Using Virtual Reality Effectively in Scientific Data Exploration - Perception, Usability and Design in Immersive Displays

by

Johannes Novotny

Department of Computer Science

Brown University

A dissertation submitted in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy
in the Department of Computer Science at Brown University

Providence, Rhode Island

May 2021

This dissertation by Johannes Novotny is accepted in its present form by
the Department of Computer Science as satisfying the dissertation requirement
for the degree of Doctor of Philosophy.

Date: ____________________ ____

David H. Laidlaw, Advisor
Brown University

Recommended to the Graduate Council

Date: ____________________ ____

Doug A. Bowman, Reader
Virginia Polytechnic Institute
and State University

Date: ____________________ ____

James Tompkin, Reader
Brown University

Approved by the Graduate Council

Date: ____________________ ____

Andrew G. Campbell
Dean of the Graduate School

Curriculum Vitae

Johannes Novotny is the son of Anton and Sylvia Novotny. He graduated from the Vienna University of Technology with a Bachelors of Science in Medicine and Computer Science (2007) and a Master of Science in Visual Computing (2013). He worked as a research scholar at the Visualization and Interface Design Innovation Group, University of California, Davis until 2014, when he joined the Ph.D. program at the Brown University Computer Science Department. He earned his Masters of Science there in 2018, and successfully defended his dissertation in May 2020.

Acknowledgements

I would like to thank my colleagues at Brown University and other institutions whose ideas, feedback, and support have been critical for the completion of this dissertation. In particular, I would like to thank Joshua Tveite, Wesley Miller and Fumeng Yang of the Brown visualization research lab for their collaboration in discussing new visualization concepts, adding to our pool of VR applications, and pulling things together for publication. Likewise, I am grateful to Stephen Gatesy and Morgan Turner from Brown's Ecology and Evolutionary Biology Department, as well as Fritz Drury from the Rhode Island School of Design for their valuable feedback and enthusiasm in our dinosaur footprint project. This gratuity also extends to the Brown and RISD students of the course Virtual Reality Design for Science, who were instrumental in completing this work.

I would also like to acknowledge our collaborators from the Rhode Island Hospital, Francois I. Luks, Derek Merck and Scott Collins for bringing their interest and expertise into our surgical planning pilot project. From Virginia Tech, I want to thank Wallace S. Lages and Bireswar Laha, who were a tremendous help in forming my initial ideas for this project and performing our first experiments on visual fidelity components.

I am particularly grateful to my advisor David Laidlaw for his insightful advice and our countless fruitful discussions that guided me through my research work at Brown. This work wouldn't have been possible without his mentorship and his effort to provide a nurturing research environment. My appreciation also goes to the readers of this dissertation, Doug A. Bowman and James Tompkin for their support throughout my time working on this dissertation.

Finally, I would like to thank my friends and family for their endless support in solving academic and, importantly, non-academic problems. Thanks to all members of the board game, movie, softball groups for being amazing companions throughout this journey.

To my wife, Eun Young, thank you for your encouragement and patience over the past seven years. I could not have reached this point without you.

Attributions

David H. Laidlaw is a Professor of Computer Science and head of the Visualization Research Lab at Brown University. As my PhD advisor, he was my collaborator and co-author on all papers forming this dissertation.

Wesley Miller and Joshua Tveite are Masters in Computer Science and Research Assistants at Brown University's Visualization Research Lab. Both have contributed to the execution of the user study on VR Fidelity Components presented in Chapter 4. Wesley Miller also contributed to the design of the user study on medical data exploration in Chapter 5 and was co-author in the resulting publication [Novotny et al., 2020]. Joshua Tveite contributed to the implementation of our VR particle visualization application and was co-author in the resulting publication [Novotny et al., 2019].

Wallace S. Lages is Assistant Professor of Computer Science at the Virginia Tech School of Visual Arts. He contributed to the design and execution of the user study on VR Fidelity Components presented in Chapter 4 and presented the initial study results at IEEE VR [Lages et al., 2016].

Chapter 5 is based on:

J. Novotny, W. R. Miller, F. I. Luks, D. Merck, S. Collins, and D. H. Laidlaw. Towards Placental Surface Vasculature Exploration in Virtual Reality. *IEEE Computer Graphics and Applications*, 40(1):28-39, Jan-Feb 2020.

Chapter 6 is based on:

J. Novotny, J. Tveite, M. L. Turner, S. Gatesy, F. Drury, P. Falkingham, and D. H. Laidlaw. Developing virtual reality visualizations for unsteady flow analysis of dinosaur track formation using scientific sketching. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):2145-2154, May 2019.

This work was in part supported by the National Science Foundation under grant no. IIS-1319606.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

"A picture is worth a thousand words", a simple proverb with many interpretations. In the study of human perception, it leads to a host of questions, as not every picture conveys the same amount of information it is of importance to identify which qualities of an image have an impact in this information transfer. Vision is the dominant sense with which most humans perceive their environment and our brains have adapted to reliably process large quantities of incoming optical stimuli. Our intuitive understanding of visual information has been used to great effect throughout the ages, for example in the form of paintings and sculptures that can capture the nuances of a moment in visual poetry. However, classic artworks are ultimately static projections of reality. Advances in computer-display technology have allowed us to break the limitations of static images. First through 2D Desktop displays and now via Virtual Reality (VR) devices, we can now create dynamic visual stimuli representing a plausible environment users can directly interact with.

Since their inception by Sutherland [1968], VR devices have seen steady development. What started as a heavy ceiling-mounted headset showing vector-graphics cubes, aptly named the Sword of Damocles, was iteratively improved upon to reach the Head Mounted Displays (HMDs) and CAVE (CAVE Automatic Virtual Environment) Display rooms we know today (see Figure 1.1.

During their development, the specific features that make up a VR display were improved and reworked to provide increasingly accurate visual representations of virtual worlds. On the hardware side resolution, field of view and other characteristics of displays have improved from small windows into virtual worlds to stereoscopic visualizations covering a users entire visual field. At the same time improvements in 3D position tracking allows us to create VR scenes that cover large walkable areas without the restrictions of cables and other tethering. While advances in graphics programming allows us to create highly plausible looking simulations of real world scenes that provide users with an accurate perception of being physically present (immersed) in the presented

(a) Sword of Damocles [Sutherland, 1968]   (b) Oculus Quest 2 [2020]

Figure 1.1: The progress of over 50 years in VR hardware development, from Sutherland's Sword of Damocles installation, to the untethered Oculus Quest 2 Head Mounted Display (HMD).

environment.

Many fields have much to gain from accurate immersive simulations, for example medicine, which uses them to train new practitioners and to enhance therapeutic treatments. However, not every VR visualization needs such a high level of realism to be an effective way of conveying information to users. Which led Bowman and McMahan [2007] to ask "how much immersion is enough?" for a given task. In their work they spearheaded research into how hardware and software aspects of VR displays benefit user immersion in a virtual environment, and how these benefits interact to enhance application effectiveness for users. Multiple works have since shown the advantages of increased immersion on certain data exploration tasks, yet one major problem was always the generalizability of such results to multiple VR devices. This problem forms the motivation for the first sections of this thesis, which explore the effects of several VR display fidelity on user performance to examine the cross-platform validity of such experiments. Here our goal is to uncover guidelines for the effective use of VR devices to aid users in scientific data exploration.

Apart from studying perception aspects of VR, it is also important to consider user requirements and practical applications of immersive visualizations. Domain scientists may now have access to VR devices, but not every part of their research workflow necessarily benefits from their use. Entering data into a spreadsheet for example might currently be even more difficult in an immersive environment when compared to a 2D desktop visualization, while analyzing spatially complex 3D biological data is significantly eased by such immersion [Laha et al., 2014].

Due to the relative novelty of VR systems, domains scientists are often unaware of their advantages. This highlights the need for inter-disciplinary collaborations between visualization researchers and other academic fields. The latter sections of this thesis are dedicated to the study specific applications of VR, their benefits they provide to users, and the process of designing immersive visualizations in general.

1.1 Thesis Statement

This brings us to the central thesis of this dissertation:

The study of human perception within VR displays can lead to insights benefiting the creation of effective and intuitively usable immersive visualizations.

The research presented in this work aims to advance the understanding of visual fidelity and interaction aspects of VR displays with regards to scientific data exploration. Our specific contributions are as follows:

- **Evaluation of Effects and Interactions of Visual Fidelity Components**
  In multiple experiments we evaluated how hardware and software features of VR visualizations influence user effectiveness and behavior during immersive data exploration. These features included display resolution, field of view and specific rendering choices within applications.

- **Insights into the Generalizability of VR Displays**
  In our experiments we studied multiple competing VR devices to verify the generalizability of our outcomes and gain insights into cross-platform effects applicable to VR vision studies in general. The Brown YURT [Kenyon et al., 2014], a CAVE-like system with a display resolution at the limit of human perception and an artifact-minimizing screen layout was our main host systems for experiments. The high display-fidelity of the YURT allowed us to evaluate visual conditions beyond the current capabilities of wearable hardware and therefore draw conclusions about HMD devices yet to come.

- **Studying Applications of Immersive Visualization**
  In two case studies we evaluated how immersive scientific visualizations could help domain scientists not previously acquainted with VR displays. We report our insights from observa-

tions and user feedback and lay out requirements and design considerations for the development of novel VR applications.

- **Providing Actionable Guidelines for VR Use and Development**
  Finally we combined insights from our vision experiments and a long-term of the *Scientific Sketching* application design methodology into a set of recommendations and guidelines for the effective use of VR in scientific visualization, particularly considering the trade-offs of different display characteristics and how they should inform visualization design.

1.2 Outline

The outline of this dissertation is as follows. In chapter 2, we will provide a more thorough background on existing VR display systems, their specific hardware properties and overview of the concept of fidelity components from which we draw insight and inspiration. Chapter 3 will provide a detailed description and results of our experiments to evaluate the influence of specific display fidelity characteristics on user performance in scientific data exploration. Chapters 4 and 5 shift the focus on evaluating human perception and interaction within VR systems through experiments analyzing text reading performance and feature detection in immersive visualizations. Chapter 6 presents the results of a large scale design study evaluating the *Scientific Sketching* design methodology and the novel immersive flow visualization techniques gained from it. Chapter 7 lists a summary of our research contributions, describes potential avenues of future work, and ends with the general conclusions of this dissertation.

Chapter 2

Background

The field of VR research has been very active over the past few years. Hardware breakthroughs in both, traditional Cave Automatic Virtual Environments (CAVE) [Kenyon et al., 2014, Papadopoulos et al., 2015] and portable Head-Mounted Displays (HMD) have enabled new ways for media consumption and content creation [Roberts et al., 2014]. Naturally, these displays are of growing interest to the scientific visualization community. The wider availability of VR devices makes it possible to integrate immersive data analysis and exploration into a wide field of scientific disciplines and opens up new possibilities for visualizing and interacting with data.

While VR technology has been under development since the 1960s [Sutherland, 1968], up until the last few years only a small selection of fields used immersive displays on a larger scale. A prime example is the medical field where VR has been successfully employed since the early 1990s. Applications range from educational and therapeutic systems to medical image analysis tools and training simulations [Pensieri and Pennacchini, 2014]. Positive impacts of virtual reality and robotics on surgical planning have been summarized by McCloy and Stone [2001], who also highlighted the effectiveness of practice in simulated environments on surgical outcomes. For similar reasons, VR technology was also adopted in military applications early on for the simulation, training and planning of combat operations as well as uses in guided field surgery [Lele, 2013].

The advent of sufficiently powerful and affordable VR displays, opened up their use from immersive simulations to the more general field of scientific visualization. Initial pushes towards these broader applications of VR technology started in the late 1990s. Articles by Bryson [1996] and Brooks [1999] introduced the potential benefits of immersive systems over regular desktop setups and called for a wider adoption in the sciences.

Over the course of continuous improvements, VR systems have since proven their effectiveness in data visualization of a variety of fields, such as biology [Laha et al., 2014], archeology [Kim et al., 2015], astronomy [Hanula et al., 2015], geology [Caravaca et al., 2020] and data science [Donalek et al., 2014], to name a select few examples. In this thesis we focus on tasks and requirements falling into the field of 3D scientific visualization, particularly on visualizations of datapoints with direct correspondence to 3D locations in the real world (e.g. medical scan data and 3D flow fields).

2.1 Scientific Visualization and Immersion

Several studies have shown the benefits of virtual reality environments on the analysis of complex 3D datasets. Immersive systems can give a better understanding of spatial relationships [Schuchardt and Bowman, 2007, Laha et al., 2014], 3D paths [Ragan et al., 2013], and shapes [Demiralp et al., 2006]. These properties are especially useful in those engineering and scientific domains that require users to deal with datasets containing intricate spatial structures. Tasks involving the analysis of data from computerized tomography (CT) scans or magnetic resonance imaging (MRI) of biological samples [Laha et al., 2014, Zhang et al., 2001], for example, can be very challenging.

These initial results establish the beneficial effects of VR displays, yet the principles underlying these performance improvements are not yet well understood. Uncovering the relationship between display characteristics, interaction methods, and user performance is critical for the development of novel visualization techniques. Additionally it would provide valuable input for upcoming VR hardware generations [Swan et al., 2003].

Visualization researchers have approached the analysis of display properties by directly comparing user performance between VR systems [Demiralp et al., 2006, Forsberg et al., 2008]. Even though this approach can help to establish usage guidelines [Swan et al., 2003], the results are harder to generalize to new VR platforms and are less likely to expose the underlying cause of the observed effects [Laha and Bowman, 2012]. To overcome these issues it is necessary to analyze the influence of specific characteristics of VR systems on user performance.

In his work on general terminology of perceived presence in VR, Slater [2003] discussed how the visual fidelity of a VR system might affect the level of immersion experienced by its users. He listed a set of features that could provide a way to objectively compare VR systems on specific hardware and software characteristics. These features, termed visual fidelity components by Bowman and McMahan [2007][Bowman and Raja, 2004, Bowman et al., 2012], include:

- *Field of View* (FOV), the size of the visual field that the display covers during use.
- *Field of Regard* (FOR), the total size of the visual field that can be covered by the display, taking user movement and head orientation into account.
- *Display Density*, the denseness of pixels on the screen area.

- *Stereoscopy*, whether the display provides monoscopic or stereoscopic images.

- *View-dependent Rendering*, whether or not, and to which degree the VR system takes the viewers head and body location into account during rendering.

- and *Image Quality*, a combination of specific rendering choices (e.g. texturing and lighting) and hardware capabilities (e.g. display flickering) affecting the fidelity of displayed images.

Additional fidelity components, like 3D sound and haptic displays, are actively studied by perception researchers. These non-visual components are however beyond the scope of this thesis.

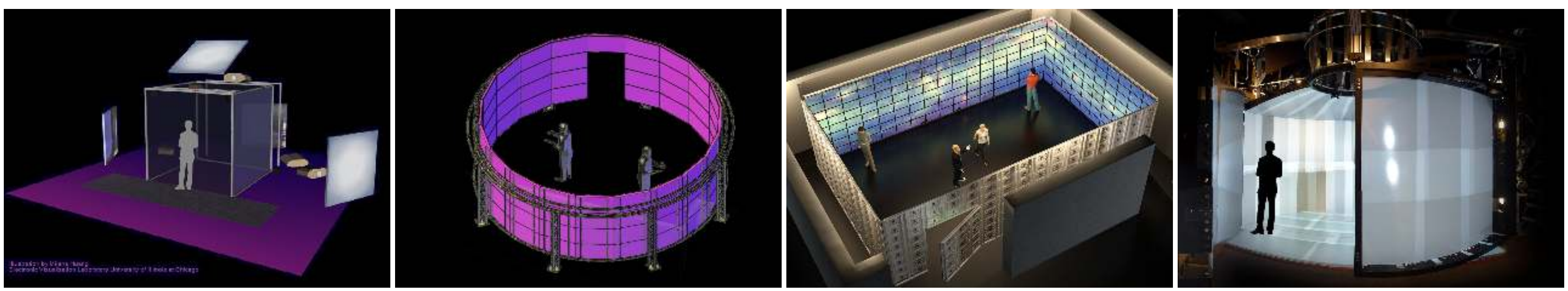
A number of studies have already shown how fidelity components affect a users sense of presence within a scene. Cummings and Bailenson [2016] surveyed 83 experimental evaluations and found that stereoscopy, head-tracking and field of view were of particular importance for the creation of immersive scenes. However, as Bowman and McMahan [2007] remarked earlier, heightened perceived presence does not necessarily entail improved performance in data exploration tasks. Studies by [Zhang et al., 2001] as well asLaha et al. [2014], have already established some effects of stereoscopy and head-tracking on exploration accuracy in a variety of tasks. A detailed description of this line of research can be found in section 3.1. In this thesis we extend the set of evaluated fidelity components and also address the problem of generalizability between VR systems and architectures.

Modern VR systems of the HMD and CAVE varieties can vary greatly in their provided hardware and software features. CAVE displays in particular are often built with a particular task in mind, leading to fidelity trade-offs during their construction. The Reality Deck [Papadopoulos et al., 2015] for example does not use stereo displays to prevent a reduction in display contrast. Figure 2.1 shows several representative examples of CAVE display rooms, to exemplify the range of different screen layouts, while Table 2.1 compares a set of their fidelity components to the human eye and a current HMD device. The variety of possible feature combination underlines the importance of the experimental studies performed in the remained of this thesis, in order to understand how to effectively use VR visualizations independent of specific display hardware.

	Eye	CAVE [Cruz-Neira et al., 1993]	CAVE 2 [Febretti et al., 2013]	Reality Deck [Papadopoulos et al., 2015]	YURT [Kenyon et al., 2014]	HTC Vive Pro [2018]
Architecture	-	CAVE	CAVE	CAVE	CAVE	HMD
Resolution [arcmin/pixel]	1	4	1	$<<1$	1	2.8
Stereo display	yes	yes	yes	no	yes	yes
FoV (horizontal) [°]	160	140	140	160	140	110
FoR (horizontal) [°]	$\pm$ 180	$\pm$ 135	$\pm$ 180	$\pm$ 180	$\pm$ 180	$\pm$ 180
FoR (vertical) [°]	$\pm$ 90	+45, -90	+15, -35	$\pm$ 30	+75, -90	$\pm$ 90

Table 2.1: Comparison of visual fidelity components between multiple VR display systems.

∞



(a) CAVE Cruz-Neira et al. [1993] (b) CAVE 2 Febretti et al. [2013] (c) Reality Deck Papadopoulos et al. [2015] (d) YURT Kenyon et al. [2014]

Figure 2.1: Schematic representations of the four CAVE environments described in Table 2.1.

Chapter 3

Effects of VR Fidelity on Volume Exploration Performance

Understanding the advantages of using VR displays in scientific data exploration is key to build effective immersive visualizations. In this chapter we focus on evaluating the effects of specific fidelity aspects of VR displays on the performance of users engaged in the visual analysis of volumetric iso-surface visualizations.

The analysis of volumetric data is a common task in a variety of scientific fields. Volume datasets can be obtained from medical imaging (e.g. Computer Tomography, Ultrasound), geologic/geographic measurements (e.g. Ground Penetrating Radar, Lidar) or many other many other domain specific imaging methods. It can also be the result of computer simulations or abstract 3D visualizations of information visualization data. One common feature of this kind of data is spatial complexity, which makes it difficult to represent on 2D desktop screens and therefore a prime target for immersive VR visualizations. In this chapter, we present our evaluation of user effectiveness in visual analysis tasks of volumetric biological datasets in VR environments. Our experiment focuses on the stereo and FOR capabilities of three different VR displays, one head-mounted display (HMD) and two CAVE displays [Cruz-Neira et al., 1993]. We chose to follow the experimental design of Laha et al. [2014], as it attempts to cover a large variety of different exploration scenarios. Basing our work on an existing study, but re-performing it on multiple new VR platforms allows us to and find effects applicable to VR systems in general, independent of individual display architectures.

The contributions of this study are insights into:

- **Effects of varying fidelity conditions on exploration task performance.**
  Here we confirmed that the stereo fidelity component has a significant beneficial effect on task completion times across all tested platforms.

- **Influence of user interaction strategies**
  User behavior differed across different hardware systems in response to hardware limitations of given systems, which influence performance metrics.

- **Design recommendations for application and study design**
  Our gathered quantitative and qualitative results were compiled into a list of actionable recommendations for future experiments and VR visualization development.

Initial findings of this work were published in collaboration with Wallace S. Lages, Bireswar Laha, Wesley Miller, Johannes Novotny, David H. Laidlaw, John J. Socha and Doug. A. Bowman at IEEE VR 2016 under the title "Effects of field of regard and stereoscopy and the validity of MR simulation for visual analysis of scientific data" [Lages et al., 2016].

3.1 Related Work

Several recent studies have attempted to characterize the effects of immersive environments on complex data analysis tasks in detail. Early works focused on comparing user performance in different display archetypes. Qi et al. [2006] compared how response time and accuracy varied in an HMD, fish tank VR, and fish tank VR with haptics using four generic tasks in the visualization of volumetric data; their results indicated that HMD users had larger response time and error than the other platforms. Schuchardt and Bowman [2007] reported significant benefits of a CAVE environment over a projection wall on the task performance of complex search tasks in spatially complex underground cave structures. Forsberg et al. [2008], in an empirical experiment comparing the effectiveness of desktop, fishtank VR, and CAVE systems in analyzing volume-rendered confocal microscopy datasets, reported that more immersive VR environments provided significant benefits in user performance. Contrary to these results, Demiralp et al. [2006] presented study results for an abstract visual search task that favored fishtank VR, a system with a lower level of immersion, over a CAVE setup. However, while direct comparisons among VR systems provide valuable data for establishing basic usage guidelines [Swan et al., 2003], it is difficult to tie their results to individual fidelity components of the evaluated VR systems. This drawback makes it hard to generalize their results to VR systems outside the original experimental designs [Laha and Bowman, 2012].

Addressing these limitations, research shifted towards the experimental evaluation of individual fidelity components of VR displays Bowman and McMahan [2007]. These experiments were mostly based on mixed reality (MR) simulations, which simulate VR displays with lower levels of system fidelity within host-displays of higher fidelity [Bowman et al., 2012]. Such controlled simulations provided a more rigid experimental setup and provided deeper insights into the effects of visual fidelity components.

In multiple studies, Laha et al. evaluated the effects of FOR, stereoscopy, and head-tracking on user performance in volume data analysis tasks using CAVE environments [Laha et al., 2012, 2014] or HMDs [Laha et al., 2013]. Their experiments covered a wide variety of data-exploration tasks in direct volume rendering and isosurface representations of scientific datasets. Among their results, they reported benefits of enabled stereo and active head-tracking on the task speed of participants. They also found significant FOR effects in tasks involving volumetric rendering. A related study by Ragan et al. Ragan et al. [2013] on small-scale spatial judgment tasks within underground cave datasets reported similar effects.

Work by Chen et al. evaluating the effects of stereoscopy and display size on user performance in the analysis of DT-MRI data reports main effects of stereoscopy on task completion time as well as combined effects of size and stereoscopy on task accuracy. Surprisingly, users performed better in monoscopic display conditions, a finding that the authors attributed to the participants' familiarity with lower-fidelity displays Chen et al. [2012].

In addition to hardware-dependent fidelity components, several researchers have analyzed how rendering quality and application performance affect users of VR systems. In an experiment using an HMD display, Lee et al. [2013] compared a virtual environment with varying levels of visual realism to an augmented reality visualization. They reported only minimal differences in participant performance, further supporting the validity of MR simulations. Studying the trade-off between DVR image quality and framerate, Hänel et al. [2016] report the best user performance in a system that adaptively lowered rendering quality to maintain a high framerate; this system also ranked highest in subjective participant preference.

In this paper, we further evaluate the effects of stereoscopy and FOR on performance in complex volumetric data exploration, as these are fidelity components that still vary between modern VR systems. Our aim is to overcome generalizability limitations found in prior research, which often evaluated only a single VR system. Building upon the results of Laha et al. [2014], our study combines MR simulation with a multi-platform experimental setup to gather results applicable to a wide range of VR applications. To reduce the number of experimental conditions, we chose to not evaluate the head-tracking fidelity component, as modern immersive VR systems vary more often in stereo (e.g. monoscopic phone-based AR/VR applications, stereo HMDs) and FOR characteristics (e.g. VR tables, CAVEs, display walls) than in head-tracking capabilities. Our decision was also informed by Laha's initial results, which showed fewer head-tracking effects on task performance.

3.2 Experiment

We studied how stereo and FOR affect user performance in three different HMD and CAVE virtual reality display platforms. Our main goals were to analyze the potentially beneficial effects of higher-fidelity display conditions, and investigate if matching effects can be found across multiple VR systems. This lead to the following sub-hypotheses:

- *H1. Participants will perform better in higher-fidelity conditions than in lower-fidelity conditions.*
  Prior experiments provide evidence that specific data exploration tasks can be solved more effectively in high-fidelity VR environments, with stereo capabilities showing a more pronounced beneficial effect than increased FOR. H1 aims to verify previous results and to extend them to new VR systems. Our main point of reference is a study by Laha et al. [2014] on which our experiment is based.

- *H2. Participants in conditions with the same FOR and stereo assignments will achieve similar performances across platforms.*
  With this hypothesis we aim to evaluate if the results of fidelity experiments are consistent between VR systems with different hardware characteristics and display architectures. Similarity of performance metrics between groups of the same FOR/-stereo condition across VR platforms would would prove the generalizability of of VR display-fidelity experiments. However, we expect a platform-specific offset between CAVE and HMD performance results.

3.2.1 Experimental Design

We performed a between-subject experiment with three independent variables: VR platform, stereoscopy, and FOR. Each platform was evaluated using two stereoscopy levels, stereo and mono, as well as two field of regard levels, 90° and 270°. This resulted in a factorial design with 12 different conditions. Table 3.1 lists all conditions and their respective group numbers.

Participants were asked to perform fifteen exploration tasks, each of which required them to search for, and estimate features in a biological dataset. To prevent any task from giving insight for future answers, they were presented in a specific, predefined order. Before trial data from these tasks were collected, each participant performed five training tasks using a different but structurally similar dataset. This allowed participants to acclimate themselves to the system without introducing learning effects. A written script was followed to ensure consistent phrasing during the whole experiment to minimize variations between experimental sites.

Table 3.1: Group numbers of our twelve experimental conditions.

FOR	Stereoscopy	CAVE A	CAVE B	HMD
90	Stereo	1	5	9
90	Mono	2	6	10
270	Stereo	3	7	11
270	Mono	4	8	12

We chose to focus our experimental tasks on the exploration of spatially complex volumetric isosurface visualizations. To achieve results generalizable to other VR systems, we decided to explore a wide variety of potential user interactions with these VR visualizations. The fifteen selected tasks cover the majority of user task categories of visual volume data analysis based on the taxonomy introduced by Laha et al. [2015]. The core concepts of our tasks; visual search, pattern recognition, spatial understanding, quantitative estimation, and shape description, are relevant to numerous volumetric visualization applications with similar underlying data. Examples include MRI scans of blood vessels and spatial network datasets. We believe that these task concepts are also relevant for other VR visualization methods, such as direct-volume rendering with sharp object boundaries.

3.2.2 Apparatus

The three VR display platforms used in this experiment were two four-sided CAVEs located at Brown University and Virginia Tech, and an nVisor SX111 head-mounted display (HMD). These systems represent common VR setups at the time of our data collection. Table 3.2 presents a general overview of the hardware characteristics of each system. CAVE A and B are cubic display systems with side lengths of about 3m (10ft) and 2.4m (8ft) respectively; three rear-projected walls each, a front and two sides; and front-projected floors. Passive stereo is provided in CAVE A using INFITEC wavelength multiplexing stereo filters (Fig. 3.1a), while active stereo is provided in CAVE B using CrystalEyes LCD-shutter glasses (Fig. 3.1b). Each wall screen in CAVE A and CAVE B has a resolution of 1920×1920 pixels, for a horizontal and vertical angular pixel size of 2.81 arcmin when standing in the center. The effective angular resolutions are equal because when standing in the center of the CAVE the field of view towards the screen is the same for each, despite the larger size of CAVE A.

