: A Framework for Conversation Analysis using Situated Analytics in Extended Reality
Principal Investigator:	Andrew Johnson
Study ID:	STUDY2023-1074
Exempt Category:	2
Funding:	None
Documents Reviewed:	• Informed Consent, Category: Consent Form; • Interview Questions, Category: Other; • Post Study Survey, Category: Subject Survey; • Pre-Study Survey, Category: Subject Survey; • Protocol, Category: IRB Protocol;

The IRB/OPRS has determined that this protocol meets the criteria for exemption from IRB review on 9/26/2023.

To document consent, use the consent documents that were approved and stamped by the IRB. Go to the Documents tab to download them.

In conducting this protocol, you are required to follow the requirements listed in the Investigator Manual (HRP-103), which can be found by navigating to the IRB Library within the IRB system.

This determination applies only to the activities described in this submission and does not apply should any changes be made. If changes are made and there are questions about whether these activities impact the exempt determination, please submit a Modification to OPRS for review.

Sincerely,

Office for the Protection of Research Subjects

EXEMPT DETERMINATION – MOD002

January 25, 2024

Ashwini Naik
anaik3@uic.edu

Dear Ashwini Naik:

On 1/25/2024, the IRB/OPRS reviewed the following Claim of Exemption application:

Type of Review:	Modification / Update
Study Title:	Improving Conversation Analysis pipeline through Personal Situated Analytics in Extended Reality
Principal Investigator:	Ashwini Naik
Study ID:	STUDY2023-0018-MOD002
Exempt Categories:	2, 3
Funding:	None
Documents Reviewed:	• Informed Consent, Category: Consent Form; • Interview Questions, Category: Other; • Media Consent, Category: Consent Form; • Post Study Survey, Category: Subject Survey; • Pre-Study Survey, Category: Subject Survey; • Protocol, Category: IRB Protocol; • Recruitment Email, Category: Recruitment Materials;

The IRB/OPRS has determined that this protocol meets the criteria for exemption from IRB review on 1/25/2024.

STUDY2023-0018-MOD002 PI Summary: The following documents have been modified/added:
1. Protocol
2. Media Consent Form
3. Informed Consent Form
4. Pre-Study Questionnaire
5. Post-Study Questionnaire
6. Recruitment Email
7. Interview Questions (Added)

In conducting this protocol, you are required to follow the requirements listed in the Investigator Manual (HRP-103), which can be found by navigating to the IRB Library within the IRB system.

This determination applies only to the activities described in this submission and does not apply should any changes be made. If changes are made and there are questions about whether these activities impact the exempt determination, please submit a Modification to OPRS for review.

Sincerely,

Office for the Protection of Research Subjects

APPENDIX

PART1 - SURVEYS

D.1 Pre-Study Survey

1. Do you have any uncorrected visual impairment?
2. Do you have any uncorrected motor impairment?
3. Please rate your expertise with HoloLens 1?
4. Please rate your expertise with HoloLens 2?
5. Please rate your expertise with Oculus Quest 1?
6. Please rate your expertise with Oculus Quest 2?
7. Have you used an Augmented Reality application on a smartphone before?
8. Have you used an Augmented Reality application on a headset before? If yes, which headset did you use?
9. Have you used a Virtual Reality application before? If yes, which headset did you use?
10. Have you watched any visual content in 3D before?
11. Are you prone to Virtual Reality sickness?
12. Which is your Dominant Hand?

APPENDIX (Continued)

D.2 Post-Study Survey

The survey questions were answered at the end of each of the two parts of the user study. This survey also consists of a subset of the Witmer-Singer Questionnaire[94]. However, as questions 10(a) and (b) were designed to compare users' experiences between the two parts, they were excluded from the Post-Study Survey Part 1.

* Indicates a required question

1. Participant ID *
2. Device Used *

   HoloLens2 OR Quest2
3. The device was comfortable to use.

   Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree
4. On a scale of 1 to 7 please rate the appearance of room size in the application with 1 being very small to 7 being very big in comparison with real room size.

   Very Small 1 2 3 4 5 6 7 Very Small
5. Menu Interaction

   (a) To touch the main menu button I mostly used

   Left Hand or Right Hand

   (b) To touch the buttons (other than the main menu button) I mostly used

   Left Hand or Right Hand

APPENDIX (Continued)

(c) To interact with the slider I mostly used

Left Hand or Right Hand

(d) The buttons were appropriately placed

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

(e) Other Comments about Menu

6. Understanding and analyzing the Data

(a) It was easy to determine important words and phrases in the conversation.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

(b) It was easy to determine important words and phrases in the conversation.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

(c) It was easy to determine important words and phrases in the conversation.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

(d) It was easy to access important words and phrases in the conversation.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

(e) The chat bubbles helped understand the conversation better.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

(f) Being co-located with the participants helped me perceive the conversation better.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

(g) Changing viewpoints helped my analysis.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

APPENDIX (Continued)

(h) The best viewpoint was _____? Please state the reasons if possible.

7. Experience in the environment

(a) I was able to control the appearance of events in the conversation with the help of UI.

Not At All 1 2 3 4 5 6 7 Completely

(b) The environment was responsive to actions that I initiated (or performed)?

Not Responsive 1 2 3 4 5 6 7 Completely Responsive

(c) I was able to anticipate what would happen next in response to the actions that I performed.

Not At All 1 2 3 4 5 6 7 Completely

item I was able to actively survey or search the environment using vision.

Not At All 1 2 3 4 5 6 7 Completely

(d) I was able to closely examine objects.

Not At All 1 2 3 4 5 6 7 Very Closely

(e) I was well involved in the virtual/augmented environment experience.

Not Involved 1 2 3 4 5 6 7 Completely Engrossed

(f) I could concentrate on the assigned tasks rather than on the mechanisms used to perform those tasks?

Not at All 1 2 3 4 5 6 7 Completely

8. Experience with Sounds in the environment

APPENDIX (Continued)

(a) How much did the auditory aspects of the environment involve you?

Not at All 1 2 3 4 5 6 7 Completely

(b) How well could you identify sounds? Not at All 1 2 3 4 5 6 7 Completely

9. Experience with Haptics in the environment

(a) How well could you actively survey or search the virtual environment using touch?

Not at All 1 2 3 4 5 6 7 Completely

(b) How well could you move or manipulate objects in the virtual environment? Not at All 1 2 3 4 5 6 7 Completely

10. Feedback about the application/experience

(a) On a scale of 1-7 please rate your experience in comparison to the first part with 1 being Not good to 7 being very Good.

(b) Please share comments w.r.t. to the above question if any.

(c) What did you learn from the conversation?

(d) What did you most like about the experience?

(e) What did you least like about the experience?

(f) What capabilities would you like added to the application?

(g) Are there applications or areas where using such system would be beneficial? If so, please list or describe briefly.

(h) Any further comments about the application used during this study

APPENDIX

EXPERT EVALUATION - SURVEYS

E.1 Pre-Study Survey

1. Age Group
2. Gender
3. Can you provide some context about the tasks you perform in relation to Conversation Analysis/ related area
4. How long have you been doing this job?
5. What is your main area of expertise in language and social interactional analysis?
6. If other please specify:
7. How do you usually start your work/analysis? Are there any specific steps you need to follow to accomplish a task?
8. What tools or software do you use to complete your tasks?
9. Are there any features you find particularly helpful or frustrating?
10. Do you work with others or collaborate on tasks?
11. Can you describe your physical work environment?

APPENDIX (Continued)

E.2 Post-Study Survey

1. Are there any improvements or changes you would like to see in the tools?
2. Do you have any suggestions for the tool that can help making your tasks more efficient?
3. Are there any additional use cases for the system you just used?
4. How would you envision this workflow evolving in the future?

