Toward Effective Multimodal Interaction in Augmented Reality

Matt Whitlock
Daniel Leithinger
Daniel Szafir
Danielle Albers Szafir
University of Colorado Boulder
Boulder, CO, USA
{matthew.whitlock,daniel.leithinger,daniel.szafir,danielle.szafir}@colorado.edu

ABSTRACT

Immersive analytics (IA) applications commonly visualize data in AR or VR using stereo rendering and embodied perspective providing new opportunities for data visualization. Efficient IA systems need to be complimented with effective user interfaces. With this position paper, we discuss the importance of effective mapping of interaction modalities to analytics tasks and to prior approaches in previous AR interaction literature. We use this synthesis to identify often overlooked aspects of AR multimodal interfaces. These include transitions between interactions, the importance of field of view, issues with traditional text entry, and employing complementary display types. In identifying these challenges, we hope to facilitate and guide future work toward interaction best practices for IA.

KEYWORDS

Augmented reality; Multimodal interaction; Immersive analytics

ACM Reference Format:
Matt Whitlock, Daniel Leithinger, Daniel Szafir, Danielle Albers Szafir. 2020. Toward Effective Multimodal Interaction in Augmented Reality. In *Proceedings of ACM CHI Workshop on Immersive Analytics (CHI'20).* ACM, New York, NY, USA, 6 pages. https://doi.org/10.475/123_4

CHI'20, April 2020, Honolulu, HI USA

INTRODUCTION

Immersive analytics (IA) applications move beyond mouse and keyboard interaction, offering multimodal data exploration. Effective multimodal interaction (i.e., employing multiple channels for input, such as gesture and gaze) in AR requires sensibly mapping modalities to tasks and considering synergies between modalities. However, we do not yet have concrete guidance for designing effective multimodal AR experiences. In this paper, we synthesize recommended mappings from the HCI literature and identify open challenges for IA interaction. As research in AR multimodal interfaces (MMI) has gone in disparate directions, we discuss how these mappings have taken shape, where they need to be rethought and their relevance to tasks in IA. We posit that: a) legacy bias toward menus and default interaction has hindered optimal mapping of a wider breadth of modalities to tasks, b) many tasks (e.g., interface-level commands & text entry) may be mismapped, and c) considering transitions between interactions and displays in MMIs is critical to the integration of novel techniques. This position paper discusses directions for future AR MMIs for improved usability of IA systems.

TRENDS IN MULTIMODAL AR

To ground our discussion, we synthesized trends in how interaction modalities are mapped to particular tasks. While this is not necessarily a systematic review of the entire design space, our synthesis provides preliminary grounded insight into potential modality $\times$ task mappings. While prior work in IA has begun to empirically explore mappings [2], our synthesis builds on broader AR and HCI interaction literature to provide generalized insight into current knowledge and challenges as they apply to IA. Three common standards emerged in our survey that summarize current practices in IA interaction: the use of freehand gesture for object manipulation, lack of controllers, and popularity of menus.

Freehand Gestures for Transform Manipulation: Freehand gestural interaction has frequently been employed for transform manipulation in AR, allowing users to pick place, scale and rotate objects directly in the physical space. This type of research has seen considerable attention in AR content creation—a common testing bed for multimodal interaction in AR [9, 20]. In this context, users will pick and place virtual objects in the physical environment. In IA, transform manipulations are particularly important as they allow embodied exploration of data through direct manipulation. For example, ImAxes allows users to manipulate the positions and orientations of virtual axes representing data columns to change the 3D visualization layout in VR [6].

While early AR interaction relied heavily on direct gestural interaction, multimodal gaze+gesture use has dramatically increased since the release of the MS Hololens in 2016. The Hololens' default gaze-tap interaction, likened to a mouse click, has dominated modern interaction studies [5] and UI design for AR system contributions [20]. However, this interaction is not grounded in prior interaction or gesture elicitation studies [12], leaving IA systems reliant on interaction modalities that are not optimized for efficient user experiences. IA systems should consider how to overcome the legacy bias introduced by the gaze-tap to explore a fuller space of multimodal interactions for object manipulation.