The nVisor HMD places two separate small screens in front of the user's eyes, working on a similar principle to consumer HMDs like the Oculus Rift and the HTC Vive (Fig. 3.1c). The HMD has a resolution of 1280×1024 pixels per eye, its horizontal angular pixel size is 4.78 arcmin and

Table 3.2: Overview of hardware characteristics of the three VR environments

	CAVE A	CAVE B	HMD
Horizontal Pixel Size (arcmin)	2.81	2.81	4.78
Vertical Pixel Size (arcmin)	2.81	2.81	3.75
Field of View	$110° \times 80°$	$110° \times 75°$	$102° \times 64°$
Accommodation Distance (m)	˜1.5	˜1.2	∞
Stereo Technology	Passive	Active	Separate Screens
Tracker latency [ms]	4	8.3	4
Stereo Technology	Passive	Active	Separate Screens
Headpiece Mass (g)	85	79	1300
Seams	Visible	Visible	None
User Body Visibility	Visible	Visible	Not visible

its vertical angular pixel size is 3.75.

In order to track user head positions, both CAVE A and the HMD are equipped with an ultrasonic wireless Intersense IS-900 system; CAVE B uses an OptiTrack optical tracking system. These systems provided 6-DoF tracking of headpieces and the wand tools used for interaction. The tracking attachments were mounted on top and side of the glasses used in CAVEs and to the front of the HMD. Both tracking systems operated on a 120 Hz update rate.

Participants could grab or release the dataset using a button on the wand, permitting them to rotate the dataset about or translate it along any axis. Special care was taken to make the interaction comparable on all both platforms; the dataset and the environment were scaled identically, the front trigger button was used for interaction on all wand tools, and the available physical space was equivalent. We used a cluster of four machines with Nvidia Quadro FX 5800 GPUs to render images in CAVE A and a single machine with two Nvidia Quadro K5000 GPUs to render the images in CAVE B. One machine with an Nvidia Quadro FX 5800 graphics card was used for the HMD. All platforms ran the same software for 3D isosurface rendering, Meshviewer, from KeckCaves lab [Kreylos, 2017], built using the Vrui VR Toolkit [Kreylos, 2008] on Linux. General application performance was equivalent between all experimental conditions with frame rates above 60 FPS. This was confirmed by a personal inspection of the Brown setup by the study leader at Virginia Tech.

The horizontal FOR was controlled by displaying two black walls that extended from the participant's head to form the desired condition, either 90° or 270°. The walls translated with each user's head but maintained a fixed orientation in space and did not rotate (Fig. 3.2). The vertical

(a) The CAVE A environment located at Virginia Tech. It uses magnetically tracked INFITEC passive stereo glasses.

(b) The CAVE B environment located at Brown University. It uses optically tracked CrystalEyes stereo shutter glasses.

(c) The nVisor SX111 HMD system located at Virginia Tech.

Figure 3.1: The three VR systems used during the experiments. The two CAVE systems differed in size (CAVE a being larger), stereo, and tracking technology. See Table 3.2 for a list of detailed VR system characteristics.

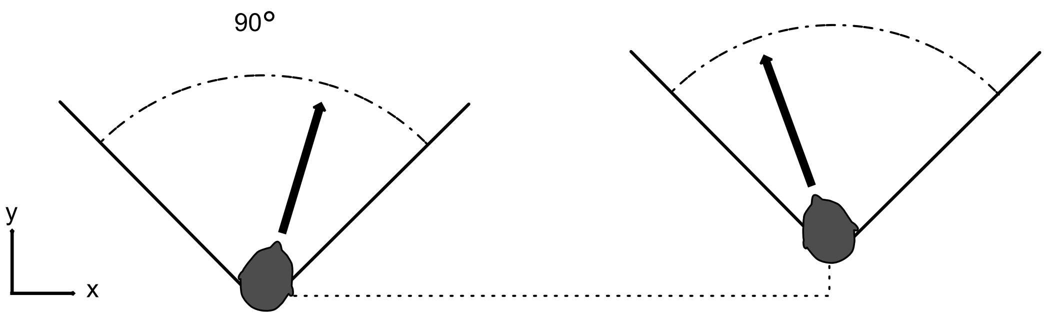


Figure 3.2: Schematic of the virtual walls used to control field of regard (in this example 90 degrees). The virtual walls translate with the user's head but do not rotate.

FOR was not controlled; however, we observed that users in both the CAVEs and the HMD used mostly the wall areas, and the CAVEs have no ceilings, so this resulted in an effective vertical FOR of approximately 90°. Mono conditions were produced by rendering both images from a single camera position.

By nature, the HMD occludes the physical body of participants from their vision, while they remained visible in CAVE systems. The HMD headpiece (1300g) was also heavier than the glasses worn in CAVE A (85g) and those worn in CAVE B (79g). Because the screens in the HMD are fixed relative to the user's eyes, there are no visible seams; by contrast, the seams between walls of the CAVEs are visible to users. Each display used a different method to provide stereo view (separate screens in HMD, passive stereo in CAVE A, and active stereo in Cave B) with varying accommodation distances. The optical lenses of the HMD places the accommodation distance at infinity, while in cave systems it is the distance of the viewer from the projection screen; approximately 1.5m (5ft) in Cave A and 1.2m (4ft) in Cave B. These differing accommodation distances could have led to differing levels of eye strain and fatigue onset [Hoffman et al., 2008].

3.2.3 Datasets

In the experiment we used two datasets of carabid beetles obtained through Synchrotron Radiation Microtomography (SR-µCT); datasets were acquired with a bending magnet beamline (2-BM) at the Advanced Photon Source of the Argonne National Laboratory. A scan of a *Pterostichus* beetle was used for training, and one of a *Platynus* beetle was used for testing (Fig. 3.3). These beetles are commonly called ground beetles and are of interest to the biomechanics community due to the peculiar dynamics of their respiratory system [Socha et al., 2008, Socha and De Carlo, 2008].

The respiratory system of beetles and other insects consists of several branching tubes (tracheae and tracheoles) emerging from openings in the surface of the insect's body called spiracles.

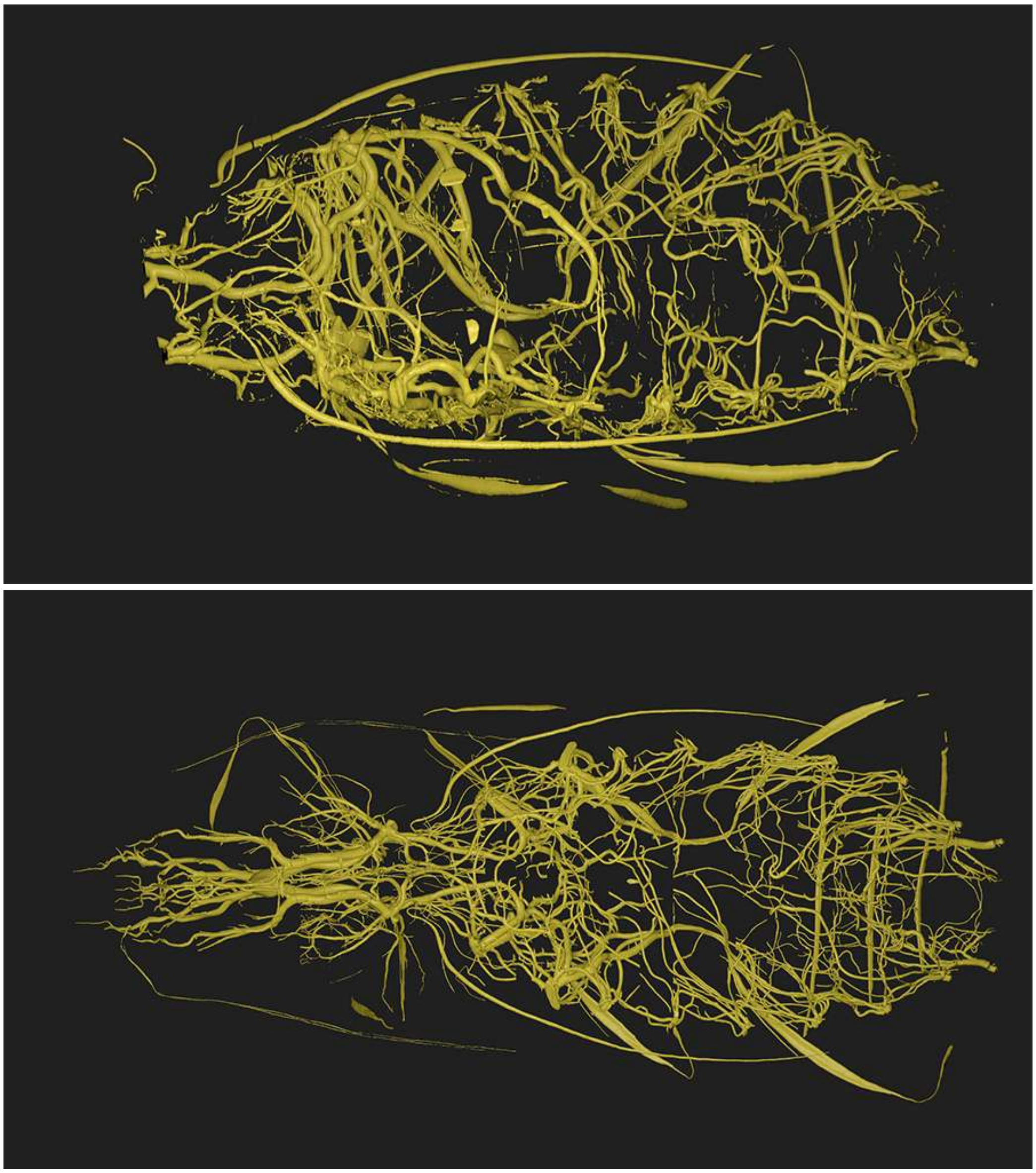


Figure 3.3: Overview visualization of the isosurfaces used in the study. Top: *Pterostichus* (used for training), Bottom: *Platynus* (used for main study).

These tube networks can be very complex and difficult to understand, largely because of the sheer amount of branching, twists, and occlusion in the data. For this experiment, a mesh of the tracheal system was created from raw scans using automatic segmentation in Avizo[1], followed by manual refinement.

3.2.4 Tasks

To maintain ecological validity, we asked a collaborating biomechanical researcher to review the list of common tasks performed in studying these datasets used by Laha et al. [2014]. Most often, analysis involves looking for symmetries, counting or estimating structures, and looking into topological features like branching and connection between structures. The tasks used in the experiment mimic questions that the researcher has when studying a specimen for the first time, for instance:

> "Does the number of spiracles on the left side match the number of spiracles on the right side? If not, what is the difference?"
>
> "Which region of the body appears to have the highest volume density of tracheal tubes?"

Each task consisted of listening to a question from the experimenter, analyzing the dataset for the answer, and answering the question orally. Because some tasks required a basic understanding of insect anatomy, we instructed participants on how to identify the head, thorax, and abdomen, as well as the dorsal and ventral sides of the beetle, tracheal tubes, and spiracles. An abbreviated description of all tasks used in this experiment can be found in the Appendix.

3.2.5 Metrics

For each task, we measured the completion time, accuracy, the user's confidence in his/her answer, and the user's subjective assessment of the question's difficulty. In order to measure the time, we defined as the start time the instant the participant acknowledged the instructions. After that, the participant was told to begin the interaction. The task ended after the participant said "done." The time between these two moments was measured in rounded seconds, using a regular stopwatch. Since most tasks took approximately one minute, potential errors due to the reflex time of the experimenter were not significant. In addition, to prevent participants from beginning a task early, we asked them to turn away from the dataset while they received the instructions. The accuracy metric was computed by grading each answer against a score sheet created by the domain expert.

[1] avizo3d.com

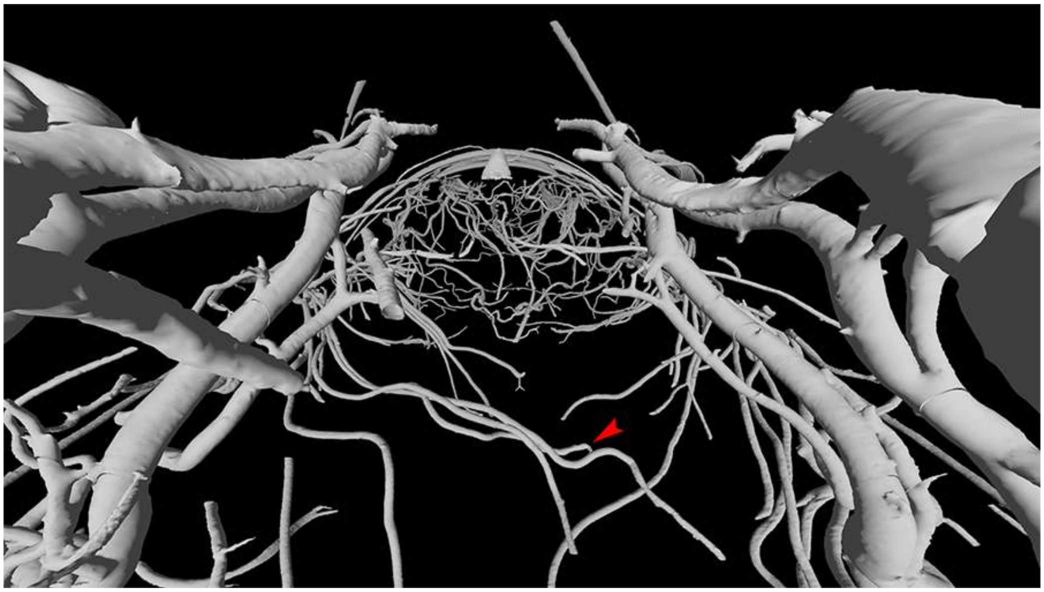


Figure 3.4: Example of a participants' view on the dataset during the experiment. In this trial participants were asked to follow the tube marked by the red arrow to its end point, a spiracle below the beetles exoskeleton. Tracing was done visually, without required wand interaction.

Possible accuracy values varied from 0% to 100%, with multiple intermediate values for tasks with non-binary answers (e.g. counting tasks). We also collected subjective confidence and difficulty ratings reported orally after each task, using a 7-point Likert scale [Ekstrom et al., 1979]. These metrics were chosen to stay in line with previously published studies by Laha et al. [2012, 2014] and allow for a comparison of results.

3.2.6 Participants

We recruited 60 participants, 19 females and 41 males, from the campus communities at Brown University and Virginia Tech. Announcements were made on mailing lists and social media, and volunteers were distributed into the 12 groups. The participants were from different backgrounds and had ages ranging from 18 to 63, with a median age of 25. None of them reported prior knowledge of biomechanics or experience in the analysis of volume datasets. Participation in the study was voluntary, with no incentive given. All participants had an equivalent level of experience with the data and tasks. We provided them with all the required domain knowledge to complete each trial tasks during the instruction and training phase of the experiment.

3.2.7 Procedure

Upon arrival, participants were asked to read and sign an Informed Consent Form describing the experiment and their rights as subjects. They then filled in a background questionnaire on demographic information such as age, gender, familiarity with computers, and prior knowledge in biomechanics. After completing an initial spatial ability test [Ekstrom et al., 1979], participants received an introduction to the experiment, explaining what they would see, as well as the details of the equipment and how to operate it. We also educated the participants on basic insect anatomy and functioning of the respiratory system, using diagrams and photos of beetles and tracheal scans.

When they were ready, participants stepped into the CAVE or put on the HMD and began a series of five training tasks. The training tasks were similar to the tasks included in the main study but instead used the *Pterostichus* training dataset. During this phase, the participants were educated on how to use the wand and report the answers, confidence, and difficulty for the tasks. They were also instructed on how to move around and manipulate the dataset to get better views. Figure 3.4 shows an example of how a task would look like from a participants point of view.

After a short break, the participants continued to the main part of the experiment, completing 15 exploration tasks within the *Platynus* dataset. Participants were allowed to begin each task only after the corresponding task instructions had been read aloud by the experimenter and confirmed as understood by the participant. For each task, after the participant claimed to have found the answer by saying, "done," the experimenter stopped the timer and recorded the response along with the time spent carrying out that task. Next, the experimenter asked about the perceived difficulty of the task and the user's confidence in his/her answer. The participant then reported a number from 1 to 7 for each metric. Before starting the next task, the view was reset to a standard position. Participants using the HMD were offered a 3-minute break in the middle of the task sequence.

After completing the tasks, participants answered a final questionnaire designed to capture their overall experience and any specific strategy they may have used. The study was approved by the Institutional Review Board at Brown University and Virginia Tech.

3.3 Results

We performed data analysis on three levels. First on the complete dataset to find high-level effects found across all platforms, to evaluate our hypothesis on the benefits of high-fidelity VR conditions. Then we analyzed the data split by VR device to identify platform-specific effects between the two differently-sized CAVEs with magnetic and optical tracking respectively, and the HMD (See system characteristics in Table 3.2), investigating the similarity of participant performance between

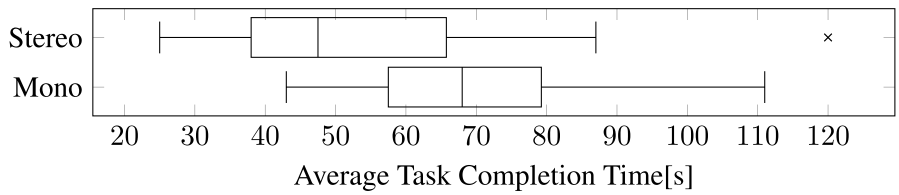


Figure 3.5: Boxplot showing the effect of stereo conditions on average task completion time (Whiskers indicate highest and lowest values within 1.5 interquartile range of the corresponding quantiles, datapoints outside this range are marked as outliers). Participants were significantly faster in stereo on conditions.

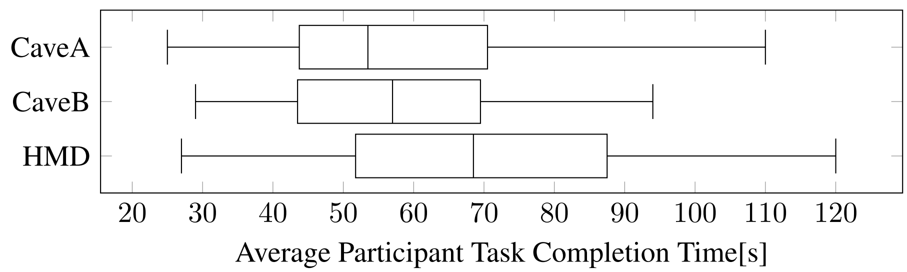
different display types. Finally, we analyzed individual tasks to draw connections to previously published results by Laha et al. [2014]. In the following sections we discuss all statistically significant findings. Comprehensive result tables of the statistical tests can be found in the supplemental material.

3.3.1 Summary Results

To understand the data and verify our hypothesis, we conducted a three-way mixed model ANOVA for the time metric and fit a generalized linear mixed-effects model (GLMM) to the ordinal accuracy metric, as well as the user-reported difficulty and confidence scores. These methods allowed us to take our fixed effects (stereo, FOR, platform) and the random effects of our experimental setup (participants, tasks) into account. Questionnaire responses were collected only once per participant were analyzed using proportional odds logistic regression (POLR), a regression model for ordinal dependent variables such as Likert responses [Venables and Ripley, 2002]. All tests in this study assumed a significance level of 5%. Testing these summary effects allowed us to combine data from multiple conditions to increase the overall statistical power of our findings.

We found a significant main effect of the stereo condition on task completion time ($F(1) = 20.754$, $p < 0.001$) across the entire dataset (Fig. 3.5). Participants in stereo conditions completed tasks significantly faster than those in mono conditions. A similar effect was found in the GLMM on self-reported difficulty ($Estimate = -0.990$, $Std.Err. = 0.345$, $z = -2.869$, $p = 0.004$). Here participants reported lower difficulty in stereo conditions. We did not find significant effects of stereo on accuracy or self-reported confidence.

Task accuracy was only affected by the display platform. The GLMM showed a significant main effect of platform on accuracy ($Estimate = -0.686$, $Std.Err. = 0.276$, $z = -2.484$, $p = 0.013$). Pairwise comparisons showed that participants in the CAVE A system were significantly less accurate than those on the other two platforms (Fig. 3.6). Better scores on the initially spatial



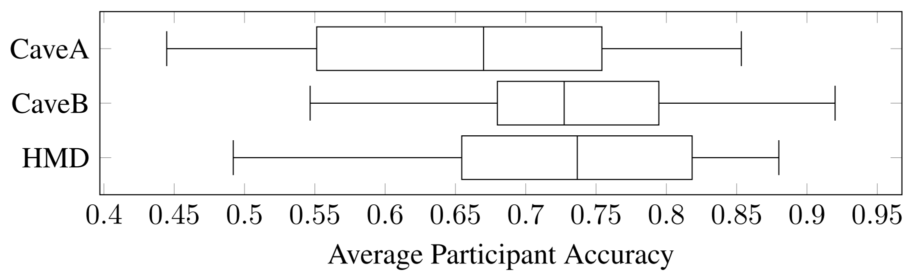


Figure 3.6: Boxplots showing the overall performance distribution of participants for each platform; (top) average task completion time, (bottom) average accuracy per participant over all 15 tasks. While HMD users appear to be slower no statistical significance between groups was found. We found statistical significance between groups in the accuracy metric. In pairwise tests, participants in the CAVE A achieved significantly lower accuracy that the ones in CAVE B and HMD.

ability scores of participants did show statistically significant effects on any of the other collected metrics.

Finally, analysis of the feedback collected in the post-experiment questionnaire showed effects on self-reported wand interaction ($LR\chi^2 = 10.228$, $Df = 2$, $p = 0.006$) and walking responses ($LR\chi^2 = 15.878$, $Df = 2, p < 0.001$). For both we found significant effects of VR platform, with the participants of CAVE A reporting that they interacted with the dataset more by walking and with less wand actions than participants of the other systems (Fig. 3.7). Participants in stereo conditions also reported a higher score for movement interactions ($LR\chi^2 = 12.954$, $Df = 1, p < 0.001$). For the wand interaction metric we found an additional interaction of FOR and Stereo ($LR\chi^2 = 5.636$, $Df = 1, p = 0.018$).

3.3.2 Platform Results

The collected accuracy results and completion times by platform are visualized in Figure 3.8. We plotted average participant accuracy over all 15 tasks over the corresponding average completion time for each participant, coded by their experiment condition. Points in the top-left area of the plot have higher performance levels (higher accuracy, shorter completion times), while points in the

Table 3.3: Anova Results - Significant effects on time (all tasks)

Response	Source	Df	F ratio	Pr(> F)
Time	Stereo	1	20.754	< 0.001
Log(Time)	Stereo	1	11.803	0.001

Table 3.4: GLMM Results - Significant effects on accuracy and perceived difficulty (all tasks)

Response	Source	Estimate	Std. Error	z value	Pr($> Z$)
Accuracy	Platform (CAVE B)	-0.686	0.276	-2.484	0.013
Difficulty	Stereo	-0.990	0.345	-2.869	0.004

lower-right have worse performance (lower accuracy, longer completion times).

In this stage we evaluated whether the effects found in the overview analysis could be found on the tested platforms individually and if there are differences between matching FOR and stereo conditions across platforms. We followed the same procedure of statistical tests we used for the overall summery analysis, i.e. three-way mixed model ANOVA for time (Table 3.3) and GLMM for accuracy and user-reported metrics (Table 3.4). We tested our hypotheses by directly comparing subsets of our twelve experimental conditions. For H1 (higher fidelity conditions improve performance), we performed six pairwise comparisons within each platform. For H2 (in matching FOR and stereo conditions task performance will be similar between platforms), we compared the groups with the same levels of FOR and stereoscopy from each platform. To keep the analysis concise we focus our report on collected time and accuracy data.

Within individual platforms we found effects as expected based on the summary analysis. On

Table 3.5: POLR Results - Significant effects on walking and wand interaction (all tasks)

Response	Source	LR χ^2	Df	Pr($> \chi^2$)
Wand interaction	Platform	10.228	2	0.006
	FOR:Stereo	5.636	1	0.018
Walking	Stereo	12.954	1	< 0.001
	Platform	15.878	2	< 0.001

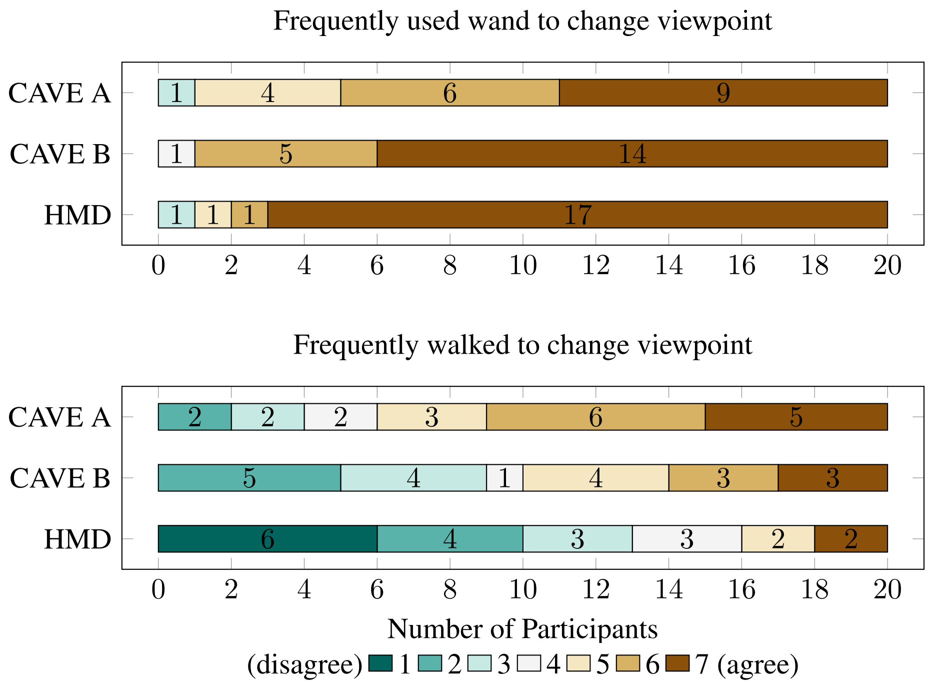


Figure 3.7: Distribution of participant feedback on the frequency of (top) wand interaction and (bottom) walking interaction based on the platform.

both CAVE platforms we found an effect of stereo on task completion time (Cave A: $F(1) = 9.415$, $p = 0.007$; Cave B: $F(1) = 11.779$, $p = 0.003$). However, on the HMD platform this was not significant (Fig. 3.9). We did not find any significant effects of stereo or FOR conditions on accuracy in any of the evaluated platforms.

3.3.3 Task Results

Finally we followed the analysis steps laid out by Laha et al. [2014] and analyzed effects of display conditions on the level of individual tasks. For the direct comparisons between condition groups in specific tasks we used Cliff's method [Cliff, 2014] with an extension to control the familywise type-I error (Hochberg's method). Cliff's method is a robust, rank-based, non-parametric method that tests the hypothesis that groups have identical distributions, similar to the Wilcoxon-Mann-Whitney test. It is an appropriate method for small numbers of participants per group. This let us relax the assumption of normality and homoscedasticity without losing much statistical power.