APPENDIX

PART2 - SURVEYS

F.1 Pre-Study Survey

1. Do you have any uncorrected visual impairment?
2. Do you have any uncorrected motor impairment?
3. Have you used Hololens1 or 2 before?
4. Have you used any Oculus Quest before?
5. Have you used an Augmented Reality application on a smartphone before?
6. Have you used an Augmented Reality application on a headset before? If yes, which headset did you use?
7. Have you used a Virtual Reality application before? If yes, which headset did you use?
8. Have you watched any visual content in 3D before?
9. Are you prone to Virtual Reality sickness?
10. What is your age group?
11. What is your gender? Male/ Female/ Do not wish to answer
12. What is your department/major?

APPENDIX (Continued)

F.2 Post-Study Survey

Modified SUS Scale [12] Questions common to both S1 and S2.

1. I think that I would like to use this workflow to analyze conversations frequently.
2. I found the workflow unnecessarily complex.
3. I thought the workflow was easy to use.
4. I think that I would need the support of a technical person to be able to use this workflow.
5. I found the various functions in this workflow were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this workflow very quickly.
8. I found the workflow very cumbersome to use.
9. I felt very confident using the workflow.
10. I needed to learn a lot of things before I could get going with this workflow.

F.2.1 S1

1. It was easy to determine important words and phrases.
2. It was easy to access important words and phrases.
3. The workflow was easy to use and understandable.
4. It was easy to answer the data-specific questions?
5. What did you most like about the experience?

APPENDIX (Continued)

6. What did you least like about the experience?
7. What capabilities do you think this workflow was lacking?
8. What capabilities do you think helped accomplish the tasks?
9. Any further comments about the application used during this study?

F.2.2 S2

1. It was easy to determine important words and phrases.
2. It was easy to access important words and phrases.
3. The chat bubbles were helpful. How?
4. Can you describe the workflow you used to arrive at answers to data-specific questions?
5. Having mobility in the environment helped me get a better understanding of the data. Why?
6. What is the best point of view for conducting analysis?
7. Would you prefer conducting the analysis through the first workflow (video, transcript, and images), the prototype using HoloLens2, or using both approaches? Why?
8. I was able to observe that the two participants were engaged in exploring the dataset collaboratively.
9. I was able to observe that the two participants arrived at conclusions collaboratively.
10. Being immersed in space gave me more room for exploration.
11. Being immersed in space helped me through the thinking process.

APPENDIX (Continued)

12. Being co-located and getting close to the participants helped me gain various perspectives on the data that could not have been accomplished without it. Why?
13. Being able to freely navigate through the environment improved my perception and understanding of the conversation. How?
14. How did zooming in/out help your analysis?
15. Being embedded and co-located with the participants helped me get a deeper understanding of silences and pauses in the conversation. Why?
16. I preferred to move around to conduct my analysis.
17. I preferred to stand in a particular position to conduct my analysis.
18. I preferred to use a mix of both of the above approaches.
19. Based on the gaze data, what can you say about engagement, agreement, and disagreement in the conversation?
20. Having a see-through mixed reality environment helped the analysis process. How?
21. What are some of the tasks that this environment particularly helped accomplish easily?
22. Where do you think you can integrate this application into your data analysis workflow?
23. Are there any limitations of the device that disrupted your experience?
24. What are some of the limitations of the application that made it challenging to complete a task?
25. Did you have to develop any workarounds during analysis to overcome the limitations of the environment? What were they?