Lack of Controllers: Use of a handheld remote [16] or video game controller [15] could provide a familiar means of interaction. Despite very heavy usage in VRHMDs, use of handheld remotes in AR interactions research is almost non-existent. We hypothesize there are two possible reasons for this: the fact that recent popular ARHMDs do not have an integrated remote or that AR affords more embodied direct hand manipulation, replacing some of the need for a controller. Gestures and controllers are not mutually exclusive, but constantly holding a controller may impede use of hands freely in concurrent tasks (e.g., manipulating real-world objects). Quantified issues of gorilla arm effect [10] make the haptic feedback and ergonomic comfort of the controller appealing.

Complex visual analytics systems often use many 2D GUI elements for data exploration tasks like filtering data or changing encodings. Within IA, the familiar use of a remote for interaction with 2D GUI elements (as on a television or video game console) may be helpful. Without use of a controller, UIs could consider shortcuts to provide haptic feedback, including attaching 2D menu-based GUIs to physical surfaces [18]. This would allow the analyst to smoothly transition between expressive freehand gestural interaction and constrained, precise 2D touch interaction.

Menus: A staple of the WIMP interaction paradigm (Windows, Icons, Menus, Pointers) is the heavy use of hierarchical menus. These make sense for desktop interfaces and are used in visual analytics tools like Tableau and PowerBI. With integrated dictation recognition, ARHMDs could in many cases replace the need for a menu: users can just say what they would want to select. Though this does not allow for serendipitous discovery of the UI's capabilities, it could be more efficient, particularly for expert users, more accessible, and better tailored to embodied interaction [2]. We hypothesize that with well-integrated voice-based interaction, IA systems could redundantly encode interface level commands typically presented in menus. This would allow analysts to either visually explore options to manipulate the view or describe the action to take using natural language query systems like Eviza [14]. Considering the heavy use of GUI elements such as sliders, radio buttons and menus, IA offers considerable opportunities for understanding how to design multimodal dynamic query systems.

KEY ISSUES FOR MULTIMODAL IA INTERFACES

While the above synthesis identifies key patterns in interaction design, IA offers unique design challenges for immersive and multimodal interaction.

Context Transitions

An important consideration for MMIs is how to manage context switches within the UI. MMI design assumes that particular modalities or combinations of modalities are well-suited to particular tasks (see Badam et al. for a proposed mapping of affordances [2]) However, as analysts explore data, the

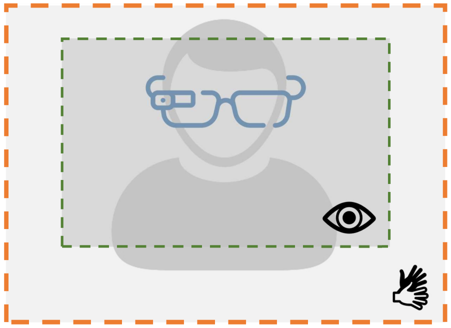

Figure 1: ARHMDs have two fields of vision which users need to manage. Visual search aids should help users deal with limited field of view where virtual content can render. Additionally some visual, audio, or haptic feedback could help users manage the limited field of view for freehand gestures.

tasks they need to engage in may change. For example, MMIs can encourage analysts to switch from primarily gestural interaction when manipulating visualization layouts to primarily voice-based interaction when filtering or changing encodings.

Context switches are also important to managing interactions at different scales. Research in particular techniques for dealing with distance in AR have considered remapping the dimensions of the room [13] and employing raycast techniques [16]. Effective IA displays should leverage these techniques to facilitate task transitions. This could involve transitioning between the typical room display and a remapped room with warped dimensions to bring objects within reach or smoothly transitioning from raycasting techniques to direct hand manipulation. For example, removing a raycast laser or cursor when the user is within arm's reach of data points of interest could encourage users to switch to direct, freehand manipulation. Future MMI research should consider not only the efficiency of interactions themselves but how to design for context switching.

Dealing with Limited Field of View

Though future headsets could resolve the existing limitation of field of view, interaction designers should consider designing for narrow field of view when crafting MMIs. Prior work has explored helping users find specific objects outside of the headset's field of view [3]. However, these techniques focus on finding targets, not encouraging exploration. Solutions could include conveying summary statistics for points outside field of view or guiding users toward unexplored regions of the visualization.