The pairwise analysis between experiment conditions revealed differences that were not consistent across platforms. Table 3.6 summarizes the pair-wise differences found across conditions for time and accuracy metric. While we found multiple significant effects on time and accuracy

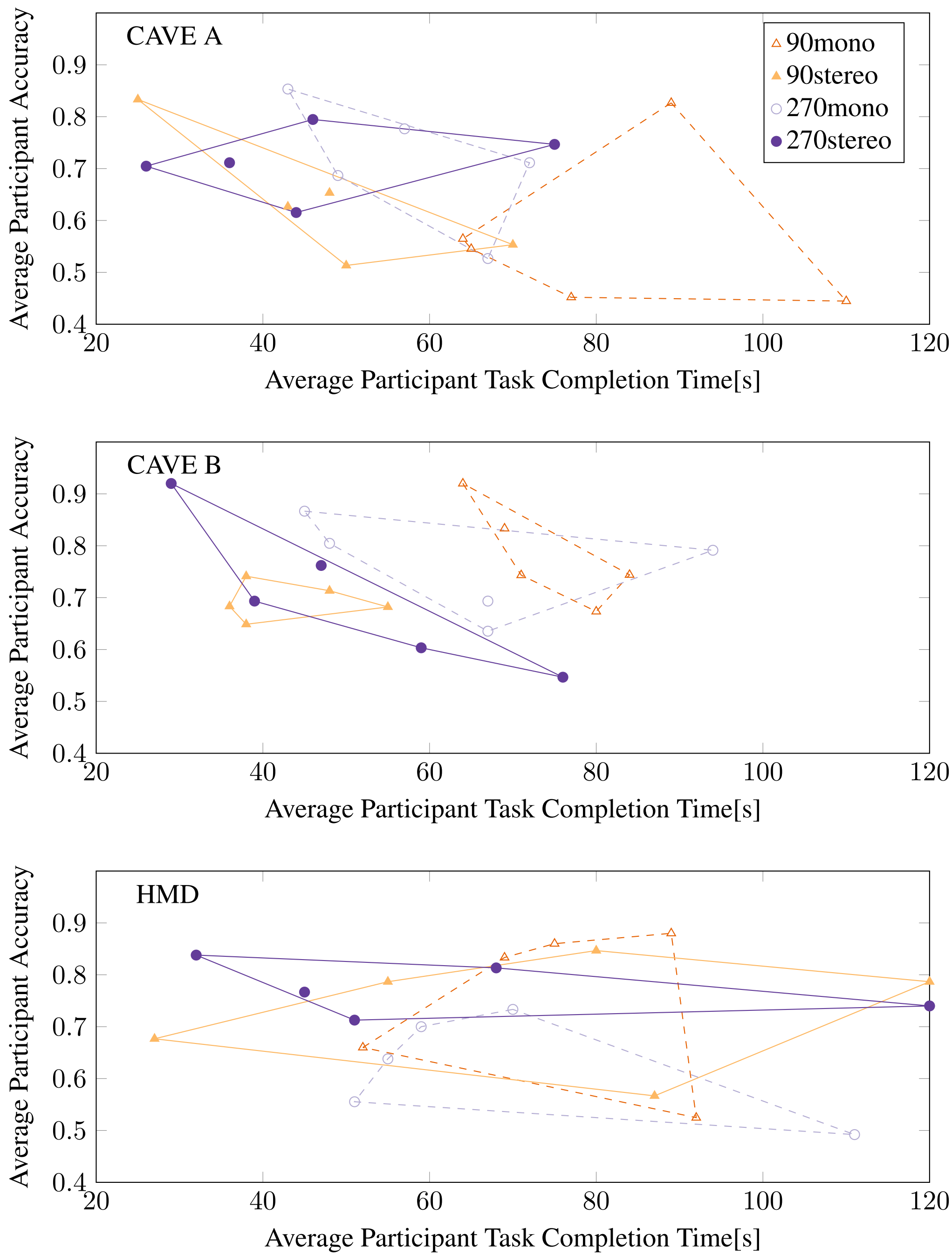


Figure 3.8: Distribution of average completion time and accuracy results of each participant over all 15 tasks, separated by platform and experimental condition. (top) CAVE A, (middle) CAVE B, (bottom) HMD. To visualize the variability in each group the convex hulls around their respective data points are plotted.

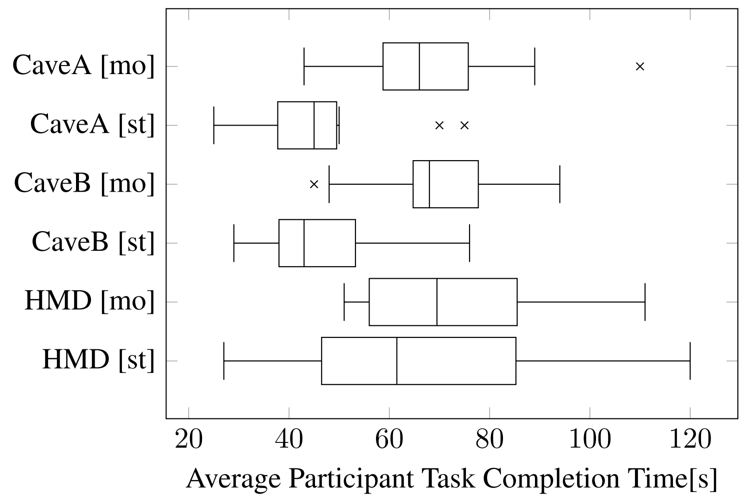


Figure 3.9: Box plot visualizing the platform-specific effects of stereo on average task completion time over all 15 tasks. Stereo improved the completion time on all three platforms, but statistical significance was only found for the two CAVE systems.

in each platform individually, we found no corresponding effects on other platforms for the same task. For example, in Task 03 we found that in CaveB, participants in the stereo 90 condition took significantly longer to complete the task than those in the mono 90 condition, while in the HMD we observed the opposite effect. Implications of these results are discussed in Section 3.4.3. Numerical results of the statistical tests can be found in the supplemental material.

3.4 Discussion

In this section we discuss how the significant effects found in the analysis of our collected metrics relate to our two initial hypotheses. While we have found evidence that participants achieve better performance in high-fidelity conditions (H1), the experiment revealed significant differences between platforms (H2) on the level of specific tasks. As our results deviate from our expectations based on prior work [Laha et al., 2014], we also discuss possible causes of these outcomes by coupling them with collected qualitative feedback.

3.4.1 H1 - Beneficial effect of high-fidelity conditions

We observed that participants in stereo conditions were able to complete given tasks faster and perceived them as less difficult than participants in mono conditions. Overall, participants took on average 20 seconds less per task in stereo conditions. While we only found statistical significance

Table 3.6: Differences of performance between conditions on each platform. Cross glyphs denote the four experimental conditions; columns describing the FOR and rows the stereo conditions. The top left quadrant marks mono 90 and the bottom right quadrant marks stereo 270. Arrows between quadrants indicate significant difference between groups from worse to better conditions (e.g. in Task01 on CAVE A, participants in the mono 90 condition were significantly slower than those in the stereo 90 and stereo 270 conditions).

90mo	270mo
90st	270st

	Time			Accuracy		
	CAVE A	CAVE B	HMD	CAVE A	CAVE B	HMD
Task01						
Task02						
Task03						
Task04						
Task05						
Task06						
Task07						
Task08						
Task09						
Task10						
Task11						
Task12						
Task13						
Task14						
Task15						

of this effect in the two CAVE environments, we still observed a matching trend in the HMD data (Fig. 3.9). This supports our hypothesis H1: "Participants will perform better in higher-fidelity conditions than in lower-fidelity conditions".

We did not find significant effects of FOR on time or accuracy. While unexpected, this outcome is supported by previous findings in the literature [Schuchardt and Bowman, 2007, Laha et al., 2012, Ragan et al., 2013]. In particular the study by Laha et al. [2014] suggested that FOR affects user performance only in specific tasks, often in interaction with head-tracking. However, our findings underline the importance of stereo visualizations in biological isosurface analysis.

3.4.2 H2 - Similar performances across VR platforms.

Comparing condition groups between platforms, we found evidence against our hypothesis H2: "Participants in conditions with the same FOR and stereo assignments will achieve similar performances across platforms". Our GLMM analysis showed a significant effect of platform on accuracy (Section 4.1). As visualized in Figure 3.6 it is evident that the accuracy for CAVE A was worse than that of CAVE B and the HMD. The mean accuracy values were 67%, 75%, and 75% for CAVE A, CAVE B, and the HMD respectively.

The differences between participant performance in the two CAVE systems were unexpected. We found a significant effect of platform on participant accuracy with users of CAVE B showing better performance in this metric. While some variability between these two platforms existed (e.g. system size, projector brightness and contrast, controller latency, etc.), we could not tie these differences to specific hardware characteristics. Since no consistent offset between condition groups on these platforms could be observed, we hypothesize that these variations are caused by differences in participant populations and individual exploration strategies of participants (see section 3.4.4).

When comparing matching FOR/stereo condition groups across platforms we only found a significant difference between the mono 90°-FOR groups in average completion time. Here the HMD group was the outlier due to the generally wider spread of average completion times on the platform. Limited field of view, higher wearable equipment weight and platform specific artifacts were hardware-features with potential influence on this outcome. The fact that users cannot see their own bodies or other visual reference points within the walking area in the HMD may have influenced their interaction behavior. During the experiments we noticed that users wearing the HMD avoided moving around and preferred to use the wand to rotate the dataset, which may have affected the results. This was confirmed by the responses in the post experiment questionnaire (Fig. 3.7). We also observed a difference in interaction preferences between the two CAVE environments. Participants in CAVE B reported more wand interaction and less walking preference, which we explain by the smaller walking area within the VR setup.

Several participants reported an onset of mild motion sickness during or after the experiment. In the HMD group, three participants accepted the offer to take a break after the seventh task. Many mentioned that the HMD was getting heavy by the end of the experiment.

It is important to notice that, due to time limitations on the study examiners, we could only evaluate a total of 60 participants, which resulted in five participants per group in our between-subject experimental design. Due to the confounding factors found in our study design, we chose not to extend the pool of users beyond the initial 60 participants. Different designs with more participants per condition are required to uncover more details about task-level effects.

3.4.3 Platform differences in task-level analysis

Our analysis of participant performance on the level of individual tasks showed numerous differences between specific condition groups (Table 3.6). We mainly observed effects on task completion times with lower completion times in higher-fidelity conditions (except for CAVE B - Task03 and HMD - Task05). We also found a smaller number of significant effects on the accuracy metric. However, our task-level results were not consistent across the evaluated platforms and did not directly match the results of the experiment by Laha et al. [2014].

In addition to the effects within a platform, there were also several task-dependent effects across platforms. For example, in the mono 90 condition the HMD was faster than both CAVEs on task 13 but slower than them both on task 3. There are several tasks (6, 10, 14) for which HMD users were more accurate than those in CAVE B in various conditions, but also tasks (11, 13) for which the opposite was true. We believe that these differences are the result of confounding variables influencing the outcomes of our study, as discussed in detail in the following section.

3.4.4 Influence of user interaction strategies

We hypothesize that the performance differences between the platforms under similar levels of FOR and stereoscopy have been caused by confounding variables we did not control for in the experiment. While analyzing the qualitative feedback about user experience during the VR experiment we found a correlation between the self-reported movement and interaction grades ("I frequently walked around the dataset to look from different viewpoints", "I frequently changed the viewpoint by rotating or grabbing the dataset", recorded on a 7-point Likert scale) and task accuracies of participants. Additionally, our qualitative observations during trials indicate that participants in stereo conditions were moving more actively within the VR system than those in mono settings. This suggests that participant movement (i.e. exploration and interaction strategy) might be an intermediate variable in the causal pathway between fidelity condition and user performance. Bowman and McMahan [2007] describe such an effect as an immersion benefit that directly influences the application effectiveness.

To evaluate the hypothesis of interaction strategy as an intermediate immersion benefit affecting task participant performance we analyzed the effects of wand interaction and walking on time and accuracy. Running POLR analysis of self-reported walking and wand interaction revealed a significant effect of walking on accuracy ($Pr(> Z)0.0473$) across all platforms (Table 3.5).

In combination with the observed effect of stereo on walking (Section 4.1), this indicates a potential mediating effect of walking on user performance and has potential implications on future experiments comparing user performance in room-scale display environments. Experiments will have to consider participant-specific interaction strategies and control for them through strict study design or recording of interactions and include them as metrics into the result analysis. This might also influence the interpretation of prior experiments that did not control for participant interactions on that level of detail (e.g. [Laha et al., 2014, Ragan et al., 2013, Chen et al., 2012]).

A follow-up experiment to our study by Lages and Bowman [2018] has investigated the effects of walking and wand interaction methods in more detail. Their study shows that walking can enable higher relative performance for users with low spatial ability and users who report little experience with video games. However, experienced video game players may perform better with Non-walking interfaces, especially if they lack spatial ability.

3.4.5 Design recommendations

Our findings suggest several guidelines for research as well as application development in immersive VR visualizations:

- High display fidelity-levels (e.g. stereo) can increase the performance of users in spatially complex isosurface exploration tasks.
- Building an inclusive environment that invites users to explore the data by walking around it may have a positive impact on visualization effectiveness.
- We recommend to integrate user interaction methods and exploration strategies as a controlled variable into the structure of upcoming display fidelity experiments.

3.5 Conclusion

This chapter presented the results of a study on the effects of stereoscopy and field of regard in the performance of biological dataset analysis. We measured completion the time and accuracy of participants in tasks using two CAVEs and an HMD. We found that stereo improved completion time for a given level of FOR; we also found an interaction between stereoscopy and choice of platform in the 270° FOR cases. Our findings indicate that there exist platform-specific effects on user performance that can not be explained by stereo and FOR settings alone. Furthermore we hypothesize that the walking behavior of participants acts as a mediating factor between stereo conditions and task accuracy of participants. This suggests that the use of environments that invite users to explore data by physically moving around it might improve the effectiveness of room-scale VR applications. Overall our results reinforce the need for deeper analysis into user interaction in VR spaces in order to build more effective data analysis applications.

Chapter 4

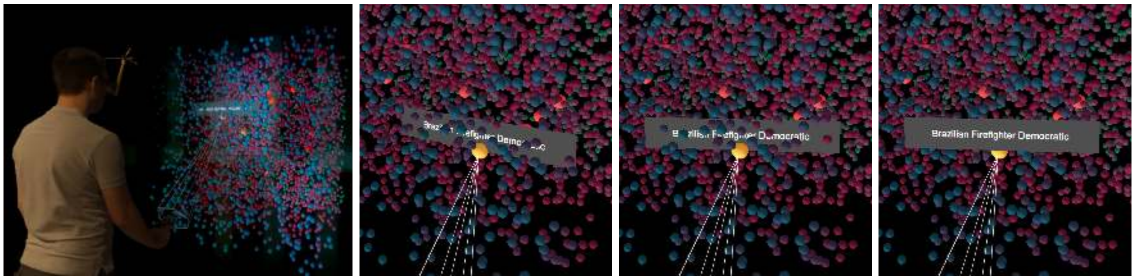
Text Perception in Immersive Particle Visualizations

In this chapter we present our empirical study evaluating specific representation characteristics of text panels embedded within VR visualizations and their influence on user reading performance. This continues our research into perception aspects within immersive environments started in chapter 3.

It is not always possible to completely encode the information of a dataset into intuitive geometric objects or object properties (e.g., color). Specific data values and labels are usually displayed in textual form within a given scene. In visualizations on 2D displays, this is usually done in the form of label or tooltip overlays. The use of labels typically involves a trade-off between legibility of the text itself and the occlusion it introduces into its host visualization. In addition, the label placements need to clearly convey connections to the objects they correspond to while maintaining frame-coherence to avoid visual discontinuities [Ali et al., 2005]. Balancing these requirements has led to a number of different interactive labeling techniques often tied to specific visualization types and use cases [Oeltze-Jafra and Preim, 2014].

In 2D displays, tooltips are a common way to incorporate text information into 3D data visualizations. Through simple click or hover interactions they allow users to obtain detail information about specific regions of the visualization. Tooltip text panels typically overlay rendered 3D visualizations on the near plane of the view frustum to avoid occlusion of the text. Like most text, they are aligned with the pixel matrix of the desktop screen, which allows sharp font rendering. This anchoring to the physical display gives them a stable position with regards to the visualization and increases the readability of the shown information.

Due to their proven effectiveness, it is a reasonable approach to replicate tooltips within im-



(a) (b) (c) (d)

Figure 4.1: This figure shows a study participant within our experimental setup to evaluate text perception (a). Examples of three experimental text representation conditions used within this study: (b) a static text panel embedded into a 3D particle dataset, (c) a text panel dynamically rotating towards the user, (d) and a text panel with removed occlusions, similar to tooltips in classic desktop applications.

mersive visualizations. However, the specific hardware characteristics of VR displays, in particular HMDs, require several specific design decisions and trade-offs, which warrant detailed evaluation [Sadana et al., 2016]. HMDs for example, the currently most wide-spread VR display architecture, lack the spatially stable display surface for text panel orientation and placement that is usually provided by desktop displays. While rendering text panels on the image plane of an HMDs physical display is possible, their fixed size and orientation could be perceived as inconsistent with the motion of other objects in a VR scene during every head motion, leading to reduced immersion.

In practice, the stereoscopic view of HMDs requires text panels to be rendered as 3D objects within a given scene.This opens up a wide space of design considerations for effective placement, orientation and rendering of such embedded text panels. In this study we evaluate a set of specific design choices, to help researchers and practitioners choose the useful representation parameters for their own text panels. By using the immersive particle visualization application described in chapter 6 as host environment for our experiments, we gain results that are also applicable to visually similar immersive scatter plot visualizations. The main goals and contributions of this study are:

- **A comparison of visual acuity and text perception across three state-of-the-art VR devices**
  Here we perform a baseline evaluation of reading speed differences within CAVE and two HMD systems based on their display resolutions.

- **An empirical analysis of label rendering and orientation strategies within immersive visualizations**

By experimentally comparing reading speeds of static or user-facing text boxes embedded within particle visualizations of varying density and the addition of occlusion removal as experimental variable, we provide insights into effective rendering choices.

- **A set of recommendations and guidelines of effective tooltip display methods** The Guidelines informed by quantitative results of our reading experiment and qualitative participant feedback, provide practitioners with actionable advice to improve text presentation within VR data visualizations.

In the remainder of this chapter, we describe our user study analyzing the performance of text panel reading tasks in 3D particle visualizations displayed in three different VR devices. Section 4.1 covers related work on the topics of text representation and legibility in VR, as well as hardware related experiments. Section 4.2 describes the experimental setup and procedures, while Section 4.3 gives an overview of the collected results. Finally, in Section 4.4 we discuss our findings and recommendations for effective VR text representation.

4.1 Related Work

Our work builds on prior 2D and 3D user interface research in the areas of text representation, label orientation, and VR application evaluation.

Reading text is one of the most common tasks computer screens are used for, and transferring this task into VR environments is the main goal of our work. Designing effective ways of representing text and evaluating their effect on readability and user comfort has been a core concern in the field of human factors since the earliest computer screens [Mills and Weldon, 1987]. Historically, display resolution was one of the key factors affecting reading speed and text comprehension, however, in recent years the pixel densities of desktop and hand-held displays have reached a level at which higher resolution does not significantly increase reading effectiveness. A study by Mayr et al. [2017] evaluated the use of displays with angular pixel sizes of 1.68 and 0.86 arc minutes (132 and 264 pixels per inch at ~38 cm distance) in proofreading tasks. While the collected quantitative metrics did not show significant differences between the two displays, users reported subjective discomfort when reading on lower resolution systems. Part of our work focuses on finding parallels to this study by evaluating systems at varying levels of angular resolution. In terms of size and text placement, studies by Dobres et al. [2018] have shown that how font size and placement affect text legibility in single word reading tasks. Larger fonts and wider padding proved beneficial, while random placement increased reading time. Our work extends on this by evaluating a scenario that replicates a real-world application of text display in VR visualizations.

Several labeling methods have already been integrated into VR visualizations and extended to make better use of the visual cues available in immersive environments, effectively forming an *information-rich virtual environment* as defined by Bowman et al. [2003]. Stereoscopic depth-cues for example can reduce the ambiguity of overlapping labels and the wide field of regard offered by modern VR devices provides additional space for potential label placement. Just as in the case of 2D displays, the effectiveness of a specific labeling method depends on the use case and the visualization type [Bell et al., 2001]. However, we believe it is possible to gain generalizable insights into effective label usage by analyzing the readability within a single type of visualization under varying rendering characteristics across multiple VR devices, as shown in previous work studying perception in VR [Laha et al., 2014].

A study by Jankowski et al. [2010] analyzed a variety of ways to embed text panels into 2D videos and 3D scenes in desktop environments, with a focus on font rendering styles and color. They found that embedding text panels with dark background into scenes provided users with the best reading experience. A similar result was found by Debernardis et al. [2014] in an evaluation of text panel color choices in augmented reality (AR) HMD devices. Their study also recommended white text, yet their panel color recommendations were dependent on the real background present in optical and video AR. These studies informed our choice of text panel representation, and we extend their work by evaluating text representation in purely virtual immersive settings. Further experiments in AR settings often focus on text legibility over real-world backgrounds. Work by Manghisi et al. [2017] for example suggests that the legibility of overlayed text can be predicted by analyzing background image information. While our study is mainly focused on occlusions between viewer and text panel, these insights informed out study design to reduce confounding effects from background objects.

The resolution of virtual reality systems (in particular HMDs) matches and sometimes exceeds that of desktop displays. However, it is usually spread over a larger fields of view, leading to a lower angular resolution. This reintroduces some of the design challenges of earlier 2D displays. A study by Dittrich et al. [2013], investigating the legibility of text in physical form, 2D and stereo projection, found that stereo environments with low angular resolution (6-13 arc-minutes per pixel) required larger font sizes than similar 2D display conditions. A similar experiment in high-resolution 2D and stereo CAVE displays (0.03 arc-minutes per pixel) by Iyer et al. [2017] did not show significant differences between stereo and mono text representation. This indicates that the resolution of immersive VR displays also has diminishing effects on reading performance, a topic that we aim to investigate by evaluating multiple VR devices.

Hardware resolution partially limits the design and placement choices of text panels within immersive applications. A study by Grout et al. [2015] evaluated reading performance in two HMD

VR environments with varying resolution. Their results show that text displayed on a flat virtual panel in peripheral regions of an HMD's screen suffers from distortions that impair reading performance. They suggest the use of curved text panels to display screen-filling amounts of text. However, small text panels displayed at the center of the field of view did not benefit from a curved representation. Our work extends upon this by evaluating the impact of user-facing flat text panels.

Dingler et al. [2018] discuss a qualitative method to determine guidelines for text panel size and placement parameters. In their study, participants adjusted text panels into comfortable reading positions for size, distance, and content. While the resulting placement ranges provide useful boundaries for comfortable reading positions, the variance is relatively high and since only one HMD was evaluated, the results might be platform dependent. Based on these previous studies on reading on high-resolution 2D displays and lower-fidelity VR displays, we aim to evaluate the combined effects of these factors in our high-resolution YURT display room [Kenyon et al., 2014] to collect quantitative measurements about readability and panel placement parameters.

Finally, our experimental design employs standardized evaluation methods from the fields of ophthalmology and human factors. We measure visual acuity using LogMAR charts [Elliott, 2016], sentence reading speeds based on Radner test sentences [Radner and Diendorfer, 2014], and perceived mental and physical workload using the NASA TLX Questionnaire [Hart, 2006].

4.2 User Study

The aim of this study was to evaluate the reading speeds of text panels in an immersive scientific data analysis under varying text display conditions. We experimentally evaluated orientation and rendering parameters of text panels in VR particle visualizations. The tasks were designed to uncover strengths and weaknesses of different panel representations for the effective integration of tooltips into immersive visualizations (Figure 4.1). In particular, we evaluated three factors:

- **Static vs. user-facing text panel orientation**
  Orienting a 3D text panel in a VR environment towards the user's head position has potential benefits for readability. Displayed text will match up more closely to the pixel matrix of VR device leading to improved font rendering. However, objects rotating on their own without direct user control potentially interfere with the sense of presence a user experiences within a VR scene. While this might lead to some discomfort for users, we expect that the sharper text representation in user-facing text panels will outweigh this downside, leading us to our first hypothesis:
  H1. In VR environments, user-facing text panels will allow for a higher reading speed

than panels with static orientation.

- **Occluded vs. unoccluded rendering**
  Placing 3D textboxes in complex scenes often introduces occlusion problems, with 3D objects covering up parts of the text and reducing its readability. Displaying the text panels on top of the visualization, as typically done in 2D desktop environments, by removing occluders between the user's head position and the text panels improves the text visibility. However, since panels have a 3D location within the scene, this might cause visual artifacts and a loss of immersion, as parts of the visualized data may disappear in front of the text when a user changes their viewing position. Despite this downside, we state our second hypothesis:
  H2. In VR environments, rendering text panels without occlusions will allow for a higher reading speed than rendering text panels fully embedded in the scene.

- **Display hardware**
  One of the defining features of VR systems, and computer displays in general, is the resolution of their respective displays. Higher resolution allows the rendering of finer details within a given 3D scene without losing visual clarity. In desktop environments, this results in a sharper representation of fonts and is often linked to improved reading comfort. In current generations of commodity HMDs, resolution has greatly improved with every new iteration, nearly doubling the angular display resolution in only six years (e.g. Occulus DK 1, 2012 - Occulus Rift S, 2017). Immersive CAVE display rooms can achieve even higher resolutions, by placing displays farther from the user's head to increase the angular resolution. Some CAVEs can reach and exceed the visual acuity of the human eye, which can recognize details down to one arc minute viewing angles [Williams et al., 2004]. We expect that resolution improvements have a significant impact on the readability of VR text panels and evaluate this in three different VR devices with increasing visual fidelity. Table 4.1 lists resolution parameters of the systems used in our study. This leads to our third hypothesis:
  H3. Displaying text panels in higher-resolution VR displays will allow for a higher reading speed than in lower-fidelity displays.

4.2.1 Virtual Reality Apparatus

In this experiment we used three state-of-the-art VR devices as display apparatus, two consumer-grade HMD systems, and a high-fidelity CAVE display. The selected HTC Vive and HTC Vive Pro systems represent two current HMD systems that see common use in private, academic, and industry settings. Each system offers a 110° field of view (FoV) at 90Hz refresh rate. Their main

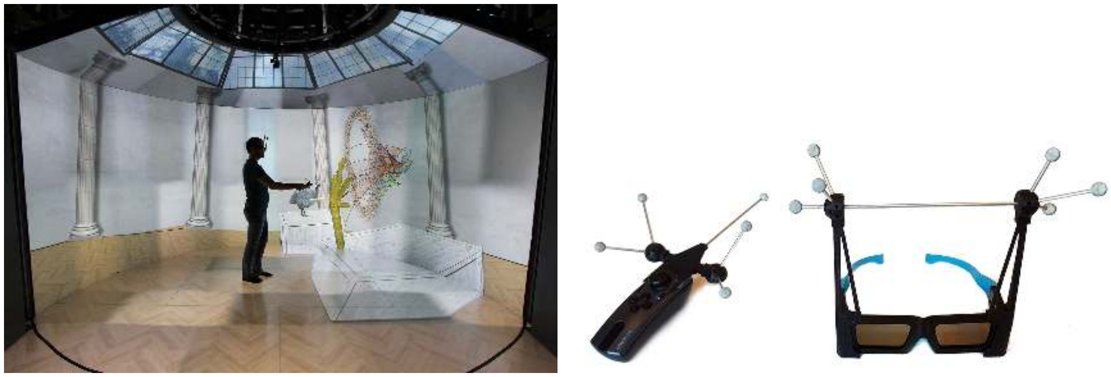
Table 4.1: Characteristics of the three VR displays used in our experiment, from lowest to highest resolution display.