APPENDIX (Continued)

26. What did you most like about the experience?

27. What did you least like about the experience?

28. What capabilities do you think this workflow was lacking?

29. What capabilities do you think helped accomplish the tasks?

30. Any further comments about the application used during this study.

31. The overall experience of Part II was better than Part I.

VITA

NAME Ashwini G. Naik

EDUCATION M.S. in Computer Science, University of Illinois Chicago, 2011
B.E. in Information Science, Visvesvaraya Technological University, 2007

TEACHING Virtual, Augmented and Mixed Reality (CS428, Fall 2022)
CS 425 Computer Graphics I (CS425, Fall 2018)
CS 211 Programming Practicum II (CS211, Summer 2022)

PUBLICATIONS A. G. Naik and A. E. Johnson, "PSA: A Cross-Platform Framework for Situated Analytics in MR and VR," 2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Sydney, Australia, 2023, pp. 92-96, https://doi.org/10.1109/ISMAR-Adjunct60411.2023.00027

Ashwini G Naik and Andrew E Johnson. 2023. Using Personal Situated Analytics (PSA) to Interpret Recorded Meetings. In Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23 Adjunct). Association for Computing Machinery, New York, NY, USA, Article 42, 1–3. https://doi.org/10.1145/3586182.3616697

Naik, Ashwini G et al. "V-NeuroStack: Open-source 3D time stack software for identifying patterns in neuronal data." Journal of Neuroscience Research vol. 101,2 (2023): 217-231.
https://doi.org/10.1002/jnr.25139

A. G. Naik, H. HuynhNguyen, S. Jones, J. Patton and R. V. Kenyon, "Virtual Slots Game for Rehabilitation Exercises," 2019 IEEE Games, Entertainment, Media Conference (GEM), New Haven, CT, USA, 2019, pp. 1-4, https://doi.org/10.1109/GEM.2019.8811546

SfN 2019 - V-NeuroStack: 3-D time stacks for finding patterns in spontaneous activity of neurons in mouse brain slice
https://www.abstractsonline.com/pp8/#!/7883/presentation/49635

PROFESSIONAL EXPERIENCE

- **University of Illinois at Chicago**
  Jan 2017 - Apr 2024
  Graduate Researcher, Electronic Visualization Laboratory

  – Teaching Assistant for CS 428 – Virtual, Augmented and Mixed Reality; CS 425 Computer Graphics; CS 211 Programming Practicum II; AR in journalism.
  – Developed the Personal Situated Analytics (PSA) framework for sensemaking & analysis of conversations for MR and VR using Microsoft HoloLens2 and Meta Quest2 respectively.
  – Conducted research on the role of photorealism in the perceived authenticity of photogrammetric visuals in journalism using Augmented Reality.
  – Created V-NeuroStack application showing dynamic communities using 3D point cloud in Three.js.
  – Conducted research on Machines Assisting Recovery from Stroke (Paper published at IEEE GEM 2019).

- **Epsilon Data Management LLC**
  May 2022 – Aug 2022
  Data Visualization Ph.D. Intern

  – Conducted research on Macro Path to Purchase - Event Sequences.
  – Implemented a new visualization component on a Peta Byte-level data analysis system.
  – Created interactive/responsive web components using state-of-the-art UI design principles.
  – Created API routes to access data for the visualization component.

- **Scientific Games**
  Jul 2012 - Dec 2016
  Senior Test Engineer

  – Set up an Engineering Customer Support team for India location.
    Created training plans and trained associate test engineers.

  - Worked and resolved issues from the field that improved revenue in several casinos across jurisdictions.

- **Olenick and Associates**
  Jul 2011 - Mar 2012
  Client: Chicago Mercantile Exchange
  Automation Engineer

  - Enhanced the existing Automation framework.
  - Maintained the automation testing framework for EOS Trader CME Globex Trading Application.
  - Performed Acceptance and Regression Tests.

- **Birlasoft Ltd.**
  Aug 2007 - Mar 2009
  Bangalore, India; Danbury, CT, USA
  Client: General Electric – Commercial Finance
  Software Engineer

  - Worked in the full life cycle (SDLC) in various stages of the project including analysis, user acceptance testing, SQL data mining, and defect tracking.
  - Provided troubleshooting in various domains across multiple Siebel applications including Oracle databases (Production Servers).