Many headsets make use of outward-facing, on-board sensors for gestural recognition. This design is beneficial as the headset becomes the only wearable equipment needed for freehand interaction. However, this design introduces an angular field of view where the user's hand(s) need to be in order for the headset to pick up the interaction (Fig. 1). Future interfaces relying heavily on gestural interaction should help users to understand when hands are in trackable range and when tracking has been lost. Interface components such as a visual indicator, small vibration, or audio clip could subtly alert users to lost handtracking. This would help users understand system state, circumventing issues where users do not understand why continued hand movement does not affect virtual content or why a gesture performed outside the field of view does not initiate an interaction.

Text Input

Visualization systems often let users freely annotate visualizations or use text for targeted search and filtering; however, text input is a key roadblock to viable interaction in IA. QWERTY-style keyboards have long been the accepted means of text input on desktop displays. Legacy bias has preserved keyboard input in AR despite the lack of haptic feedback and more efficient key layouts [4]. In applying this keyboard metaphor to HMD-based text entry, some have used the haptic affordances of the headset, such as the touchpad on the Google Glass [7], to provide a surface against which to interact.

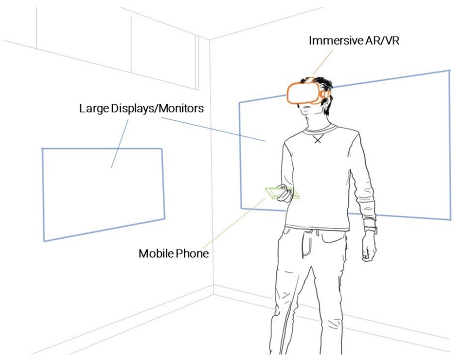

Figure 2: Immersive analytics systems can make use of multiple display types in order to support different analytic tasks and efficient multimodal interaction. For example, manipulations of the dataset may be performed on a mobile phone GUI where the dataset renders in immersive AR or VR.

Others have adapted the keyboard metaphor in novel layouts like a ring [19], better optimized for HMDs but lacking any haptic feedback.

In many cases, voice input may be well-suited to AR text entry; however, systems would inevitably need to support corrections [11]. Given the lack of haptic feedback on the virtual keys themselves when using an HMD and the potential integration with gaze, gestural or remote interaction, this approach to text entry could be an effective solution for HMD text entry.

Multi-display Systems

Just as some tasks are better suited to interaction modalities, tasks could also be suited to different display types (Fig. 2). Within the context of IA, pronounced differences in visualization perception and user behavior indicate tradeoffs of AR, VR and desktop displays for visual analytics [17]. Additionally, sketching interactions for visualization annotation may be more precise on a secondary display to complement expressive but imprecise freehand gestures [1]. With proper consideration of tradeoffs between different immersive displays (similar to the Vistribute framework [8]) and context switches between them, visual analytics tasks could be bolstered by complimentary display types.

CONCLUSION

Despite significant research on AR multimodal interactions, we do not have accepted best practices for IA interaction design. With this position paper, we discuss prominent trends and important considerations of research on AR multimodal interfaces. Continued AR MMI research will need to consider sensible and suitable mappings of modalities to tasks as well as higher level design considerations that will allow for effective switching and viable long-term use. For immersive AR to see widespread adoption for analytics tasks will require continued research to consider the balance of existing work on interactions with proposal of novel interaction methods and paradigms.

REFERENCES

[1] Rahul Arora, Rubaiat Habib Kazi, Tovi Grossman, George Fitzmaurice, and Karan Singh. 2018. Symbiosissketch: Combining 2d & 3d sketching for designing detailed 3d objects in situ. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–15.

[2] Sriram Karthik Badam, Arjun Srinivasan, Niklas Elmqvist, and John Stasko. 2017. Affordances of input modalities for visual data exploration in immersive environments. In *2nd Workshop on Immersive Analytics*.

[3] F. Bork, C. Schnelzer, U. Eck, and N. Navab. 2018. Towards Efficient Visual Guidance in Limited Field-of-View Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics* 24, 11 (Nov 2018), 2983–2992.