	HTC Vive	HTC Vive Pro	YURT
Architecture	HMD	HMD	CAVE
Horiz. Pixel Size (arcmin)	4.8	2.8	~1.0
Vert. Pixel Size (arcmin)	3.7	2.8	~1.0
Diagonal Field of View (°)	110	110	170
Refresh Rate (Hz)	90	90	60
Accommodation Dist. (m)	∞	∞	~1.2
Stereo Technology	Split Screen	Split Screen	Shutter
Headpiece Mass (g)	470	555	79

difference is display resolution, with the HTC Vive at 1080 $\times$ 1200 per eye, and the HTC Vive Pro at 1400 $\times$ 1600 per eye. Both HMD systems offer physical lens adjustments to accommodate for individual interpupillary distances (IPD) of users. Both HMD's used the same set of positional trackers to avoid differences in tracking latency and calibration (Vive Lighthouse 1.0). Additional characteristics of the HMD systems can be found in Table 4.1. The base system for both HMD headsets was an MSI GE63VR with a quad-core Intel i7-7700HQ CPU clocked at 2.80GHz, 16GB DDR4 RAM and an NVidia GForce GTX 1070. The operating system was Windows 10 Home with all updates at the time of testing.

For our high-fidelity CAVE condition we used the YURT (YURT Ultimate Reality Theater) VR display room located at Brown University [Kenyon et al., 2014] (Figure 4.2). The YURT is equipped with 69 high-definition stereo projectors that use rear projection to illuminate a curved wall with approximately 5m diameter, curved doors, a conical ceiling, and a 12.5 m^2 floor. When standing in the center it effectively provides retina resolution on its 190° front wall. In that position, the YURT covers 95% of the users' field of regard. Additional characteristics are listed in Table 4.1. Stereo was provided by Volfoni active stereo glasses with a shutter frequency of 120Hz. Users interacted with the YURT environment using an Aimon PS wireless wand controller. Glasses and wands are tracked by an OptiTrack Prime 13W optical tracking system with an array of 8 infrared cameras mounted in the ceiling of the YURT.

To provide a realistic experimentation environment, we adapted an existing VR particle visualization application used to visualize fluid dynamic simulations of substrate deformations [Novotny et al., 2019] to view text panels as particle tooltips. Core parameters like particle size and general scale of the visualization were derived from typical settings of the application and have been tested extensively in informal pilot experiments. These initial studies were also used to tune the visual



(a) (b)

Figure 4.2: (a) A student in our YURT display room, working on a VR sketch using a 3D drawing application. (b) The wand and active stereo glasses used in the YURT. The constellations of reflective balls attached to each tool allows the optical tracking system to determine their 3D location during runtime.

representations and interaction methods to be consistent across all three different VR devices. Adjustments included fixing the scale of VR objects, adjusting brightness and contrast to match the common capabilities of all three devices, and matching the interaction layouts on different controllers to the same buttons.

The VR visualization application we adapted for our experimental tasks was developed in C++ based on the MinVR 2.0 framework [Keefe, 2018]. This allowed us to run the same application code on all evaluated VR devices. In each device the application was consistently performing at the maximal framerates of the hardware displays: 90 frames per second (fps) in the HMD conditions and 60 fps in the YURT condition.

4.2.2 Stimuli

As part of this experiment we used three sets of visual stimuli. The first set aimed to collect baseline information about each participant's individual visual acuity. We used standard LogMAR charts visualized at a distance of 4 meters from the participants in each VR environment, with a physical LogMar chart at the same distance as control condition. The employed LogMAR charts feature multiple lines of standardized optotypes (i.e., test characters) at predefined angles of resolution. Character order on these charts was permutated within each line between the four evaluation conditions to avoid memorization effects.

The second set of stimuli evaluate was a set of text panels showing single sentences from the

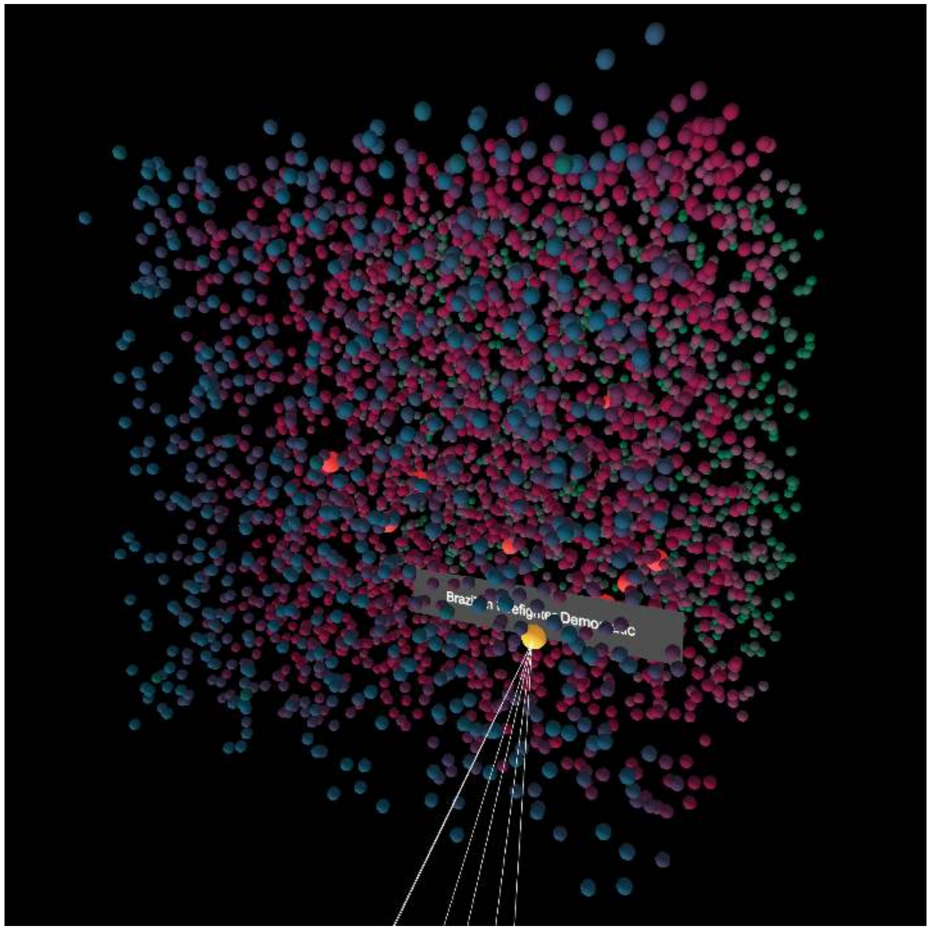


Figure 4.3: A text panel reading task from participant perspective in the HTC Vive environment. The example shows a high-density dataset condition, with occluded static text panels.

collection of English sentence optotypes by Radner and Diendorfer [2014]. These 24 well-studied sentences were designed to have matching reading difficulty and speed for effective comparison. Collecting reading speeds using these standardized sentences provides information on the baseline reading capabilities of individual the participants.

Finally, to evaluate our three main hypotheses (See Sec. 4.2) we created a repeatable point selection and reading scenario within our visualization application. We chose to simulate a reading task within 3D point cloud visualizations with relatively sparse spherical occluders. This is a typical scenario in the exploration of 3D scatter plots and fluid dynamics visualizations [Novotny et al., 2019]. Our selected visualizations consisted of synthetic 3D particle data, with each dataset filling a volume of one cubic meter and particle diameters of 4cm. Particles were randomly placed within these volumes at a density of 1000 (low density) and 4000 particles per cubic meter (high density).

The size of one cubic meter was selected based on prior interaction experiments [Dingler et al., 2018] and observations from our pilot studies. The extent of the dataset allows users to study and interact with the data in a standing position without additional walking motion.

Within each dataset, we selected a set of ten particles as anchor points for tooltips. To ensure comparable occlusion properties of these anchor-points within each synthetic dataset, all points were located within the central 66% of the volume along X, Y and Z axes. Additionally, the points were staggered into ten distance intervals relative to the user's head position, to represent reading at various depths encountered in exploration tasks.

Anchor points were highlighted as distinct red particles within the dataset with a slightly increased diameter (5cm). Bringing the virtual tip of the wand tool close to the 3D location of an anchor point (within a 5cm distance) revealed the attached tooltip text panel. After a tooltip had been visited, the corresponding anchor point changed to orange color to indicate completion of the reading task. The central placement of anchor points and their extended activation radius ensured that participants were able to change their targets in minimal time. The inclusion of the relatively minor task of manually selecting data points ensured that participants were physically engaged with the presented visualization and force intuitive upper-body movement, without deterring from the overall reading task.

Text panels were displayed with either static (orthogonal to the X-Z plane of the dataset) or user-facing facing orientation, to evaluate our hypothesis **H1**. Orthogonal to this condition, we rendered tooltips either embedded within the particle cloud or with occluding particles in front of the panel removed, to test hypothesis **H2**. Within a given dataset all tooltips were displayed with the same orientation and rendering condition. We evaluated text panel orientation and occlusion conditions in both low and high density datasets, resulting in a total of eight testing conditions. To increase robustness against outliers, each reading task condition was repeated once for each participant. Each participant completed a total of 48 text panel reading trials. An example rendering from user perspective can be found in Figure 4.3. Table 4.2 shows the combination of conditions tested within a full trial series on a VR device.

The text displayed on each tooltip was a combination of three words of similar length, syllable-count, and vowel-count. Only words between seven and nine characters length, with exactly four syllables and three to five vowels were selected from an English dictionary [YouGoWords, 2018] (e.g. "Naturally Accumulate Numerator").

During pilot runs of our experimental design we found that users used several different strategies to avoid occlusions in front of text panels. These included walking around the visualization to find the best possible reading perspective or moving their viewing position inside the dataset to put some of the occluding elements out of view. While these differing strategies were interesting observations,

Table 4.2: Trial IDs and their associated condition settings. The trial order was randomized within each VR device.

Trial ID	Density	Orientation	Overlay
1,2	low	static	unoccluded
3,4	low	static	occluded
5,6	low	user-facing	unoccluded
7,8	low	user-facing	occluded
9,10	high	static	unoccluded
11,12	high	static	occluded
13,14	high	user-facing	unoccluded
15,16	high	user-facing	occluded

they caused greatly varying reading speeds across initial participants. To avoid this confounding factor we enforced a set of movement restrictions that reduced the number of possible interaction strategies as listed in detail in the procedures section (Sec. 4.2.3).

4.2.3 Procedure

The entire study was conducted at Brown University's VR facilities, which house setups for all three VR devices in the same building. Each study participant completed the entire experiment within one session, performing tasks in all VR environments as part of the within-subject design. Upon arrival at the facilities we collected demographic information with a pre-experiment questionnaire. This survey included questions about individual experience with VR systems and scientific visualization in general. We measured each participant's interpupillary distance and eye height in standing position to customize VR visualizations for each individual. As final step before starting the three VR device trial series, we measured visual acuity with a physical LogMAR chart to confirm the 20/20 vision requirement set for each participant.

While each participant completed tasks in all three VR devices, we permutated the system order using a standard latin square design between participants. Within each system the procedure was as follows:

Visual Acuity

Participants performed standard LogMAR acuity tests at a virtual chart distance of 4 meters (13 feet). This matched the examination procedure in the preliminary physical acuity test. The visual acuity score was determined based on the number of correctly perceived optotypes at different angular resolutions.

Reading Speed

We measured the baseline reading speed of participants in each of the three VR environment using 3D text panels with standardized Radner sentence optotypes. These test sentences for measuring reading acuity and speed were designed to be "as comparable as possible in terms of number of words (14 words), word length, position of words, lexical difficulty, and syntactical complexity" [Radner and Diendorfer, 2014]. To minimize distortions, the panels showing Radner sentences were oriented to always directly face participants at eye level (i.e., billboarding). In each system, we tested five angular text resolution conditions placed at distances of 0.6 and 1.2 meters (2 and 4 feet) from the head of participants. The angular text sizes were chosen based on common LogMAR scale sizes, as suggested by Radner and Diendorfer [2014]. The tooltip distances represent distances within arms reach and fall into the effective distance range proposed by Dingler et al. [2018]. Due to the limited number of sentence optotypes available, we chose to use angular text size steps of 0.2 LogMAR and to repeat sentences in multiple conditions. To avoid confounding memorization effects, we did not repeat sentences within a given VR device and used each sentence at most twice between conditions.

To accurately measure reading time, users were first presented with blank panels matching the size and shape of the text panel. The corresponding text was shown after a participant pulled the trigger button on the interaction wand. We asked participants to read the shown sentences as quickly and accurately as possible out loud. Reading speed was then measured as the time between revealing the text to the end of the vocalization.

Tooltip panel reading

In each tooltip reading trial, a one cubic-meter volume of synthetic particle data was placed in front of participants. Ten highlighted particles had hidden tooltip text panels attached to them. Participants were asked to navigate their wand to each highlighted particle and read all ten tooltip text panels out loud as quickly and accurately as possible from a defined standing position. Participants were instructed not to step away from their standing position, indicated by a circle on the floor. Other body movement like leaning towards the dataset and crouching was allowed. A study assistant was present behind participants to ensure these movement restrictions were followed.

Before starting the 16 trial series in a VR device, participants completed a training task, which introduced them to the interaction concept of pointing the wand tool at highlighted particle locations to reveal tooltip text panels. They were also informed that they could request a break at any time, to accomodate for cases of simulator sickness. Before and between trials, participants were shown a text panel reminding them of the task instructions. With a pull of the wand trigger participants were able to start a trial, which revealed a particle dataset and the associated highlighted tooltip locations. Upon reading the final tooltip, timing was completed by the study examiner, and participants were returned to the intermediate instructions text panel. The trial order was randomly permutated between VR devicess and participants.

After completing all trials within a system, participants were asked to complete a NASA Task Load Index (TLX) questionnaire to inform us about differences in perceived workloads across environments.

The study concluded with a post-experiment questionnaire in which we asked participants if and how our evaluated conditions affected their effectiveness in text panel reading tasks.

4.2.4 Participants

We recruited 18 volunteers between the ages of 18 and 25 (Mean 20.6 years) from the student body of Brown University, forming a pool of seven female and nine male participants (two participants chose not to disclose their gender). Seven participants reported normal vision, nine used glasses and two used contact lenses to correct their vision. The majority of participants (15 out of 18) were native English speakers. Five participants reported expertise with 3D visualizations and/or video games but only one reported frequent use of VR devices. Participants on average took 65 minutes to complete the experiment and were compensated at a rate of 10 USD per hour.

4.3 Results

4.3.1 Visual Acuity

The LogMAR visual acuity scores collected in each VR system revealed interesting results tying character perception to the visual resolution of VR systems. Scores collected in the physical space control condition showed that participants had 20/20 or better vision (LogMAR score of 0 or below), with the exception of one participant with 0.2 LogMAR acuity. In the YURT, participants achieved a score of 0.18 on average. In the HMD conditions average LogMAR scores were 0.54 in the HTC Vive Pro and 0.6 in the HTC Vive system (Figure 4.4, Figure 4.5). Repeated-measures mixed-model analysis indicated significant differences between the LogMAR results in the 4 conditions. Full-factorial paired t-tests with Bonferroni correction revealed significant differences between the Physical, YURT and combined HMD conditions, but not between the HMD conditons ($t(17) = 3.41$, $p = 0.003$, $\alpha = 0.0017$ between HTC Vive and HTC Vive Pro, $p < 0.001$ in all other pairings).

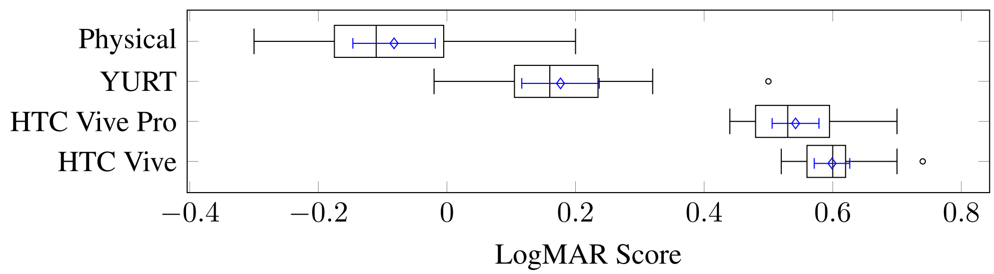


Figure 4.4: Visual acuity measurements of users in the real world and within our three evaluated VR environments. A LogMAR score of zero indicates 20/20 vision. Differences between all conditions, except between the two HMD systems, were statistically significant. Blue error bars within box plots represent the 95% confidence interval of the mean LogMAR score.

4.3.2 Reading Speed

The collected reading times of Radner sentence optotypes revealed similar differences between VR device conditions. The expected reading time under optimal conditions is 5 seconds per sentence. Our collected study results matched these reading times in all three VR devices in conditions where the effective text size was covering at least 19.5 arc minutes (LogMAR 0.8) of the participants visual field. Below that size, we measured significant reductions in reading speed in all environments until participants could no longer complete the reading tasks. The lowest readable text sizes that could reliably completed by system were 23.9 arcmin (LogMar 0.7) in the HTC Vive, 19.7 arcmin (LogMAR 0.6) in the HTC Vive Pro and 9.8 arcmin (LogMAR 0.3) in the YURT environment (Figure 4.6).

4.3.3 Tooltip Panel Reading

We processed the task completion times of the combined three-word reading trials using a full-factorial repeated measures mixed-model analysis with the 8 trial and 3 VR device conditions modelled as within-subject factors. Our analysis on log task completion times indicated several significant main effects and interactions between condition groupings, that were further investigated using post-hoc paired t-tests with Bonferroni correction.

The strongest statistical outcome was a two-way interaction between text panel occlusion and density condition (Figure 4.7). Task completion times differed significantly between occlusion conditions when panels were placed within high density particle datasets. Occluded text reading in high density data took on average 22 seconds longer to complete than in the other three conditions. Post-hoc pared t-tests between all four conditions confirmed statistically significant between the "oc-

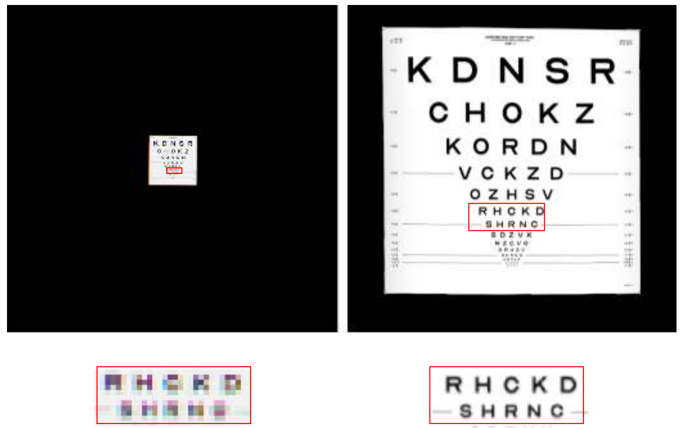


Figure 4.5: Resolution-dependent differences in screenshots of the LogMAR chart representation between the (left) HTC Vive and (right) YURT environments. The red rectangle shows the effective resolutions of LogMAR 0.5 and 0.4 lines of the chart. While the charts appeared at the same size and distance to users within the respective VR environments, the lower angular resolution of the HMD screen reduced the area covered by the chart on the actual display. It is therefore difficult to reliably read text below LogMAR 0.5 on an HTC Vive display.

cluded high density" condition and the other conditions ("occluded low density": $t(107) = 29.04$, $p < 0.0001$; "unoccluded high density": $t(107) = 26.82.73$, $p < 0.0001$; "unoccluded low density": $t(107) = 33.99$, $p < 0.0001$; $\alpha = 0.0017$).

In similar fashion we found a two-way interaction between occlusion conditions and VR platform. Post-hoc paired t-tests showed no significant differences were found across the three platforms within the unoccluded panel condition. However, all occluded conditions were significantly slower than unoccluded conditions, and reading occluded text on the HTC Vive took on average ten seconds longer than in the YURT environment (t(72)=-7.57, $p < 0.001$, $\alpha = 0.0017$, Figure 4.8). Embedded text panels in the HTC Vive Pro did not show significant differences to the corresponding YURT and HTC Vive completion times. Apart from these results, we did not find further main effects or interactions on task completion time.

4.3.4 Participant Reported Results

Finally, we collected self-reported participant responses with NASA TLX forms after each VR condition and post-experiment questionnaires. The seven questions of the TLX assess cognitive and physical workload perceived by participants and their confidence in the outcome. As Figure 4.9 shows, all three VR systems perform similarly in most of the collected categories, with a non-

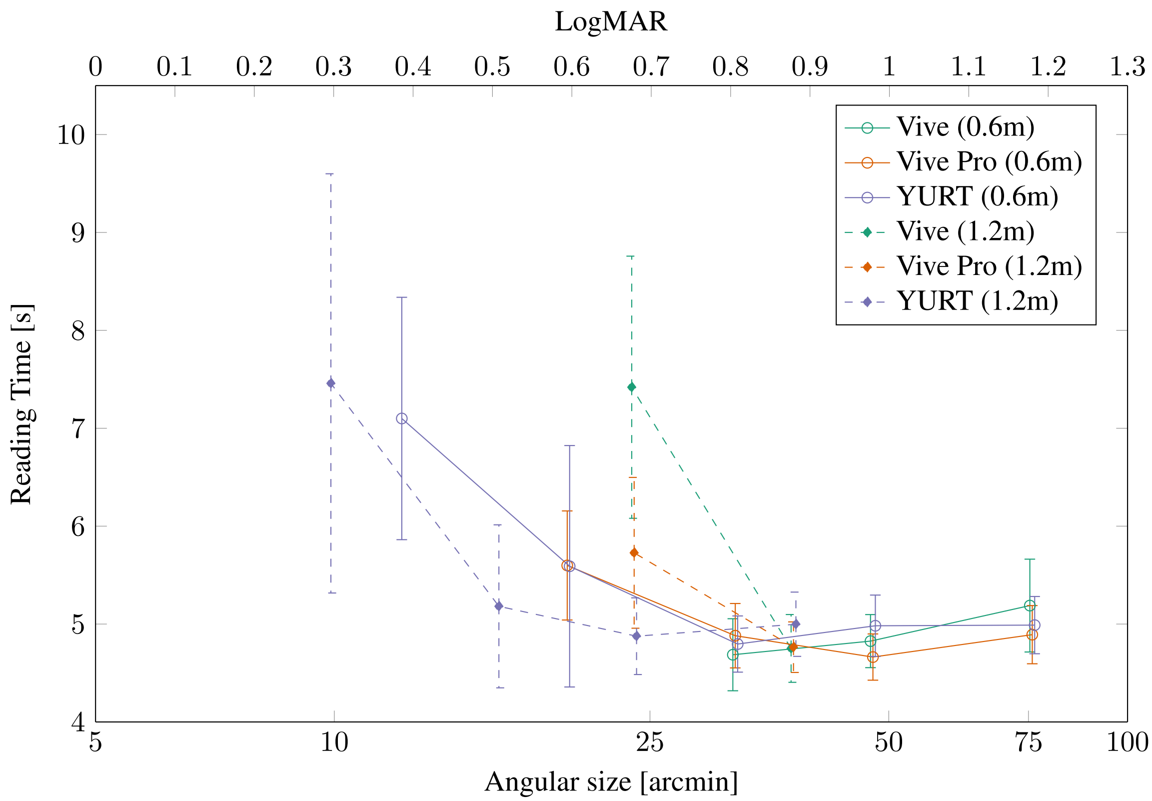


Figure 4.6: Average sentence reading times over text size in different VR Environments and text panel distance conditions. Reading speed decreased from five seconds at large angular sizes (LogMar >0.8) to up to ten seconds at smaller sizes. Users in HMD conditions were not able to complete all reading tasks and reached critical reading speed at larger angular sizes than in the YURT condition, highlighting the effect of display fidelity. Error bars represent the 95% confidence interval of the mean reading time. Vive Pro data points (blue) show the angular text size for each condition. Vive and YURT data points represent the same text size, but are shifted slightly to increase readability.

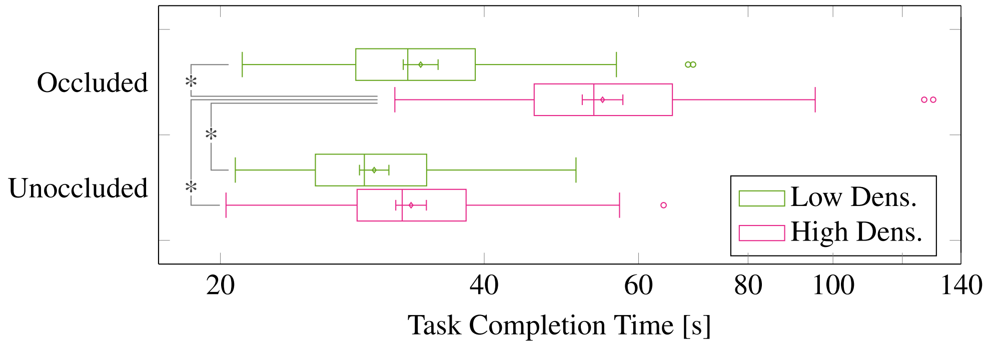


Figure 4.7: A two-way interaction was found between occlusion and density conditions. Participants took significantly longer when reading occluded text panels in high density datasets, while in low density datasets we did not find a similar effect.

significant trend towards higher-fidelity systems. This is exemplified in the mental demand category (Q1), where tasks in the YURT were rated as less demanding.

The post-experiment questionnaire indicated that users preferred the YURT as display environment for the presented reading task, followed by the HTC Vive Pro and the HTC Vive (Figure 4.10). Asked about the text panel orientation preferences, we found that there was no clear preference in low density conditions, while in high density conditions user-facing text panels were preferred (Figure 4.11).

A majority of participants preferred the use of unoccluded tooltips to embedded ones. With 14 out of 18 participants reporting that the unoccluded representation had a strong influence on their task performance. This preference was stronger in high density datasets (Figure 4.12).

4.4 Discussion

We found partial support for our initial hypothesis with the collected results, and were able to gather several key insights about effective text panel scale and placement.

4.4.1 H1. Static vs. User-facing Text Panel Orientation

Our hypothesis that user-facing panels have a significant advantage over static ones was not supported by the quantitative tooltip reading data collected in this experiment. This was likely caused by the study design which limited participant movement during the experiment. Not allowing users to walk into or around the dataset meant that the effect of user-facing panels was not as noticeable,

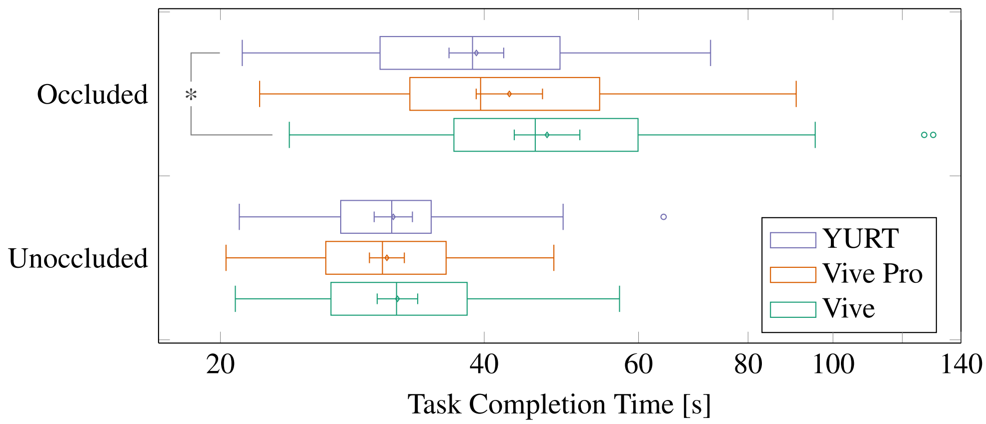


Figure 4.8: A two-way interaction was found between VR environment and occlusion mode shows differences with regard to the occlusion setting. In the occluded setting, we found a statistically significant effect between Vive and YURT conditions, which was not present in the unoccluded condition.

since the participants could not get in a situation in which they had to look a static panels from a steep viewing angle. The maximal deviation from a straight on viewing angle in the static case was 30° for text panels close the the participant. Given this outcome we can neither confirm or reject this hypothesis.