[4] Pieter Buzing. 2003. Comparing different keyboard layouts: aspects of qwerty, dvorak and alphabetical keyboards. *Delft University of Technology Articles* (2003).

[5] Han Joo Chae, Jeong-in Hwang, and Jinwook Seo. 2018. Wall-based Space Manipulation Technique for Efficient Placement of Distant Objects in Augmented Reality. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. 45–52.

[6] Maxime Cordeil, Andrew Cunningham, Tim Dwyer, Bruce H Thomas, and Kim Marriott. 2017. 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*. 71–83.

[7] Tovi Grossman, Xiang Anthony Chen, and George Fitzmaurice. 2015. Typing on glasses: adapting text entry to smart eyewear. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 144–152.

[8] Tom Horak, Andreas Mathisen, Clemens N Klokmose, Raimund Dachselt, and Niklas Elmqvist. 2019. Vistribute: Distributing Interactive Visualizations in Dynamic Multi-Device Setups. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[9] Sylvia Irawati, Scott Green, Mark Billinghurst, Andreas Duenser, and Heedong Ko. 2006. " Move the couch where?": developing an augmented reality multimodal interface. In *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*. IEEE, 183–186.

[10] Sujin Jang, Wolfgang Stuerzlinger, Satyajit Ambike, and Karthik Ramani. 2017. Modeling Cumulative Arm Fatigue in Mid-Air Interaction Based on Perceived Exertion and Kinetics of Arm Motion. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3328–3339. https://doi.org/10.1145/3025453.3025523

[11] Jiepu Jiang, Wei Jeng, and Daqing He. 2013. How Do Users Respond to Voice Input Errors? Lexical and Phonetic Query Reformulation in Voice Search. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '13)*. Association for Computing Machinery, New York, NY, USA, 143–152.

[12] Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. 2013. User-defined gestures for augmented reality. In *IFIP Conference on Human-Computer Interaction*. Springer, 282–299.

[13] Jing Qian, Jiaju Ma, Xiangyu Li, Benjamin Attal, Haoming Lai, James Tompkin, John F Hughes, and Jeff Huang. 2019. Portal-ble: Intuitive Free-hand Manipulation in Unbounded Smartphone-based Augmented Reality. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 133–145.

[14] Vidya Setlur, Sarah E. Battersby, Melanie Tory, Rich Gossweiler, and Angel X. Chang. 2016. Eviza: A Natural Language Interface for Visual Analysis. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. Association for Computing Machinery, New York, NY, USA, 365–377. https://doi.org/10.1145/2984511.2984588

[15] M. E. Walker, H. Hedayati, and D. Szafir. 2019. Robot Teleoperation with Augmented Reality Virtual Surrogates. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 202–210. https://doi.org/10.1109/HRI.2019.8673306

[16] M. Whitlock, E. Hanner, J. R. Brubaker, S. Kane, and D. A. Szafir. 2018. Interacting with Distant Objects in Augmented Reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 41–48.

[17] M. Whitlock, S. Smart, D. A. Szafir, S. Kane, and D. A. Szafir. 2020. Graphical Perception for Immersive Analytics. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*.

[18] Robert Xiao, Julia Schwarz, Nick Throm, Andrew D Wilson, and Hrvoje Benko. 2018. MRTouch: adding touch input to head-mounted mixed reality. *IEEE transactions on visualization and computer graphics* 24, 4 (2018), 1653–1660.

[19] W. Xu, H. Liang, Y. Zhao, T. Zhang, D. Yu, and D. Monteiro. 2019. RingText: Dwell-free and hands-free Text Entry for Mobile Head-Mounted Displays using Head Motions. *IEEE Transactions on Visualization and Computer Graphics* 25, 5 (May 2019), 1991–2001. https://doi.org/10.1109/TVCG.2019.2898736

[20] Ya-Ting Yue, Yong-Liang Yang, Gang Ren, and Wenping Wang. 2017. SceneCtrl: Mixed Reality Enhancement via Efficient Scene Editing. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)*. ACM, New York, NY, USA, 427–436. https://doi.org/10.1145/3126594.3126601