We originally assumed that the better alignment of the tooltip text with the pixel matrix of the used VR display would significantly increase the readability of the displayed text. However, this might not have been a deciding factor for the text representation style used in our study. The tooltips displayed only 38 characters on average, which is relatively short for a text panel. With the chosen font size and panel distance, a majority of the text could be displayed in the center of the screen of a VR headset, which is also the least distorted area of the pixel matrix. This may have reduced the impact of slightly slanted views onto text panels, due to the higher effective resolution. Work by Grout et al. [2015], partially supports this insight, as they found that small curved text panels do not lead to improved reading performance when compared to flat ones if they can be viewed at the center of an HMDs screen.

While not supported by reading speed, we did collect participant responses on the advantage of using user-facing panels. Especially in dense particle volumes, the orientation behaviour allows users to maneuver the text away from occlusions placed right in front of the tooltip, which made reading "easier due to allowing the panels to come in front of/behind objects in the scene." Our collected user preference rating for panel orientation supports this interpretation (Figure 4.12).

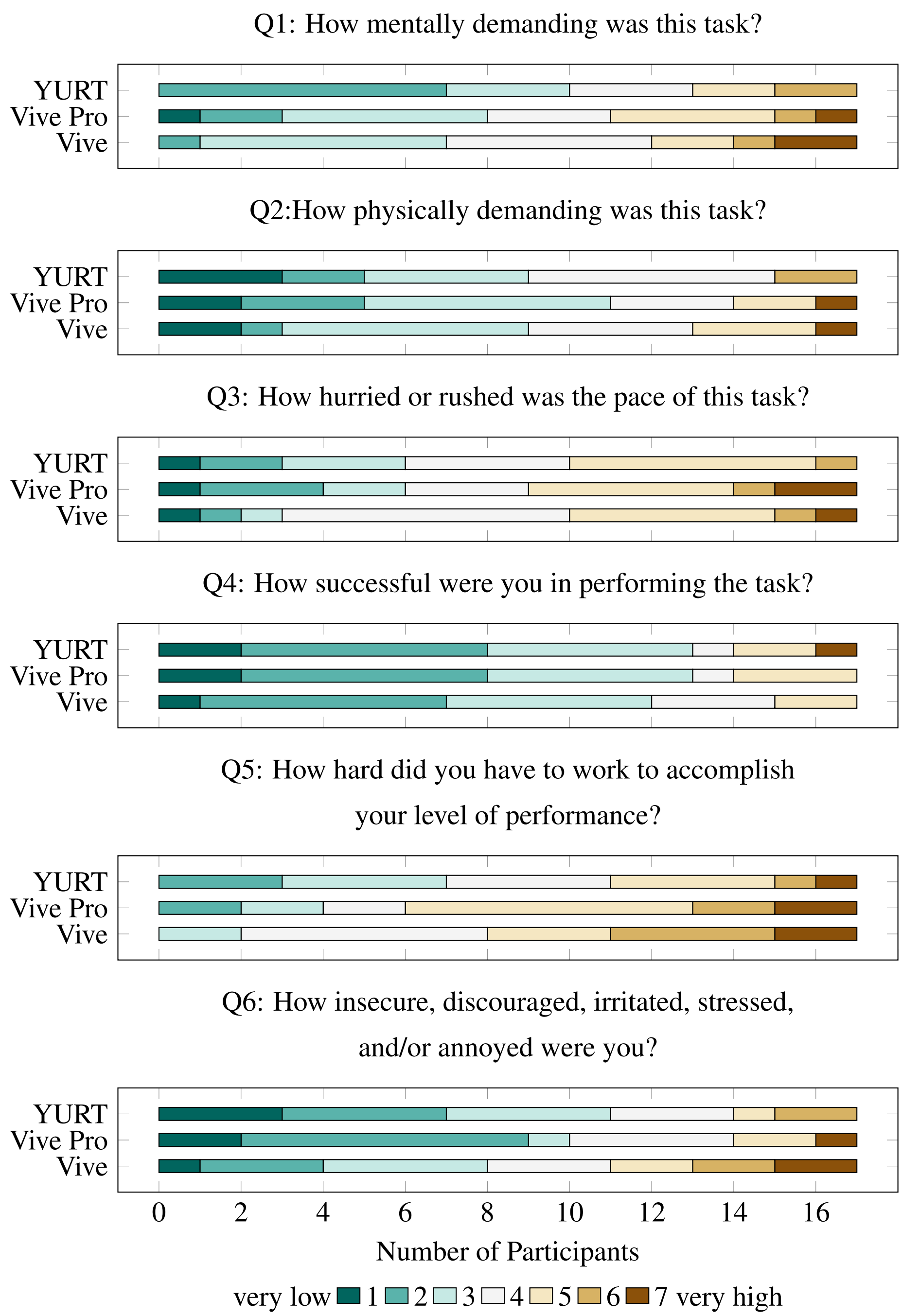


Figure 4.9: An overview of the NASA TLX survey results collected from every participant after completing tasks in each VR environment, binned into a 7-point Likert representation. (Question 4 responses range from "1, perfect" to "7, failure"). The results show a minor preference of participants towards higher fidelity VR environments, however we did not find a trend with statistical significance.

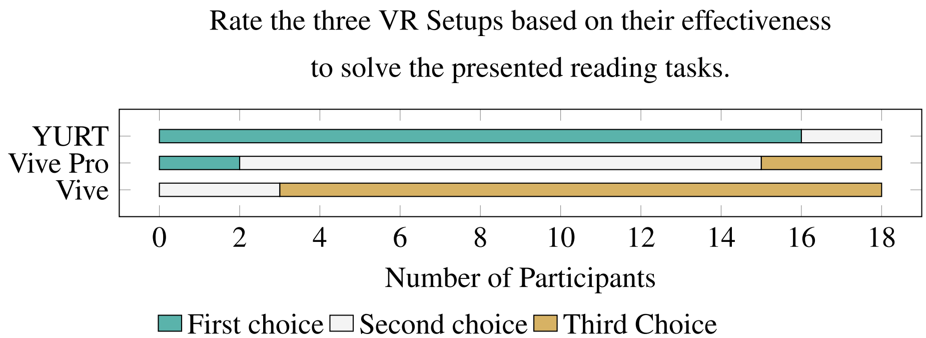


Figure 4.10: Post-experiment questionnaire results indicate a strong preference of the high-fidelity YURT environment over the two HMD systems.

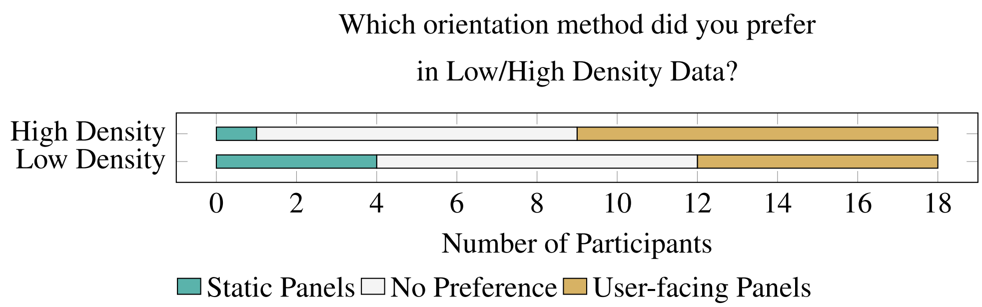


Figure 4.11: Post-experiment questionnaire results showed that a majority of study participants preferred user-facing text panels in high density datasets. No clear preference for a methods was found in low density datasets.

4.4.2 H2. Occluded vs. Unoccluded Rendering

The UI concept of overlaying tooltips over a visualization to increase readability was strongly supported by our collected quantitative and qualitative data. However, differences were only noticeable in situations with high numbers of occluding objects, such as our high density volume condition.

To our surprise, only 2 out of 18 participants reported that the visual artifacts created by removing all objects in front of a text panel caused them any visual discomfort. A possible explanation for this might be that our experimental task was mainly focused around text perception and omitted analysis tasks on the particles themselves. One participant mentioned that the unoccluded tooltips “sometimes made it hard to navigate the space looking for next spheres.”

In our study, text panels were relatively small and only covered a small part of the visualization when shown unoccluded. We could not clearly determine at which size a text panel starts to interfere with its host visualization. This size boundary likely depends on the data analysis task at hand and

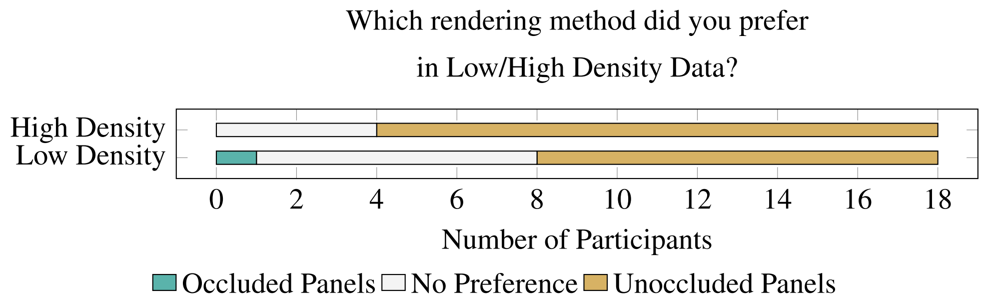


Figure 4.12: Based on the post-experiment questionnaire, we found that study participants preferred reading from unoccluded text panels in both density conditions

requires a more specific experiment to confirm.

4.4.3 H3. Display Hardware

Higher display fidelity had a significant impact on text panel reading performance in difficult reading situations, partially confirming our third hypothesis. This benefit can best be explained by the higher display resolution offered by our YURT environment, compared to the two HMD devices. The advantage of increased resolution was shown clearly in the visual acuity and reading speed part of the experiment. In the YURT environment participants were able to read text at significantly smaller angular sizes (Figure 4.4 and 4.6). This sharper font representation allows text to be more easily readable even if characters are partially occluded, as it happens in the high density condition.

However, despite the claimed retina resolution of the YURT, participants obtained lower visual acuity scores than in the physical control condition. This has likely been caused by a combination of the overall contrast of YURT projectors and distortions of the LogMAR chart by rendering functions that correct the projection for the curved screen surface.

In tooltip reading tasks we only found significant differences between environments when a high amount of occluders where present, and only between the YURT and HTC Vive systems. This was likely the result of the choice of angular text size in that part of the experiment. To provide similar reading speed conditions between all three environments, we chose a text size equivalent to 0.8 LogMar character optotypes informed by our pilot experiments. In situations with few occluding particles, all three systems provided equivalent levels of text readability. Once a high number of occluders are present, reading performance increases with display resolution. Overall, we only found a weak trend towards faster reading speeds in higher fidelity systems (Figure 4.13).

This indicates that the optimal tooltip representation parameters are dependant on the visual fidelity of a VR system. Smaller text size and tooltip panels would, for example, allow for an

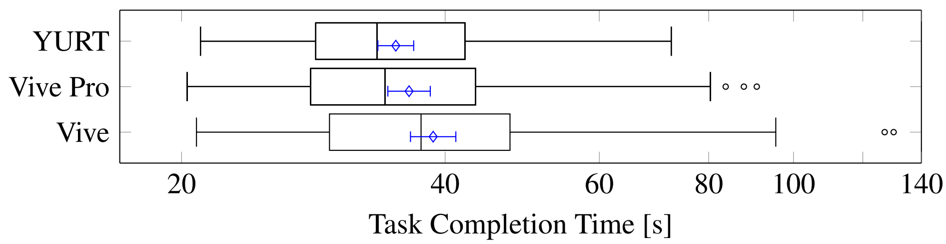


Figure 4.13: Overall comparison of task completion times between VR Environments. Due to the similar reading performance of participants in the low density condition, we only observed a non-significant trend towards faster reading speeds in higher resolution systems.

unoccluded representation that minimizes covering host visualizations.

4.4.4 Limitations

Comparing VR systems of very different architectures potentially introduces confounding variables due to platform specific hardware characteristics. In this study we attempted to minimize perceptual differences between the YURT and HMD environments. To match the color representation of 3D objects in the HMD condition to the lower brightness of the projector-based YURT setup, HMD brightness had to be reduced to 60%. However, we chose to leave contrast and black-level at the best setting for each environment as we see it as a defining characteristic of the display system. Similarly, we did not reduce the higher frame of HMD systems (90 fps) compared to the YURT (60 fps). To minimize user interaction differences, only the index finger trigger buttons of the respective wand tools were used in the experiment (Figure 4.2). Despite the higher black level and lower frame rate of the projector-based YURT, it still stood out as the preferred platform both quantitatively and qualitatively.

A majority of our study participants (11 out of 18) used glasses or contact lenses to correct their vision. While none of them reported any discomfort using VR devices during the experiment, the use of vision correction might distort the representation of the virtual environment, especially in conjunction with the optical lenses of HMDs. We did not find a significant effect of vision correction on our study outcomes, but selecting for participants with uncorrected normal vision might eliminate a confounding factor in future experiments.

4.4.5 Guidelines and Open Questions

Based on the results gathered in our study, we can make the following recommendations on effective text panel display in immersive VR:

- **Angular text size:**
  We found that an angular size of approximately 30 arcmin (LogMAR 0.8) was the lowest size that could be read without loss in reading speed across all evaluated platforms. We recommend not going below this text size in applications targeted at current HMD hardware (e.g, HTC Vive 2019). It is important to note in high-fidelity displays like the YURT even a 50% reduction of this text size did not lead to slowed reading speeds. In situations with large amounts of text or high levels of occlusion, a heigfher resolution display should be preferred.

- **Occlusion removal:**
  Our experiment showed that in environments with large numbers of occluders, removing objects in front of text panels significantly increases readability. While we can not give a direct recommendation at which occluder density removal techniques should be considered. We suggest testing readability of text panels in the most difficult occlusion conditions expected in a given VR application, and to consider occlusion removal if reading comfort is an issue.

- **Text panel orientation:**
  While we did not find a quantitative effect of user-facing text panels on reading performance, our collected qualitative responses indicated user preference of dynamic orientation in occluder-rich scenes. We recommend to at least consider dynamic panel orientation in applications which require significant user movement within a scene or environment.

- **VR environment:**
  Finally, we want to emphasize the importance of testing VR application usability across a variety of different environments. Hardware fidelity components, in particular display resolution, have a measurable impact on the usability of a visualization tool. We recommend to defining default text representation parameters based on the lowest-fidelity target VR display, and providing ways to tune these parameters in higher-fidelity environments.
  In general, if a high amount of text is expected within an application, we recommend developing for VR displays with a high-resolution to provide users with adequate reading comfort in compact text panels.

While this study lays some groundwork for effective text panel representation, further studies are required to resolve more detailed research questions. Especially the relationship between VR display resolution and angular text size and their combined effect on reading speed, could be explored in finer increments to create generalized hardware specific font size suggestions. In the case of our two evaluated HMDs we observed that the higher resolution of the Vive Pro generally provided better text readability, but not enough to reach statistically significant difference in LogMar

Scores. A high-resolution VR system could be used to simulate resolution conditions currently not covered by existing devices to generate reading speed curves that are generalizable between devices.

Likewise, future experiments could benefit from a finer gradation of evaluated angular text sizes. Our findings show that the drop in reading speed from its maximum happens at different LogMAR text sizes based on the VR System (see Figure 4.6). The reading speed falloff at this “critical print size” CPS) [Subramanian and Pardhan, 2009] occurs over a text scale difference of 0.3 LogMAR, which matches reading speed evaluations in physical settings. However, research by Radner [2017] has show that in real world paragraph reading scenarios the expected CPS would close to 0.1 LogMAR for participants of our age range, while even in the high-resolution YURT condition the observed CPS was close to 0.5 LogMAR. This difference does not match up with the initial visual acuity evaluation within our study (Figure 4.4) and warrants further investigation.

The negative impact of density and size of occluding objects in front of VR text panels warrants more investigation. Finding the density thresholds at which text reading speed starts to diminish could lead to better recommendations on when to consider occlusion removal as an option. Additionally, rather than simply removing occluding objects, other methods of occlusion reduction, such as transparency or size changes could be evaluated to avoid a loss of dataset context while still enhancing readability.

Another open question is, whether the use of readable words was a confounding factor when evaluating occlusion conditions. Due to the length of words participants have very likely guessed some of the occluded characters based on the context of the entire word. This situation might be a common occurrence that can be exploited in visualizations that use text panels to name specific regions of a visualization. However, when numeric outputs are shown it will often be necessary to remove occlusions, to ensure that every digit can be read. A study evaluating number reading would likely uncover different practical limits for text size and occlusion coverage.

Finally it would also be of interest to investigate the use of text panels with text sizes smaller than the proposed limit for a given resolution during general use. This would force specific user interactions, like moving closer to the panel, in order to read the content. Finding a balance between smaller, less intrusive text panels and the amount of effort required to read the panel could lead to a more efficient use of the available virtual space.

4.5 Conclusion

Our results show that several display concepts of tooltips can be transferred from 2D desktop applications into immersive VR visualizations. In particular removing occlusions in front of 3D text panels was very effective in increasing the text readability. In difficult reading conditions with high numbers of occluding objects, displays with higher visual fidelity offer improved reading speed, even without removing occlusions in front of the text panel. We could not confirm significant advantages of orienting tooltips to always face the user within our experimental framework. Finally, we provide angular size recommendations for effective text representation in current VR hardware systems.

Chapter 5

Feature Perception in Immersive MRI Visualizations

While the previous two chapters focused on evaluating general data exploration tasks performed by untrained study participants, it is also important to analyze performance and feedback from expert users analyzing their data in immersive settings. In this chapter we present a study evaluating the potential of VR visualizations to identify surface blood vessels in a magnetic resonance imaging (MRI) scan of a human placenta.

This work is motivated by the need for better methods for analyzing placental vessel structures and planning fetal surgical interventions. Vascular anatomy is typically explored in the diagnosis and treatment of several pregnancy disorders, including the main research area of our collaborators: twin-to-twin transfusion syndrome (TTTS). TTTS is a rare and potentially lethal condition affecting twin fetuses who share a placenta but have separate amniotic sacs; it causes disproportional blood transfer between the two through communicating placental vessels. Both fetuses are at very high risk of dying in utero [Luks et al., 2005]. In clinical practice, TTTS is diagnosed by ultrasound (US) [Lombardo et al., 2011]. While surgical planning from 3D medical imaging is possible, it is not yet possible to map out the harmful placental interconnections in advance [Luks et al., 2001]. Vascular anatomy is usually visualized by injecting radio-opaque contrast agents directly into a patients blood vessels (angiography). This approach is not appropriate in the fetus, as puncturing its blood vessels would be too invasive, and contrast agents would be potentially toxic. Thus, surgeons typically identify problematic vessels connections by examining the placental surface through a fetoscope during surgery. In general, no additional medical scans are obtained, which prevents meaningful planning of the intervention.

The goal of our project was to lay the groundwork for a VR-based medical visualization tool to support the planning of placental surgery. As a proof of concept we created an immersive direct volume rendering application which allowed expert users to trace and mark vessels within placental MRI datasets. Our reason for using a VR display was two-fold. The first was navigation. Depending on its location within the amniotic bubble, the surface of a placenta might form a relatively complex shape. We posit that the intuitive head and body motion to control the view in VR will aid medical professionals in their vessel tracing task. Additionally, we evaluate whether the enhanced depth perception of the immersive visualization allows participants identify placental vasculature accurately and consistently without the use of contrast-agents or prior vessel segmentation in the dataset.

This chapter is based on:
J. Novotny, W. R. Miller, F. I. Luks, D. Merck, S. Collins, and D. H. Laidlaw. "Towards Placental Surface Vasculature Exploration in Virtual Reality." *IEEE Computer Graphics and Applications*, 40(1): 28-39, Jan.-Feb. 2020.

J. Novotny and W. R. Miller contributed to the implementation of the VR application and the study design on equal terms. All authors provided their domain knowledge to the project and contributed to the writing of the publication.

The remainder of the chapter is organized as follows: Section 5.1 details the motivation for our work and its context in prior studies. In section 5.2 we present our experimental setup, procedures and collected metrics followed by a report of our findings in section 5.3.3. Finally, we discuss our outcomes and present our conclusions on the effective use of VR for placental MRI scan exploration in sections 5.3.5and 5.3.6 respectively.

5.1 Related Work

The size and structural features of the placenta can be important indicators for complications during a pregnancy. Analyzing in-vivo scans of the placenta from medical imaging data (usually ultrasound) is common practice in prenatal care [Elsayes et al., 2009]. The now wide availability of 3D scanning modalities offers new opportunities for a more accurate diagnosis. Luks et al. showed the benefits of volume renderings of uterine MRI scans in planning TTTS procedures on a desktop computer [Luks et al., 2001]. Their system helped users understand the spatial relationships between the placenta, umbilical cords, and fetuses. However, visualizing small communicating vessels on the placental surface was not possi-ble. Unlike previous work, our approach, which leverages higher-resolution data and 3D VR navi-gation, lets users inspect vessels that are small enough to be potentially relevant connections in TTTS cases.

Wang et al. have introduced a semi-automatic system to analyze the placenta in MRI scans. Their Slic-Seq system uses machine learning to generate segmentations of the placenta with minimal user interaction [Wang et al., 2015]. In follow-up work they augmented Slic-Seg to work on multiple scans taken from different views [Wang et al., 2016]. A recent approach by Alansary et al. [2016] presents a fully automatic segmentation framework for the placenta from motion-corrupted fetal MRIs. Their proposed framework adopts convolutional neural networks (CNNs) as a strong classifier for image segmentation followed by a conditional random field (CRF) for refinement. In contrast to these related systems, we remove only occluding anatomy (i.e. fetus and parts of the umbilical cord) from the dataset by manual segmentation, to retain the context of the uterus walls in our experimental visualization. The existence of these placenta-segmentation methods suggests that this process can be automated in future experiments.

While VR interaction techniques are generally intuitive, users often require some training to effectively use the full range of capabilities offered by VR environments. Selecting the right interaction methods for a given task is critical in creating a successful application. Several studies have investigated the advantages and disadvantages of VR systems in relation to traditional desktop setups. Pausch et al., comparing a VR interface and a stationary monitor for search tasks [Pausch et al., 1997], found that VR users were no more accurate at finding all targets in the space than stationary monitor users; how-ever, they were significantly faster at determining whether a target existed in the space because they spent much less time reexamining previously searched areas of that space. In the context of medical applications, arriving at a quick diagnosis is an important efficiency consideration and was one of the reasons for our VR experiment.

Olwal and Feiner [2003] discussed the difficulties in pointing to VR objects when they are so close together as to cause pointing ambiguity or visual occlusion and introduced a flexible pointer system that improved pointing results in their experiments. Investigating a similar problem, Keefe et al. [2007] described the difficulties of 3D tracing tasks in VR. Finding that freehand 3D tracing is difficult even when augmented with simulated friction haptic feedback, they developed a controlled tracing method to simplify drawing curves in 3D space. To avoid these problems, our annotation tool was based on separate line segments instead of continuous curves resulting in reduced difficulty and training time.

5.2 Methods

5.2.1 Dataset

The dataset used in our experiment was a uterine steady-state free precession T2-weighted MRI scan of a singleton fetus, shown in Figure 5.1, with voxel size 0.7 0.7 1.2 mm. The image slices had resolution 512 512 pixels (resampled from a 256 256-pixel acquisition matrix), and we used

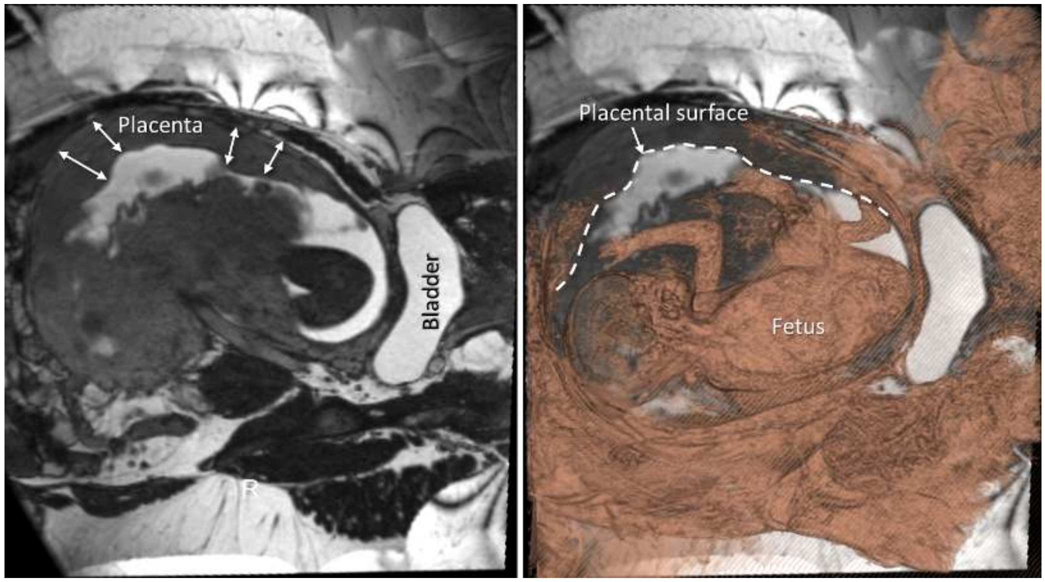


Figure 5.1: 2D renderings of the MRI scan used in our experiment highlighting placenta and bladder on the left as well as placental surface and fetus on the right. The image on the right contains a partial volume rendering overlay indicating the position of the fetus.

45 slices showing the volume containing the placenta. The placenta was aligned with the X-Y plane of the volume with minimal curvature, result figures/plots in this article are therefore project-ed to this plane. We used an anonymized MRI scan of a single pregnancy at 25 weeks gestation as a stand-in for proof of principle. The imaging data of the fetus was manually removed from the MRI scan to reduce occlusion of the placenta surface. Unfortunately, this also included areas where the fetus was in direct contact with the placenta, creating stair-shaped rendering artifacts. This manual removal step might be avoided in the future, by using automated methods as discussed in the Placental Visualization sidebar. A rendering of the original dataset, including the fetus, is shown in Figure 5.2, while the rendering shown to participants, without the fetus, is displayed in Figure 5.3. To ensure that all vessel details were visible, the data were displayed at 6.8 times their original size, a scale suggested by our medical collaborators during pilot runs of our experiment. Relevant vessels were only visible on the user-facing side of the dataset, reducing the need to examine the dataset from all sides.

5.2.2 VR Environment

We carried out our experiment within the YURT (YURT Ultimate Reality Theater), Brown University's advanced CAVE display [Kenyon et al., 2014]. It is equipped with 69 high-definition stereo

Figure 5.2: The volume-rendered MRI scan including the fetus displayed inside the YURT. The VR visualization was displayed as a 2.4 m-tall virtual object to fully utilize the YURTs available space.

projectors that use rear projection to illuminate a cylindrical wall, cylindrical doors, a conical ceiling, and a 12.5 m2 floor. For a user in the center of the YURT, the array or projectors provide effective resolution of 1 pixel per arcminute (approximately that of the human eye) and a visual surround of 3.8 radians (about 95% of complete surround). Within the YURT, users wear Volfoni 3D glasses and use an Aimon PS wireless wand controller, both tracked by an array of 8 OptiTrack Prime 13W infrared cameras on the YURTs ceiling. The visualization software used for the study was based on 3DVisualizer, a volume renderer built using the Vrui (Virtual Reality User Interface) toolkit, created by Oliver Kreylos at the UC Davis W.M. Keck Center for Active Visualization in the Earth Sciences [Kreylos, 2008]. We augmented 3DVisualizer to support illuminating the volume rendering with a light source at the user's head position and added a regularly spaced 3D grid to augment understanding of the space. Figure 5.3 shows a view of the dataset from within the YURT; for comparison, Figure 5.1 shows a typical clinical 2D rendering of the same scan. During the experiment we ensured a frame-rate of at least 50 stereo frames per second.

Figure 5.3: Close-up of a user tracing vessels in our VR setup. The red cone indicates the tip of the wand from user perspective. Annotating line segments are drawn between 2 wand positions confirmed by button presses.

5.2.3 Interaction Methods

The YURT offers basic interaction methods expected in a modern room-scale VR environment, allowing participants to move freely around the dataset to investigate it from different perspectives. The wand toolscan be used to move the dataset itself with six degrees of freedom. Annotations are created by placing separate line segments with the wand tool. A line segment is started at the tip of the conical wand marker with an initial button press and then stretches to follow the wand tip until its placement is finalized with a second button press. The visualized dataset is rendered as a solid volume, wand marker and line segments are therefore covered by the dataset surface whenever they extend into the volume, which aids the accurate placement of lines on the placental surface. Additionally, we gave participants two functions: an undo function to remove the most recently drawn segments, and a reset function to remove all annotations and return the visualization to its original state. Figure 5.3 shows a user drawing these line segments.

5.2.4 Participants

We recruited eight medical professionals from Rhode Island Hospital and Women & Infants Hospital of Rhode Island to volunteer as participants of in the study. They all had experience working with placental anatomy.

5.3 Procedure

The experiment was performed at our VR facilities at Brown University. We began by introducing participants to the equipment and the dataset. All available user interaction methods were demonstrated, and participants were given a short time to get used to controlling the environment. We also informed the participants of the camera in the YURT that recorded videos of each session and the microphone on headphones that recorded audio. Additionally, we explained that we were collecting tracings and interaction logs within the application during the experiment.

We asked our participants to practice tracing the edge of the dataset to get used to basic VR interactions. During this initial training stage, we specifically introduced participants to the concept of tracing in 3D space and instructed them to make use of spatial depth in their tracing process; we verified their understanding of the concept visually by inspecting test tracings left by the participants on a predefined target ridge feature. We asked participants to be as accurate as possible and imposed no time limit to allow free exploration of the visualization.

We then cleared all tracings and moved on to the trials. Each participant completed three trials on the same dataset, whose goal was to trace all the blood vessels on the placental surface. This task included searching the placenta for surface structures, decide what should qualify as a vessel, and marking the identified vessels with line segments, rendered as red cylinders. An example tracing is shown in Figure 5.3.

Because annotations were made with line segments, we instructed participants to use a larger number of shorter segments in areas of high vessel curvature to represent the shape accurately. Participants were also asked to express their thoughts about the data and task during running trials, so that we could record and study them to learn about the utility of the visualization and interaction. After finishing the three trials, the subject was given a free-form interview and asked about the visualization, the experience of exploring medical data in VR, and potential applications of the technology. Each participants session totaled approximately 30 to 45 minutes. The experimental protocol was approved by our universitys IRB.

5.3.1 Reference Dataset

To evaluate and analyze the obtained data, we created an expert set of tracings (Figure 5.4). This dataset was initially created by our fetal surgeon collaborator in the VR environment used for the experiment and benefited from his extensive experience with the data and equipment. However, this reference tracing still included minor offsets from the placental surface. We generated a more accurate reference dataset by projecting the tracing data onto the surface and manually retraced it with connected line segments at higher resolution using the 3DSlicer open-source medical visualization framework [Kikinis et al., 2014].

5.3.2 Metrics

To compare participant tracing data to the reference dataset it was necessary to re-project line segments onto the placenta. While some tracing segments were located exactly on the placental surface, most were drawn hovering over that surface. We recorded an average offset distance of 2.4 cm across all participants, with outliers of up to 9 cm. This can be attributed to the users inexperience with 3D VR interactions and the difficulty of simultaneously searching for vessels and keeping the cursor on the placental surface while tracing them without haptic feedback. To overcome this problem in participant tracings, we recorded their head positions at the moment they created each segment endpoint and used the lines passing through each pair of head and tracing positions to identify the intended line segments on the placental surface. This method let us use the distance between the line of sight for each tracing and the closest expert reference point to evaluate how much each participant results agreed with the expert about blood vessel locations.

To compare the obtained participant tracings to the expert dataset we defined all reference segments within 2 mm (patient scale; 13 mm within the virtual environment) of projected participant tracings as correctly identified. The error distance was chosen together with our collaborator based on the size of vessel features and the expected accuracy in our experimental VR environment. Within our study participant precision was therefore the length of line segments within error range of the reference tracing over total tracing length and sensitivity was the length of correctly identified reference line segments over total reference length.

5.3.3 Results

We found that participants were able to identify blood vessels of 1 mm diameter in our MRI VR visualization, a size relevant to the diagnosis and treatment of vascular diseases like TTTS. Figure 5.5 shows the results of all participant tracings, color-coding the segments which fell within 2 mm of the expert reference green, and those that did not red. Most participants achieved a tracing precision greater than 75% when evaluated against the expert reference tracing, with a lower bound of 59.8%.

Detailed analysis shows notable variability among participants with respect to the expert reference and to one another (Figure 5.6). Table 1 lists the individual quantitative results. While individual total coverage varied greatly, we see that most tracing results fell within the margin of error for each participant. We found that most participants were conservative in annotating vessels and often did not trace them as far the expert user, which in turn reduced their overall coverage ratio. As vessels become progressively thinner with increasing distance from the umbilical cord, their

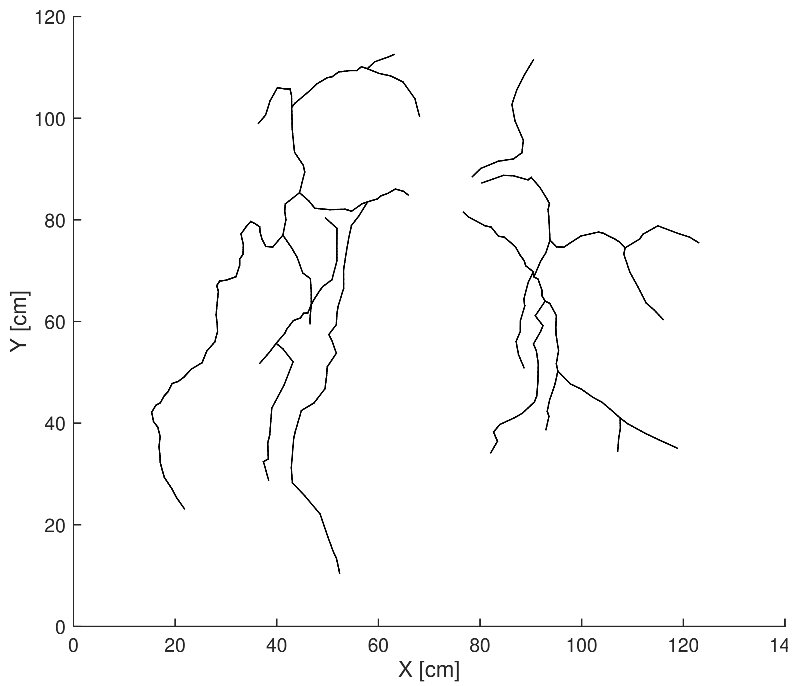


Figure 5.4: Reference tracing created by our expert collaborator projected onto the XY plane.

surface features get fainter and they start to blend in with scanning artifacts. Having less experience with placental vessel trees, most participants therefore stopped tracing early.

Some branches of the vascular tree were identified only by a subset of participants. We identified two reasons that multiple participants missed blood vessels. Several vessel branches in the lower left and lower right of the vascular tree showed comparatively faint features that were frequently overlooked. Additionally, the arcing blood vessel on the top right was reportedly difficult to identify because it followed the wall of the placenta perpendicular to the rest of the mostly planar vessel structure, making it more challenging to spot from our standard viewing position.

Beyond the reference dataset, we found that most participants marked supposed additional blood vessels around the central umbilical cord. False-positive annotations to the top left of the umbilical cord center (Figure 5.6.c, 5.6.d and 5.6.f) can be attributed to an artifact in our visualization. To give participants a sense of scale, we superimposed a 3D grid over the visualization, creating raster outlines and isolines as seen in Figure 5.3. The incorrectly marked vessels coincide with an isoline at the same location and have most likely been misidentified. Likewise, several participants misidentified a vessel right below the umbilical cord that is located at one of the visual artifacts created by the segmentation and removal of the fetus.

Apart from these two common areas, each participant annotated individual additional vessels,

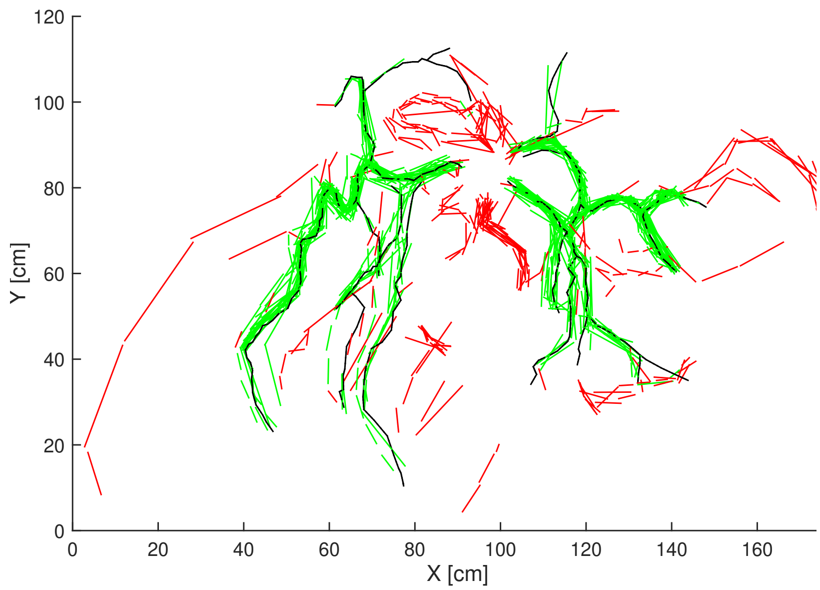


Figure 5.5: The segments of all participant trials overlaid on the expert reference in the XY plane. The expert reference is shown in black; segments with projections within 2 mm of the expert reference are shown in green; and segments outside that error margin are shown in red. Participant tracings are shown without projection to visualize the offset of tracings from the reference set.

sometimes tracing vessels farther than the expert and other times marking regions not considered in the reference data. However, most of these remaining annotations are in areas relatively far from the umbilical cord, making misidentifications more likely. The completion time for each tracing task varied greatly among participants and correlated weakly with the number of placed line segments (Table 5.1). Since no time limit was imposed, some participants spent more time exploring details of the dataset mainly during the first trial.

5.3.4 User Interaction

Analyzing video recordings and positional data of tracking data, we found that participants remained relatively stationary during the tracing task. On average, user head positions varied by 0.3, 1.2, and 0.7 m along the X, Y, and Z axes, respectively, relative to the dataset in real-world coordinates. This shows that participants did not make full use of the horizontal space the VR environment offered. We believe that this was caused by our method for placing annotation marks. To place line

Table 5.1: Quantitative results for participant performance.

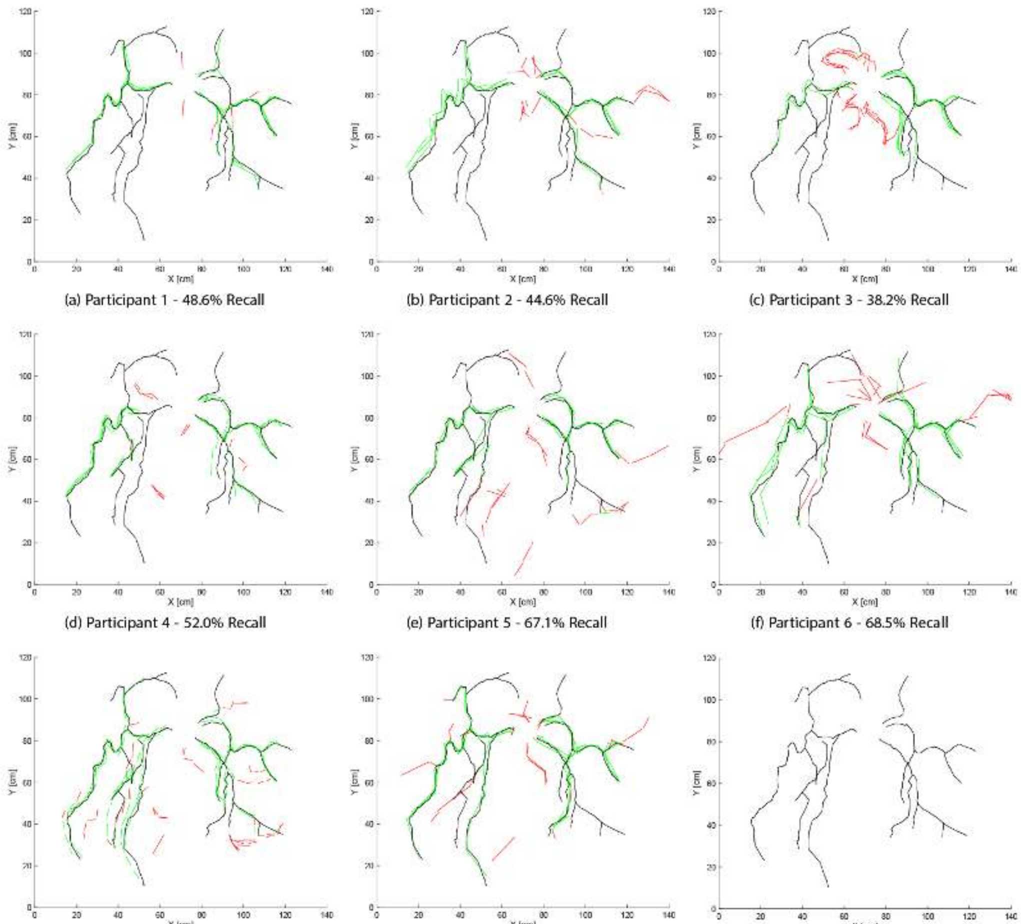
Participant	# of Segments	Precision	Sensitivity	Average Time
1	83	91.5%	48.6%	207s
2	98	74.5%	44.6%	235s
3	179	59.8%	38.3%	274s
4	161	85.1%	52.0%	384s
5	130	75.4%	67.1%	419s
6	105	65.7%	68.5%	301s
7	244	79.9%	85.2%	557s
8	166	79.5%	66.1%	338s

segments accurately, participants had to keep the wand in a stable position while moving their head to benefit from the parallax effect in VR. This is feasible when stepping forward, back-ward and while crouching, but more difficult when stepping to the side. We also report that participants rarely used the wand to rotate or translate the dataset and deduce that they deemed the default orientation good enough to solve the annotation task.

Figure 5.7 shows an example participant tracing from one trial along with lines indicating where the participants head was as each segment was drawn. These segments are shown along with a representation of the front screen of the YURT to give an idea of the space.

5.3.5 Discussion

In the context of the field of expertise of our medical collaborator, TTTS intervention, the study was an effective proof-of-concept. Participants were able to reliably identify blood vessels of 1 mm diameter, well within the vessel size targeted at surgical TTTS interventions. Additionally, they did this in an MRI dataset in which manual and automated vessel segmentation methods were not producing satisfactory results. This study is a first step towards using MRI visualizations to analyze placental vasculature without the use of a contrast agent. The expert and all participants agreed that the ability to view the entire dataset at large scale in VR is a major advantage over more confined views in desktop environments, in some cases stating that the experience was “amazing”. Based on participant feedback and experiment results we identified several topics that need further discussion:



(a) Participant 1 - 48.6% Recall (b) Participant 2 - 44.6% Recall (c) Participant 3 - 38.2% Recall (d) Participant 4 - 52.0% Recall (e) Participant 5 - 67.1% Recall (f) Participant 6 - 68.5% Recall (g) Participant 7 - 85.2% Recall (h) Participant 8 - 66.1% Recall (i) Reference dataset

Figure 5.6: Comparison of segments drawn across all eight participants; each plot is labeled by participant number and all three trials from that participant are overlaid. The expert reference is shown in black; segments with projections within 2 mm of the expert reference are green, and segments outside that area are red. Each plot shows the original user input without projection. Some green lines appear to be offset from the reference tracing; however, projecting them to the surface from their head position places them within the error margin. Tracings of participant 6 (f) were trimmed in this figure to match the scale of all plots, the full extent of false positive tracings can be seen in Figure 5.5.

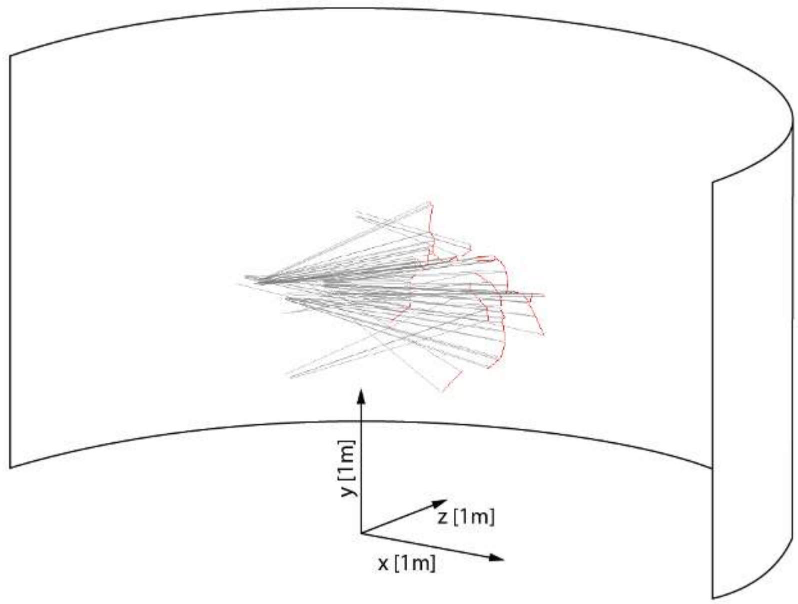


Figure 5.7: Example results from a single trial showing the tracing segments in red and the head-wand vectors in gray surrounded by a representation of the YURT screen.

User Interaction Methods

We found that an actual analysis and planning application in VR would need to consider some of our results. With the interaction tools currently available, participants in our study could not reliably keep their vessel annotations on the placental surface. As pointed out by Keefe et al. [2007], drawing on air is difficult; this is also true of precisely placing endpoints of 3D line segments. While some of this can be explained by the participants brief learning period and their inexperience with 3D VR interactions without haptic feedback, we suggest that the interaction tool should assist users in this task.

We see two ways of extending our current method to address this problem. First, modern VR controllers often include vibration motors to provide haptic feedback. Using this standard method would allow us to notify users whether they are currently touching the placental surface with the wand tool, in addition to visual feedback like highlighting the wand marker. The other option would be to use a ray interaction tool, like a virtual laser pointer, to interact with the data. This would let users point to the placental surface and have annotations snap to the intersection point between surface and ray. Similar ray-based selection techniques have already been demonstrated by Wiebel et al. [2012]. This method would allow users to step away from the dataset and use the full available space in the VR system to analyze the data while still letting them complete annotation tasks, but it would require more training to be used effectively.

Generalizing VR Environments

In this experiment we used the YURT as our base VR environment due to its high visual fidelity. Current consumer-grade VR systems, like the HTC Vive or the Oculus Rift, provide significantly lower resolution and field of view. The transfer of results between VR systems is an active re-search problem, and it is currently uncertain whether specific benefits can be transferred between different VR platforms. However, based on the interview feedback we gathered from study participants, one of the core benefits they see in this system is the ability to show a dataset at large scale and explore it with an intuitive way to select viewing positions. Most current HMD and CAVE systems can reproduce such an experience, and this strengthens our belief that our results can be applied to other available VR technology.

Applications

Our study participants commented on the potential benefits of this system in areas beyond TTTS intervention. One participant remarked that the technology might be useful in helping neurosurgeons find aneurysms, and another indicated that the same kind of visualization could help in facial reconstruction surgery by showing the surgeon a patient's anatomy in a noninvasive way. Additionally, participants also stated that this technology could be invaluable in education, whether in training surgeons by simulating the precision of movement required for endoscopic surgery or in educating students about the anatomy of the placenta or other parts of the human body. A recent overview paper by Olasky et al. shows that surgical training in VR is indeed a very active research area [Olasky et al., 2015].

Limitations and Open Questions

This work represents an initial pilot study for the effective use of VR visualizations in TTTS surgery planning. At the moment of publication, obtaining MRI scans was not part of the clinical practice for TTTS interventions. We therefore had to use the MRI scan of a singleton pregnancy. It was not possible to obtain the scanned placenta after the pregnancy, which prevented us from generating a ground truth dataset of the vascular structure via injection of a contrast agent.

While the expert dataset created within our VR application and improved with additional domain tools likely covers many major surface vessels, the lack of ground truth data prevents us from making statements about the completeness of our collected tracings. We hope that the positive outcomes of this work provide a justification to collect these required medical scans for future studies to evaluate the coverage of identified vessels and the identification of vessel intersection in actual TTTS cases.

5.3.6 Conclusion

We presented a case study evaluating our virtual reality system in finding placental surface blood vessels using a VR visualization of MRI data. We found that medical professionals can accurately identify relevant vessels of 1 mm diameter in our experimental VR visualization, a task critical to the treatment of placental diseases like twin-to-twin transfusion syndrome. Most participants achieved a tracing precision greater than 75% when evaluated against the expert reference tracing, with a lower bound of 59.8%. Our findings underline the importance of large-scale VR MRI visualizations, since we were able to visualize vessels in a scan taken without the use of a contrast agent.

On the level of user interactions, we found that study participants had difficulties placing annotations at the correct 3D depth within the VR environment. The recorded annotations exhibited view dependency: i.e., they appear in the intended location when viewed from the head position at drawing time but show depth deviation when viewed from any other point. This underlines the difficulty of 3D tracing based on visual cues without haptic touch feedback and the need for interaction methods that support users in this task. We report our insights into the VR interaction methods required to create effective immersive medical visualization applications.

Finally, interview feedback from study participants showed that the annotations generated in our experimental system can be helpful in analyzing and discussing individual vascular structures. Our application is a step towards a surgical planning VR environment for TTTS intervention.

Chapter 6

Designing VR visualizations with Scientific Sketching

One of the primary goals of studying perception in immersive displays is to understand the intricacies and benefits of VR visualizations and to use them to their full potential in actual applications. The immersion aspect of VR opens up the design space for potential visualizations. However, introducing domain scientists to these new options remains a challenge that calls for the evaluation of application design methodologies that take the novelty of VR displays into account.

In this chapter we present a design study of using virtual reality (VR) visualizations to analyze dinosaur footprint formation. Large-scale simulations of substrate flow have recently been used to explore the relationship between track morphology and foot movement by combining data from modern birds and fossilized specimens found in the field. However, the spatial complexity of these unsteady flow datasets makes it difficult to analyze them using off-the-shelf visualization tools. We designed multiple VR visualizations that help paleontologists explore their simulation data with visual metaphors tailored to their specific research questions. The iterative development process spanned a period of two years with frequent progress meetings. An integral part of the development was the inclusion of students in a VR visualization design course. These students sketched potential visualization and interaction techniques in VR, guided by our collaborators using the *Scientific Sketching* design methodology described below [Keefe et al., 2008].

We hypothesize that the engagement and immersiveness virtual reality offers can be leveraged to help knowledge workers, including analysts and scientists, with higher-level cognitive tasks. Several properties of VR environments, such as the feeling of presence of users in immersive 3D scenes, improved spatial cues, and the use of spatial movements to interact with data have potential benefits for users in complex data analysis tasks. These properties have been used in training and education

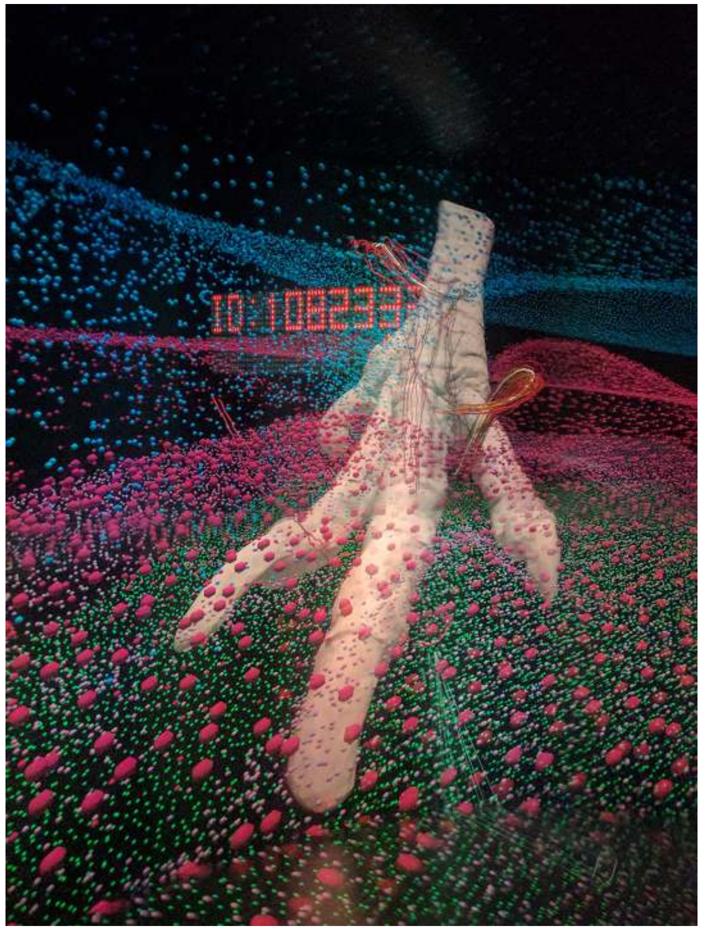


Figure 6.1: Example visualization of particle flow during dinosaur track creation. To avoid visual clutter we only show four thin layers of simulated particles, color coded based on their initial depth (from blue on top over red and magenta to green on the bottom). A tooltip display indicating selected particle ID's can be seen in the background.

applications, as well as VR games, but not as much for higher-level knowledge retrieval tasks.

Developing an effective set of visual codings and metaphors to answer specific research questions is one of the core challenges in visualization research. The main research questions are, however, rarely finalized at the start of the visualization development. They often evolve and change alongside the iterative development process of a visualization application. A better understanding of their data allows domain scientists to give effective feedback and steer the development process in the right direction. Virtual reality is a relatively recent visualization medium and few domain scientists have experience in using it to its full effect. It was therefore important for us to provide our collaborators with an overview of possible visualization styles in VR.

We based our design process mainly on the Scientific Sketching methodology proposed by Keefe et al. [2008]. This required us to introduce a third discipline to our development process, artists from the Rhode Island School of Design (RISD) with a background in visual design. Scientific sketching aims to combine the individual expertise of scientists, visualization experts, and artists to create novel solutions to visualization problems. It is separated into four phases that are inspired by artistic work processes; *Paper Sketching*, *VR Sketching*, *VR Prototyping* and finally *Implementation of visual specifications*. We realized these stages by recruiting students (visual design and computer science majors) through a course on immersive scientific visualization design that we offered two separate semesters.

This work contributes to the field of scientific VR visualization in multiple ways:

- Implementation and qualitative evaluation of multiple VR visualization methods for unsteady flow visualization. Out of a wide range of potential methods, the visualization techniques selected by our collaborators might be applicable to other flow visualization applications.
- A case study of Scientific Sketching and its effectiveness as design methodology in a long-term visualization project.
- Insights and best practices for the effective use of Scientific Sketching.

This chapter is based on:

J. Novotny, J. Tveite, M. L. Turner, S. Gatesy, F. Drury, P. Falkingham, and D. H. Laidlaw. "Developing virtual reality visualizations for unsteady flow analysis of dinosaur track formation using scientific sketching." *IEEE Transactions on Visualization and Computer Graphics*, 25(5):2145-2154, May 2019.

J. Tveite contributed to the implementation of the visualization application. The other authors provided their domain knowledge to the project and contributed to the writing of the publication.

The work is organized as follows: Section 6.1 provides an overview and related work about the scientific problem of our collaborators and 3D flow visualization in general. Section 6.2 introduces the individual stages of the design process, the simulation data used for the final implementation, and CAVE VR environment used for visualization sketching and development. Sections 6.3 and 6.4 summarize and discuss the findings of our two-year development process. Finally, we present conclusions and potential guidelines for the development of future VR visualizations.

6.1 Related Work

In this section we introduce prior work related to our visualization research and the scientific problems we are attempting to address. We provide details about dinosaur track analysis and how 3D particle flow simulation is used to infer foot motion from fossilized tracks. We then list established 3D flow visualization techniques and how they are linked to our visualization metaphors.

6.1.1 Dinosaur Track Analysis

Dinosaur tracks are relatively common records of Earth's fauna during the Mesozoic Era. Unlike a fossilized skeleton, a track is evidence left by an animal while it was alive. Such trace fossils can often be attributed to different species or groups, providing insights into population distribution and paleoecology. Even a single track holds potentially valuable information about an animals movement and behavior. However, extracting reliable inferences from a footprint's final morphology is rarely straightforward [Falkingham and Gatesy, 2014].

Tracks are not static molds of the foot, but rather the result of a dynamic interaction among anatomy, movement, and substrate. As a dinosaur contacts malleable ground, it deforms not only the exposed surface, but sub-surface layers as well. On very soft mud the foot can sink quite deeply, passing through multiple layers as it moves through the sediment volume. Displaced material can pass around the toes and collapse, or be dragged along with the motion. The amalgamation of all these events reorganizes the particles in the fossil's volume, which can be split open to reveal track surfaces at multiple levels. An understanding of sediment flow during formation is thus critical to the correct interpretation of these specimens.

Analyzing interactions below ground is hindered by foot and substrate opacity. Experiments using model indenters have been fruitful, as have observations of extant animals like birds, which have very similar feet to those of predatory dinosaurs [Allen, 1989, Milàn and Bromley, 2006, Gatesy et al., 1999]. Recently, sub-surface imaging of foot motion by multiple X-ray cameras has been combined with discrete element simulation to reproduce track formation sequences in substrate volumes [Falkingham and Gatesy, 2014]. The resulting datasets, consisting of millions to tens of millions of dynamic particles as well as a moving foot model, present challenges to visualize, explore, and interpret. Our work aims to visualize the particle flow within these datasets, providing paleontologists with insights about dinosaur locomotion and the origin of a tracks morphology.

6.1.2 3D Flow Visualization

The analysis of 3D flow data is a common task in a wide range of scientific fields and numerous methods to visualize internal flow structures have been proposed so far. Survey papers by McLoughlin et al. [2010] and Brambilla et al. [2012] provide a comprehensive overview of the state-of-the-art in flow visualization. The methods used in our VR application are based on geometric flow visualizations which represent the flow data as discrete geometric objects.

Our dinosaur track simulation data is defined as a time-varying flow field using the Lagrangian specification. This allows us to directly visualize particles in 3D (see Fig. 6.1) and show their movement as animation. To visualize particle movement over the entire simulation in a static way we use integral curve representations similar to those introduced by Zöckler et al. [1996]. In particular we utilize 3D pathlines, which connect particle positions of subsequent simulation timesteps (see Fig. 6.7).

To gain insight into particle movements over the whole volume of a flow dataset, it is necessary to visualize the relative movement of multiple particles and particle groups. Our visualization approaches are based on the concepts of 3D path and time surfaces. Pathsurfaces, as discussed by Schafhitzel et al. [2007], are the integral 3D surface created by following the movement of a connected line of particles through time. We use a similar method to visualize the movement of foot geometry during the simulations (see Fig. 6.12). To analyze relative particle movement between two subsequent timesteps we use time surfaces, introduced by Krishnan et al. [2009]. Their method defines a surface of connected particles in a single timestep of the dataset and then follows this surface through time. This effectively visualizes local changes in particle neighborhoods as surface deformations. This visual metaphor is particularly effective, since it shows similarities to the deformed substrate layers found in fossilized dinosaur tracks.

6.2 Application Design Methodology

In this section we introduce the research questions of our paleontologist collaborators and the design methodology we used to address them. We combined the collaborative efforts of students, faculty, and scientists in an interdisciplinary VR design course to create effective VR visualizations. Our design process is based on the Scientific Sketching methodology introduced by Keefe et al. [2008], which we describe in detail in the remainder of this section. This application development approach aims to efficiently coordinate the work of artists (in our case, RISD and Brown students), visualization experts (visualization majors and faculty) and domain scientists (paleontologists) in VR visualization projects. The development process is split into four successive stages; *Paper Sketch-*

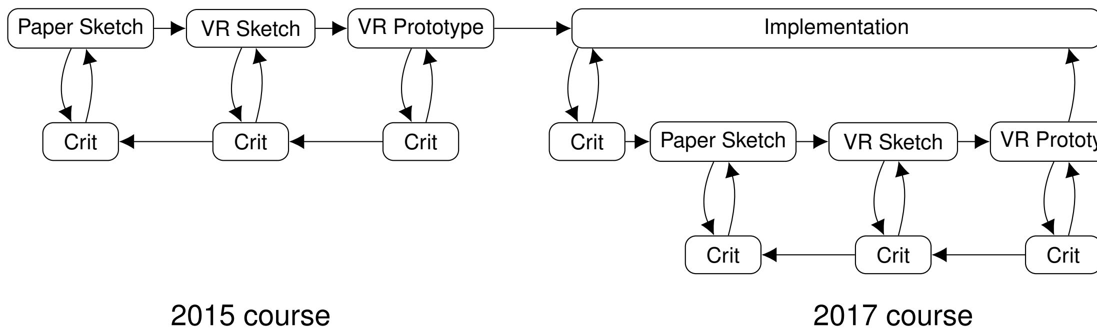
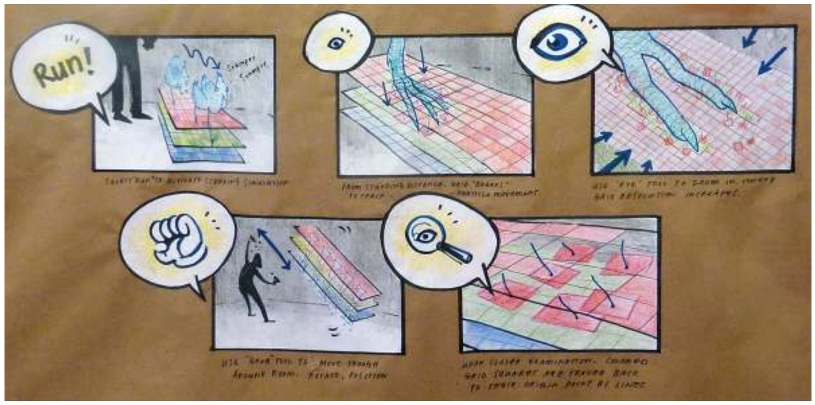


Figure 6.2: The project development timeline showing how students of two VR visualization design courses contributed to our visualization outcome. Our implementation stage started directly after the completion of the 2015 course. Insights found during the implementation stage led to improved feedback during the 2017 course. Finally we combined the outcomes of two Scientific Sketching projects into our collaborative flow visualizations.

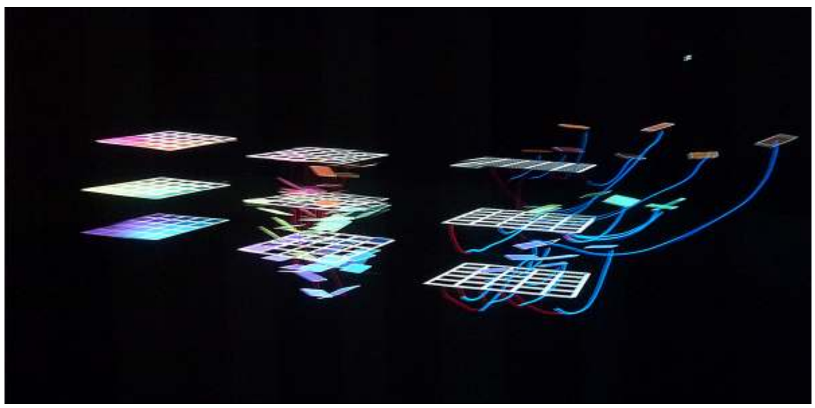
ing, *VR Sketching*, *VR Prototyping* and *Implementation of visual specifications* (see Fig. 6.2). In each stage the three participating groups have a different set of responsibilities to fulfill.

In the initial paper sketching stage, for example, the scientists' role is to provide background that explains their scientific problem. The other two groups need to be able to understand the problem and the characteristics of the underlying data, such as relationships between data variables and their relative importance [Keefe et al., 2008]. The artists' role is to take the provided information and come up with a large quantity of visual ideas, exploring the problem from a variety of different angles. The visualization experts join the sketching effort, but also mediate the discussion between artists and scientists to keep track of the overall visualization goals. In this early stage, the focus lies on covering a wide range of visualization concepts without regard for implementation complexity.

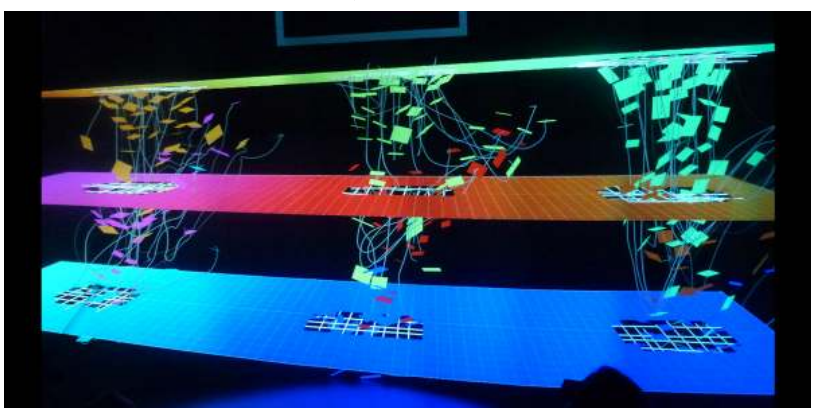
An important tool throughout all stages is the artistic critique, as used in art and design education. It is a careful and critical group discussion evaluating specific aspects of visual artworks. Any feedback given in a critique session needs to be well-founded and explained in detail (i.e. "I do not agree with this color choice, because [...]") to serve as a starting point for a constructive discussion [Trumbo, 1997]. Within the scope of this project, critiques were used to argue about the effectiveness of individual sketches and prototypes. Involving all three participating groups into these critiques ensured that each proposed visual concept was discussed from artistic and scientific viewpoints.



(a) Paper Sketch



(b) VR Sketch



(c) VR Prototype

Figure 6.3: Example of the evolution of a student idea for visualizing the deformation of different substrate layers during track formation. The paper sketch (a) includes proposed visual representation and interaction techniques. The VR sketch (b) is used to discuss the visual concept with science collaborators before creating a high-quality VR Prototype (c).

Table 6.1: Group Compositions during course the two iterations of the Virtual Reality Design course.

	Course 2015	Course 2017	Total
Scientists	2	2	2
Art/Vis Faculty	2	2	2
Brown Computer Science	8	6	14
RISD Students	5	6	11

6.2.1 Paper and VR Sketching

We developed our VR visualizations over the course of two separate semesters of the Virtual Reality Design course offered by Brown University and the Rhode Island School of Design. a local fine-arts college. A total of 11 RISD students, 14 Brown University computer science students and one faculty member from each school participated in the design of the application. Two paleontologists from Brown University presented the scientific problem, using data from their collaboration project with Liverpool John Moores University (Table 6.1). Before starting the Scientific Sketching projects, all students participated in introductory exercises covering basic 2D and 3D visualization methods as well as core VR drawing techniques.

In both iterations of the course, the paper and VR sketching stages were part of mandatory class assignments. The assignment started with an interactive lecture by the science collaborators to provide students with the background knowledge needed for the research questions at hand. The underlying data were introduced through media presentations using images and videos rendered with their desktop-based visualization tools Ovito [Stukowski, 2009] and Autodesk Maya [Autodesk, 2018]. Actual dinosaur footprint fossils and cast duplicates were used as physical examples of the visualization subject. The initial research questions posed by our collaborators were the following:

- Can we work backward from a track surface to determine the original configuration of its particles in the starting plane? Where did each particle come from?
- Alternatively, can we trace the fate of particles in the starting plane forward in time? Which particles will descend, ascend, move forward, or collapse to form the features of the final track?
- How do foot-particle and particle-particle forces move sedimentary particles? When and where do compression (push), tension (pull, cohesion), shear, or gravity dominate?
- How can dynamic, simulated data of moving particles be compared to the static final mor-

phology of a real fossil dinosaur track?

After the introductory lecture, students were given one week to create paper sketches for the first critique session with collaborators. To encourage creative designs, a wide range of visual and physical art-forms were permitted as potential hand-ins (e.g., digital art, 3D printing, and sculpting). Intermediate results of this stage can be seen in Fig. 6.3a, 6.8, and 6.12a. After two critique sessions in which each sketch was discussed at least once, students had to realize at least one of their paper sketches as 3D VR sketch. To complete their assignment, students used a VR painting application called CavePainting [Karelitz et al., 2003] in a high-resolution CAVE-like display room called the YURT (see Section 6.2.5). The application allowed students to draw simple textured primitives like tubes, planes, and text using an easy to understand wand-based interface. This allowed them to create and test immersive sketches of their proposed visualization within a short timeframe. Loading of 3D models created in external applications was prohibited, which forced students to complete the design task within the VR environment. This constraint ensured that students understood the interaction concepts and UI challenges of working within an immersive display.

The resulting VR sketches were discussed in a second round of critiques. In these sessions the paleontologists also acted as users of the sketched visualizations, and gave feedback on how they would interpret them in their current form and whether they reflected their understanding of the data. Results of the VR sketching stage are shown in Fig. 6.3b, 6.8, and 6.12b.

6.2.2 VR Prototyping

A VR prototype, as defined by the Scientific Sketching method, is a highly refined mock-up of the developed application. This includes more carefully drawn visuals, but also the use of animations to simulate interactive usage scenarios of the prototyped applications. CavePainting supports this prototype development by offering drawing layers to switch between visualization views and an animation frame system. Scenarios can then be played out by simulating the application behavior through Wizard-of-Oz interaction techniques [Buxton, 2007]. In practice, students realized this by creating a set of static scenes. During the presentation they stood in the back and switched scenes in response to users interacting with their prototypes to showcase their proposed interaction concepts.

Creating the VR prototype was the four-week final project in both classes. Students were free to pick a specific hypothesis to explore in their prototype. The assignment was to design one or more VR visualizations including a step-by-step scenario of how they would be used by scientists. To ensure that students developed their ideas in fruitful directions, the two sketching stages were repeated in the first two weeks of the final project, focusing on the interaction storyboard. Critique sessions continued during class time and students had the option to return to earlier sketching stages (e.g.

paper sketches) for idea refinement. A majority of students picked hypotheses addressing research questions in dinosaur track simulation, leading to several high-quality prototypes. Examples can be found in Fig. 6.6, and 6.8.

6.2.3 Implementation

The final projects of the 2015 course provided us with several detailed prototypes of potentially helpful VR visualization concepts. We implemented a subset of these concepts in an interactive data-driven application for exploring their most recent simulations. To keep the development on track, progress meetings with our paleontologist collaborators were scheduled multiple times every month. This tight feedback loop allowed us to extensively test and refine visualization ideas from the VR prototypes. The results of this stage with the corresponding sketches and prototypes are discussed in section 6.3.

During the Implementation Stage, the VR Design course was offered a second time, bringing in a new group of art and computer science students. Based on their experience with the VR application and new data from fossil tracks, scientists were able to provide clearer, more refined, research questions to the class. Resulting VR prototypes of that course were focused more closely on sediment and foot movement in dinosaurs as well as birds. This allowed us to use the student projects as inspirations for further refinements to the ongoing implementation process.

In the next two sections we provide details about the datasets and VR environments used during the implementation stage.

6.2.4 Dinosaur Track Simulation Data

The flow data used in our study was obtained through a combination of physical experiments, animation, and particle simulation. To study track formation in a living bird, the sub-surface walking kinematics of a chicken-like species (guineafowl) were recorded with a biplanar X-ray system as they walked through radiolucent artificial muds of varying depth and hydration, as well as a dry sand analog (poppy seeds). The recorded foot motion data were then processed using the X-Ray Reconstruction of Moving Morphology (XROMM) technique [Brainerd et al., 2010, Gatesy et al., 2010] to obtain an animated 3D model of the foot geometry. To simulate the sedimentary particle flow during footprint formation, one footfall was recreated using the discrete element method (DEM) [Falkingham and Gatesy, 2014]. The simulation was computed using LIGGGHTS [Kloss and Goniva, 2011]. The resulting simulation data (9.5 million particles, 523 frames) was initially visualized in Ovito [Stukowski, 2009], which was used for cropping and downsampling (370 thousand particles, four slabs).

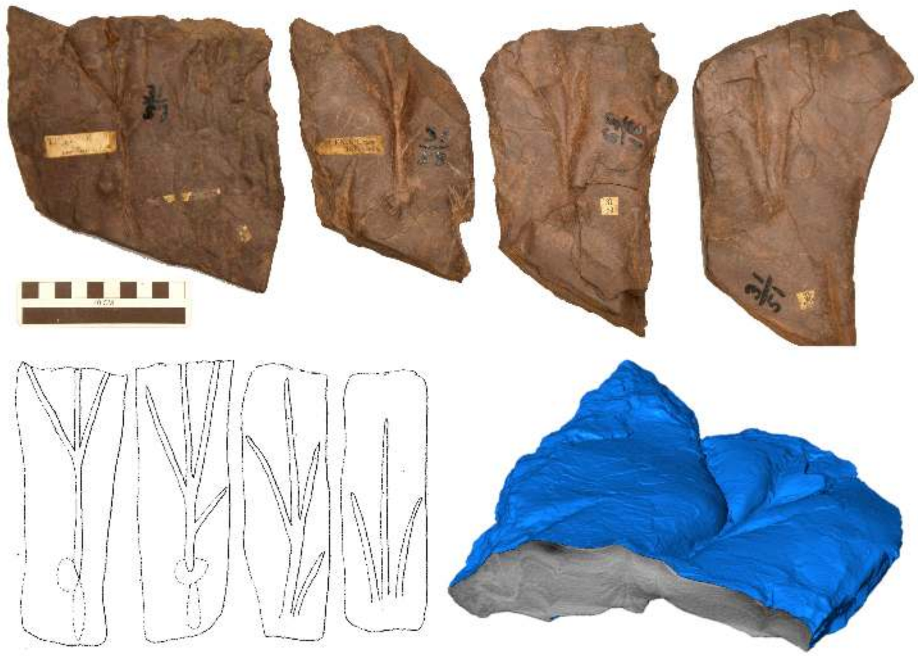


Figure 6.4: Paleontological motivation. (Top) Early Jurassic dinosaur tracks chosen for visualization (scale bar equals 10 cm). (Lower left) Interpretive illustrations of the four surfaces from this track volume highlight changes in morphology with depth [Hitchcock, 1858]. (Lower right) Deformations of the internal laminations are visible in a rendered section of the second slab based on CT scan data.

To test hypotheses of fossil dinosaur track formation, four specimens from the Beneski Museum of Natural History, Amherst College (ACH-ICH 31/51, 31/57-59; Fig. 6.4) were chosen for imaging and analysis. These four slabs, collected in the mid-1800s from the Early Jurassic (200 million years ago) rocks of Wethersfield, Connecticut, form a stack bearing a deep dinosaur track [Hitchcock, 1858]. Specimens were CT scanned to visualize internal deflected laminations, which were reconstructed in 3D with Amira (Fig. 6.4). Using these fossil data as a constraint and current understanding of guineafowl subsurface kinematics as a guide, an articulated dinosaur foot model was animated in Autodesk Maya. A LIGGGHTS DEM simulation (33 million particles, 174 frames) was first cropped (4.2 million particles) before further downsampling (70 thousand particles, four slabs) in Ovito.

6.2.5 Virtual Reality Environment

Our visualization uses the YURT (YURT Ultimate Reality Theater) VR display room located at Brown University [Kenyon et al., 2014]. It is equipped with 69 high-definition stereo projectors that use rear projection to illuminate a curved wall, curved doors, a conical ceiling, and a 135 ft^2 floor. When standing in the center it effectively provides retina resolution on its 190-degree front wall (1 pixel per arcminute). In that position the YURT covers 95% of the users field of regard.

Stereo is provided using Volfoni 3D active stereo glasses with a shutter frequency of 120Hz. Users interact with the VR environment using an Aimon PS wireless wand controller. Glasses and wands are tracked by a OptiTrack Prime 13W optical tracking system with an array of 8 infrared cameras mounted in the ceiling of the YURT (Fig. 4.2).

We developed our VR application in C++, based on the MinVR 2.0 framework [Acevedo-Feliz et al., 2014, Keefe, 2018]. This allows our application to work with a variety of different VR systems including the YURT, the HTC Vive and the Oculus Rift. It is also possible to run it as a regular desktop application with reduced functionality, due to the lack of 3D input devices in that mode.

6.3 Results

This section summarizes the outcomes of the initial three Scientific Sketching stages and discusses the iterative implementation process of selected VR prototypes. Stepping from hand-drawn sketches to interactive data-driven visualizations is a significant development effort. At the same time it is not guaranteed that sketched visualization methods work as effectively as planned once they are tied to actual data. During our implementation stage we went through several iterations to refine the resulting visualizations. Since this is an integral part of the Scientific Sketching method we not only present the final results, but also intermediate implementation stages and the original prototype visualizations they evolved from.

6.3.1 Particle Visualization

To start the VR application development stage, we implemented geometry-based particle visualization, similar to the methods available in our collaborators' Ovito environment. The goal was to provide a baseline visualization of the data as a comparison point for the upcoming VR prototype implementations. The baseline included the ability to manually step through simulation time-steps, to visualize the 3D foot models used to create the datasets, and to color particles based on their location. To visualize a high number of particles at interactive framerates we rendered particles on quad-billboard primitives, using OpenGL geometry and fragment shaders to create accurate sphere representations.

The particle visualization implementation went through several iterations to fit our collabora-

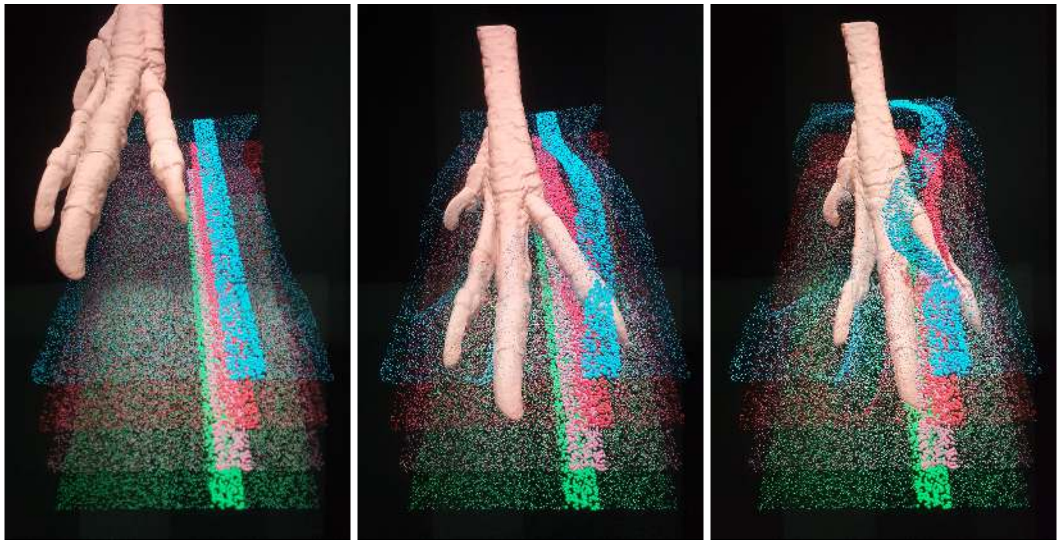


Figure 6.5: An example of the Slice-tool, highlighting the particles of a thin user-controlled slice of the dataset. Substrate deformation is visualized by keeping particle highlighted based on their location in the first simulation timestep.

tors' needs. Showing the full volume of particles in the dataset, for example, caused occlusion problems. Particles in the outer regions of the dataset would block the view of those close to the foot geometry. To uncover the moving particles of interest within the volume, we sub-sampled the data and visualized thin horizontal particle slabs as representatives of the substrates unsteady motion. The empty space between these horizontal slabs is critical for unobstructed views of both top and bottom surfaces within the volume as they deform. This reduction of the dataset greatly improves the interpretability of visualized particle motion. Another early feature request was an interactive way to adjust the diameter of all particles to control the trade-off between occlusion and particle legibility.

Even with sub-sampling and particle size adjustments, our collaborators wanted to explore smaller regions of known initial position. Based on their particle selections requirements we designed a slice tool to highlight a subset of particles of interest. With it users can control and place a 3D slice inside the dataset. Particle sizes outside the region of the slicing tool are reduced to effectively highlight selected particles and make them easier to follow throughout the animation. The slicing tool selects particles based on their starting position, which highlights their displacement over multiple simulation timesteps (Fig. 6.5). Alternatively, users can highlight particles within the slice during the actively shown timestep.

The first VR prototype implementation was particle filtering based on their total movement distance throughout the simulation. Particles that are not moving are of little interest to the research

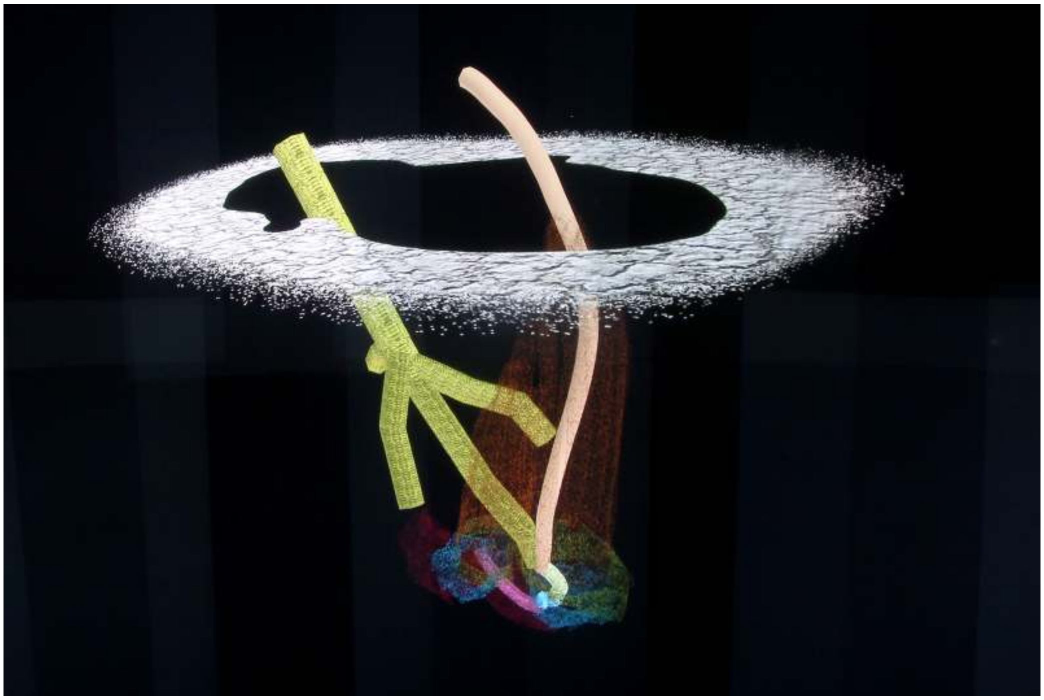


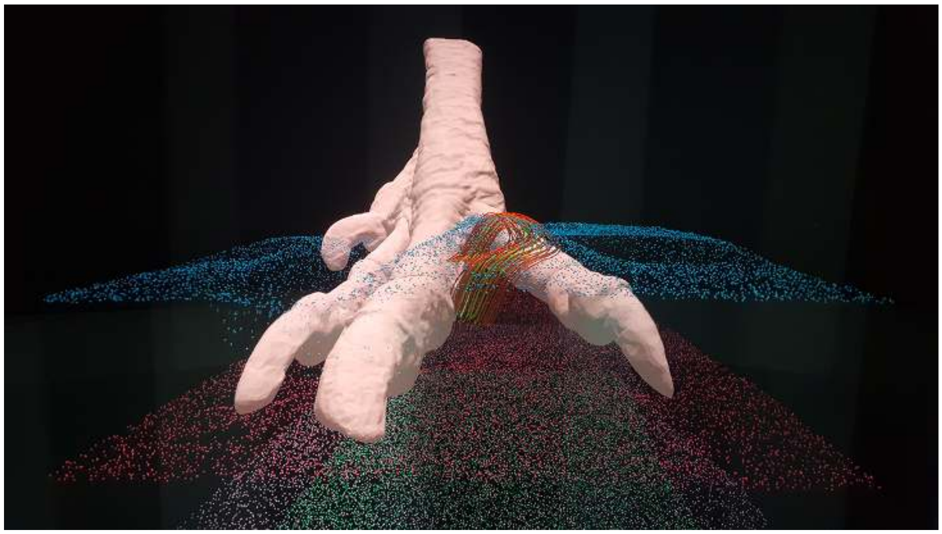
Figure 6.6: A VR prototype outlining the idea of pathline clustering around a single selected path. The student sketched out two clusters next to a selected seed pathline following the center toe tip. Pathline and clusters are colored to highlight different parts of the step sequence (foot descent, step, and foot retraction).

questions of our collaborators and we initially removed them from the visualization, as proposed in the VR prototype. However, these unmoving particles provided important context for moving particles, by representing the surrounding substrate. They were therefore visualized as small, neutral-colored particles in later visualizations.

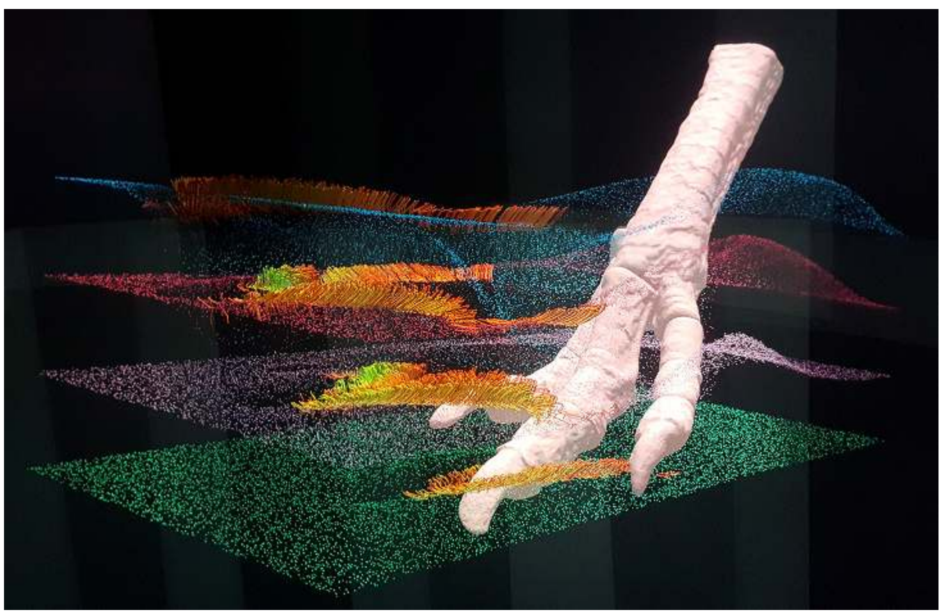
Particles alone, however, were not able to capture larger-scale particle movement patterns. Out of the numerous visualization ideas created by students during the prototyping stage in Scientific Sketching, two were of particular interest to our science collaborators: Pathline visualizations emphasizing similarity in substrate particles movement (see Fig. 6.7), and visualizing thin layers of particles as deformable surfaces, comparable to what is seen in fossil dinosaur tracks (see Fig. 6.8 and Fig. 6.9).

6.3.2 Pathline Visualization and Clustering

Following the motion of individual particles throughout the simulation is a frequent exploration task in our collaborator's datasets. Several students proposed VR prototypes based on pathlines



(a) Narrow pathline clustering between two toes.



(b) Wide pathline clustering showing multiple similar pathlines on all particle slabs.

Figure 6.7: Examples of implemented final pathline visualization. (a) Clustered pathlines of particle movement between two toes of the dinosaur foot. (b) A widely extended selection of pathlines at the borders of the dinosaur track. The relatively simple path shape is found in multiple particle slabs.

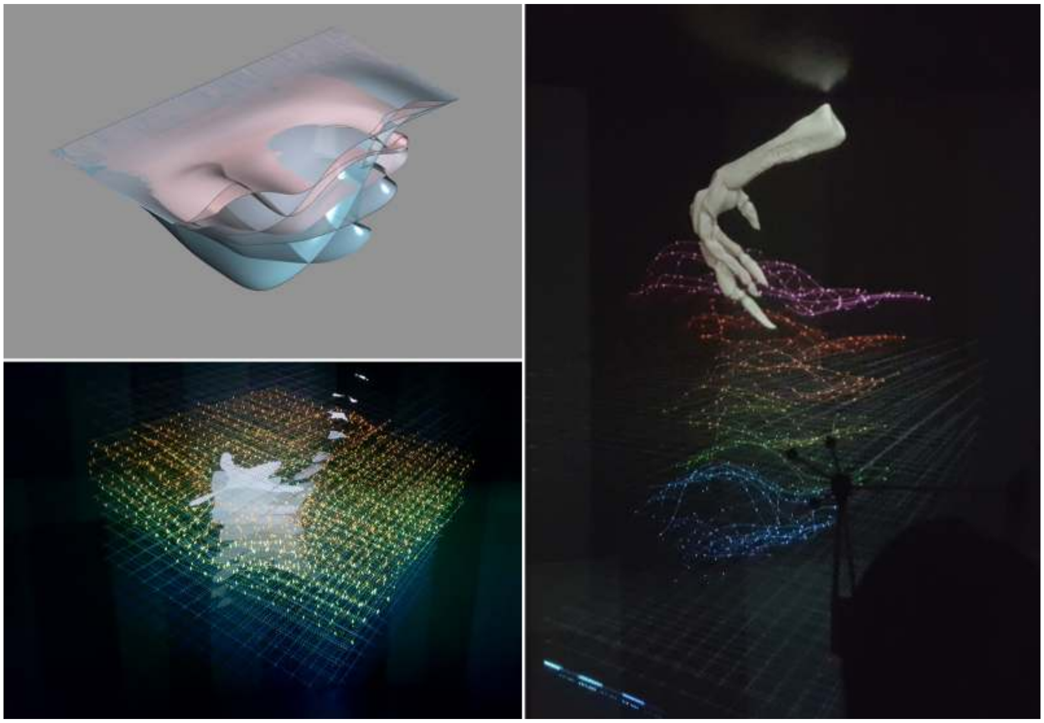
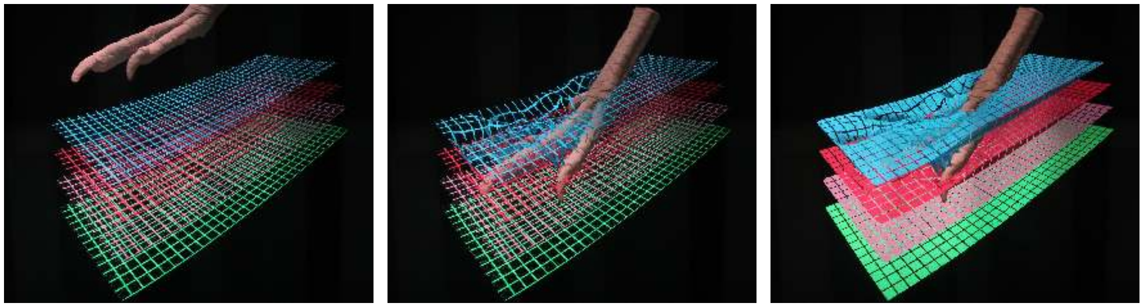


Figure 6.8: Paper and VR sketches of ideas to visualize overall substrate deformations by showing surface representations at varying depths of the particle volume. Implementation results based on that idea can be found in Fig. 6.9.

to aid users in the analysis of substrate movement. We have subsequently implemented pathline visualizations that can be activated by selecting any particle with the wand tool. However, single pathlines are rarely enough to gain insights into substrate movements as a whole.

To overcome this issue, several students have proposed to investigate multiple pathlines throughout the entire dataset at the same time (see Fig. 6.3) or to cluster multiple pathlines into an aggregate visualization (see Fig. 6.6). We implemented both methods and evaluated them with our collaborators. To create a meaningful overview visualizations with moderate levels of occlusions it is necessary to reduce the overall number of visualized pathlines. We implemented several well-known sampling techniques to select pathlines that captured critical substrate motion while still providing legible visualizations. These techniques included similarity measures based on curve properties such as critical points [McLoughlin et al., 2013], Poisson disk sampling to select a series of non-overlapping pathlines [Helgeland and Elboth, 2006], and using PCA to summarize the pathlines and then select dissimilar ones [Ferstl et al., 2016]. However, the resulting visualizations still contained many pathlines that our collaborators were not interested in, while lacking detail in areas of critical substrate movement.



(a) Regular Grid (b) Regular Grid (c) Tile Grid

Figure 6.9: Example views of the horizontal time surface visualization with a regular grid (a,b) and a tile grid (c) texture. The tile grid offers a more legible representation of surface deformation, at the cost of increased occlusion of the foot model and underlying surface layers.

(a) Initial deformed time surface (b) Intermediate state (c) Final surface position

Figure 6.10: Example views of the vertical time surface visualization with horizontal guidelines. The deformed time surface in (a) and (b) gradually converges to a flat surface in the final timestep (c), visualizing the origin of all particles on a surface in a way that mimics natural fossil cross sections (Figure 6.11).

To give them better control over pathline exploration, we implemented an adaptive selection technique that allowed them to extend a chosen ”seed” pathline to an interactively-sized group of pathlines with similar shape characteristics (see Fig. 6.7). We define pathline similarity as the least-squares distance between point pairs of two paths. To make the metric translation and rotation invariant, we align the paths using the rigid transformation that minimizes pathline similarity. The transformation is calculated using the method introduced by Sorkine-Hornung and Rabinovich [2017]. This metric is the result of empirical testing and fine-tuning over the course of multiple development cycles to fit the needs of our collaborators. It captures similar particle movement in multiple regions of the dataset and is time-dependent to avoid clustering substrate movements of different track creation stages (e.g. entry vs exit motion of the foot). We visualize the similarity of pathlines to the original selection as a color gradient from green (highly similar) to red (least similar). Our collaborators report that this method of pathline visualization has enabled them to 'learn' the dataset and explore expected and novel patterns of particle motion. The color gradients have been noted to be a particularly effective and non-overwhelming means of representing a large amount of pathline similarity among the data.

6.3.3 Time Surface Visualization

Another important analysis task in track simulation data is to investigate and compare substrate deformations within particle regions of the volume throughout the simulation. Particle visualizations in distinct slabs give a rough overview of the total deformation, but information about initial neighborhoods is often lost in the complex particle motion paths. To address this problem it is helpful to think of slabs as continuous surfaces instead of individual particles. This representation is referred to as time surface [Krishnan et al., 2009]. Several students have proposed and sketched visualization ideas based on virtual surface deformation (see Fig. 6.3 and Fig. 6.8). Our paleontologists were intrigued by the concept, since it effectively mimics the thin layer structures found in many fossil samples (see Fig. 6.4). This provides a direct comparison between simulation outcomes and field data.

Horizontal Time Surfaces

To implement this visualization concept, we reconstructed slab surfaces by using Delaunay triangulation on particles within thin horizontal slabs. Opaque flat-shaded rendering of these surfaces proved to be not effective. Each slab covered a wide area within the volume and often occluded lower layers and the foot geometry itself. Additionally, the lack of a geometric pattern on the surface made it difficult to follow more complex deformations. To solve this problem we applied a semi-transparent horizontal grid texture to the surface. Locking the texture coordinates during the initial animation step ensures that the grid deforms while following the particle motion. Overall substrate deformations can then be analyzed by observing the distortions of grid cells (see Fig. 6.9). The size and line thickness of the grid texture impacted the legibility of the visualization. More transparent grid layouts often allowed too much view on underlying surface slabs and made it harder to follow individual grid cells.

Based on promising prototypes developed during our courses, we inverted the grid texture to give it a tile-like design (Figure 6.3 and Fig. 6.9). This gave us better control over the visibility of underlying surfaces and made the visualization of deformations more legible. Our collaborators found this to be a very accessible visualization technique that provided them with new insights into patterns of substrate movement within the horizontal plane of the slab.

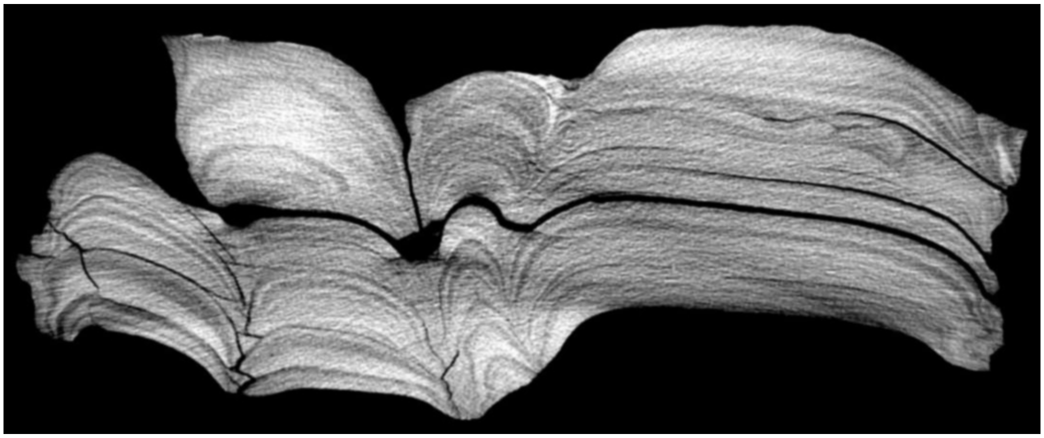


Figure 6.11: Cross-section of a fossilized track showing deformed horizontal material layers at regions the dinosaurs toes passed through. Our vertical time surface visualizations mimic this effect through surface color and horizontal guidelines.

Vertical Reverse Time Surfaces

Based on the successful implementation of time surfaces in horizontal slabs, we attempted to apply the same method on vertical slabs through the particle volume. We found that in this case, our collaborators were interested in the origin of particles within vertical slabs at the last timestep of the simulation. We implemented this by triangulating the surface in the final timestep, while selecting grid texture coordinates and color based on the particle location in the first frame. This results in a ”reverse” time surface, that starts in a heavily deformed state and gradually approaches the vertical plane over the course of the simulation (Figure 6.10). The deformed horizontal gridlines on the final flat surface correspond to layer structures found in cross-sections of fossils. This makes it an intuitive metaphor for paleontologists and allows them to effectively analyze the sediment movement that lead to the final state of the fossil. It also provides another effective comparison point between simulation outcomes and real-world fossils (Figure 6.11).

6.3.4 Foot Motion Visualization

The difference between actual foot motion and the shape of the final track is another research question of our collaborators. Critical points in a track, like toe entry points, often shift in position during substrate deformations. To gain insights into these offsets, students have prototyped visualizations of pathsurfaces created by tracking the central axis of each bone in the geometric foot model throughout the simulation (see Fig. 6.12). This visualizes the entire foot motion sequence as a simplified static 3D model. In combination with the previously introduces particle and time surface visualizations, this data representation allows researchers to explore the offsets between actual foot motion and substrate deformation in great detail (Figure 6.13).

6.4 Discussion

In this section we discuss our results and experiences with the Scientific Sketching design process. These include the advantages and disadvantages of using a large group of students to represent the artists, the overall results of the project, and lessons learned from employing this design methodology.

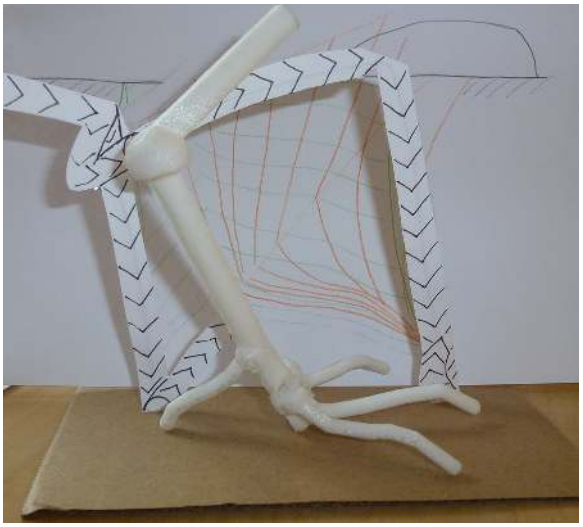
In the early sketching stages, as students began to explore the problem space, they came up with a large number of potential visualization concepts. The relatively large group size allowed us to cover a wide range of visualization ideas. However, this came at the cost of significant overhead. The artistic expertise of course participants varied and critique discussions had to include basic introductory sessions to bring everyone closer the same level of understanding. The process of evaluating and critiquing all sketches also required a considerable time commitment by our scientific collaborators.

While each student created multiple sketches for each stage of the project, only a select few sketching results were unique visualization ideas. In many cases there was a significant overlap between student designs during paper and VR sketching stages. However, thanks to the critique sessions, students were able to coordinate their works and focus on their own distinct visual concepts during the VR prototype stages.

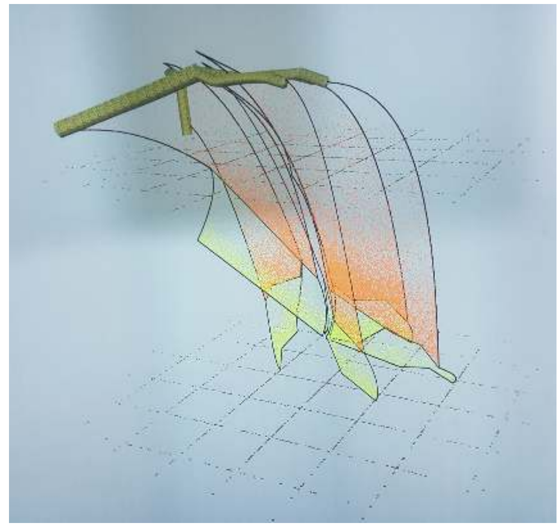
At the end of each course we collected feedback questionnaires from our students, with positive results in both cases. Students commented that the course helped to promote ”creativity and understanding of virtual reality science visualizations” and taught ”the importance of legibility and how to creatively display data”.

6.4.1 Influence of Scientific Introduction

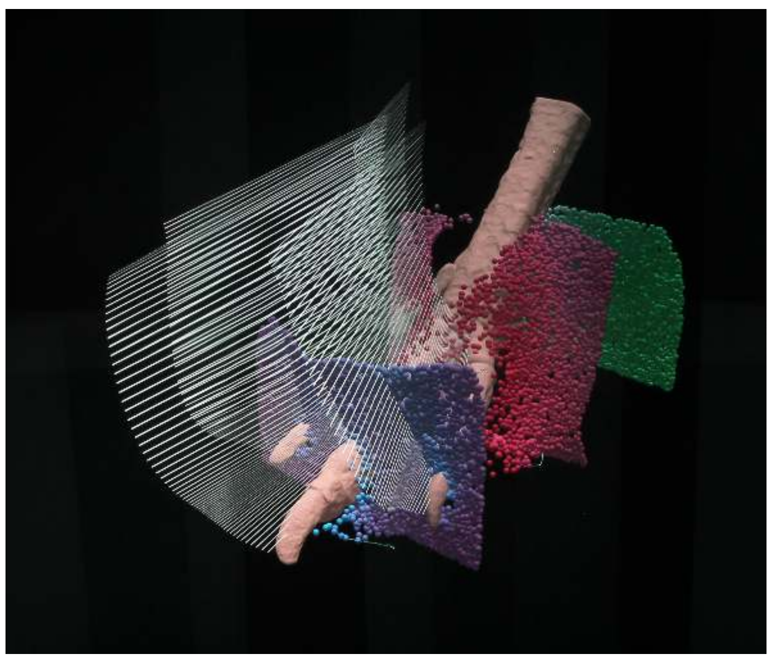
We observed that the initial presentation of the scientific problem had considerable influence on the outcomes of the sketching stages. In particular, we found differences in the variety of visual designs created by students between the two iterations of the design course. In the 2015 class, students explored a wider range of visualization ideas than in the 2017 one. This difference could have been caused by the increased expertise of the science collaborators. In 2015 the research questions presented to students were strongly exploratory, without specific expectations about the visualization outcome. In the 2017 class students were presented with more specific research ques-



(a) Paper sketch/model



(b) VR Sketch



(c) Implementation

Figure 6.12: A physical model (a) and VR sketch (b) of a concept to visualize the foot motion with pathsurfaces of the central axes of individual bones. An example of the implemented visualization showing the pathsurfaces trailing the foot motion trailing throughout simulation timesteps (c).

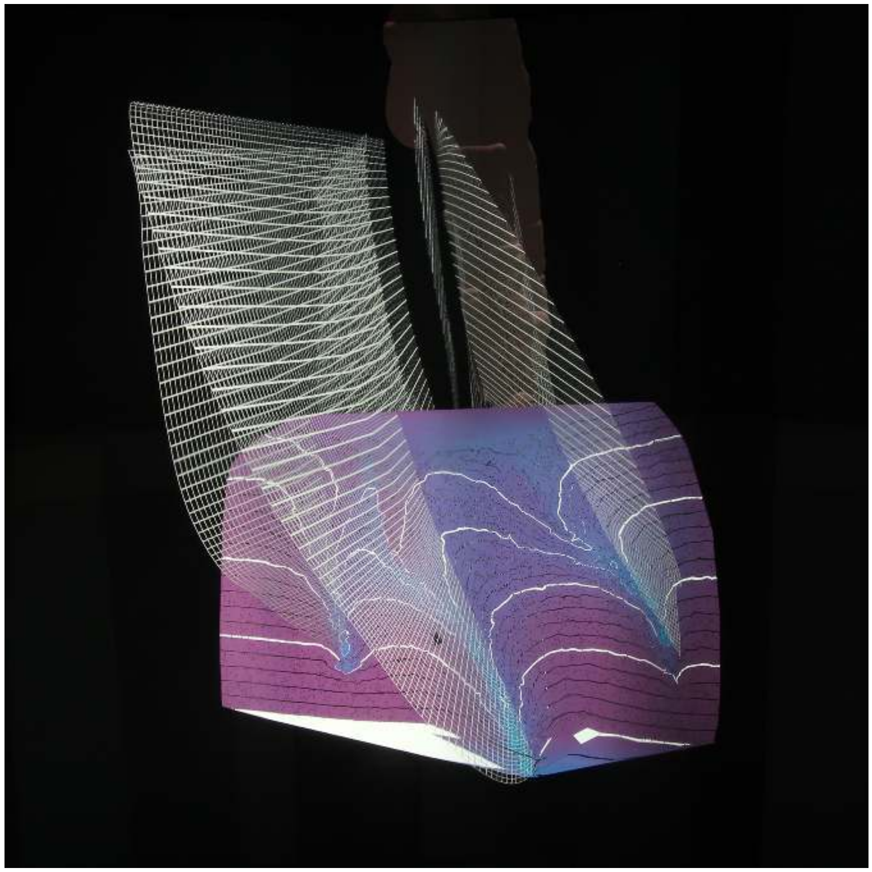


Figure 6.13: Combining foot motion pathsurfaces with vertical time surface visualizations highlights the offset between the actual motion of the foot anatomy and the trails it leaves behind. This discrepancy was one of the main take-aways of our science collaborators.

tions and improved examples of the data and existing visualizations. It also contained a second dinosaur track related assignment, providing students with deeper background knowledge. Having additional information might have caused sketches and prototypes to be visually closer to the already implemented VR visualization concepts. While the second course did not result in drastically different visual metaphors, it still provided compelling ideas for improvements to our ongoing implementation stage.

6.4.2 Benefits of VR visualization

Our paleontologist collaborators confirmed our hypothesis that VR visualization offers significant benefits compared to their standard desktop-based analysis tools. The main advantages reported over the course of the entire project were the following:

The large interaction space of VR environments was beneficial to data analysis. Our VR system surrounds users almost entirely with screen space, as opposed to a single stationary desktop-screen. This allows our visualization to show data at a larger than usual scale. Users can inspect specific regions of the visualization simply by stepping closer, which effectively zooms into the visualization. While focused on specific details users can still connect them to the context of the entire dataset by looking around in the immersive environment. However, this context was easily lost when users were completely surrounded by the visualization. In the YURT, and CAVE-like displays in general, this also distorts the visualization for viewers without head-tracked stereo glasses. We found that our collaborators mainly used the available screen and interaction space to physically walk around the entire dataset. As a response we scaled the visualization to comfortable allow this in our YURT display room.

The depth cues provided by stereo imagery and motion parallax greatly improved our collaborators understanding of their volumetric datasets. Interacting with their data through an intuitive 3D wand interface gave the visualization a feeling of physicality. These factors combined caused our collaborators to feel more present within the virtual environment, a feature that has been linked to increased visualization effectiveness [Slater, 2003]. Many of our implemented visualization techniques, would not work as effectively on a 2D display due to the spatial complexity of our visual metaphors (e.g. initial and intermediate stages of the vertical time surfaces, see Fig. 6.10).

The paleontologists also commented positively on the collaboration aspect in our YURT display room, which can host multiple people at the same time. Being surrounded by data visualizations minimized distractions and stimulated discussions. While only one person can *drive* the visualization at a given time, other observers were still able to follow the exploration process of the driver. This collaborative exploration sometimes discovered new perspectives by serendipity and generated valuable intuition about the structure of the datasets. Our successful visualization project underlines the benefits of high-fidelity VR displays and immersive visualizations for the exploration of spatially complex scientific datasets. Insight gained through data explorations with our VR application has inspired two publications by our collaborators [Turner et al., 2020, Falkingham et al., 2020]. The capabilities of our tool have been demonstrated to other paleontologists with an HMD live-demo at the International Congress of Vertebrate Morphology conference 2019 (ICVM '19) in Prague, Czechia.

6.4.3 Using Scientific Sketching

After observing Scientific Sketching in two separate design courses we observed several factors that could influence the effectiveness of this methodology:

- Scientific Background Information

  As previously stated, we found some evidence that the initial introduction of the scientific problem can have significant impact on the expected sketching outcomes. We suggest to consciously control the amount of detail given at the introduction phase to steer the creative process of the artists and influence the variety of expected visualization outcomes. Broadly defined research questions, for example, give artists a wider design space to work with, but might result in visualization ideas that do not directly address the scientists' problems.

- Group Size

  There are potential diminishing returns on the artist group size. Our courses had 13 and 12 student participants, respectively. In both cases we found a gathered a larger number of overlapping visualization ideas. This suggests that smaller groups could still work effectively while reducing organization overhead.

  We also found that using a single CAVE system for VR sketching leads to scheduling challenges at our group sizes. We overcame these problems by allowing students to work in groups of two. However, in order to ensure the accessibility of the drawing environment we suggest to use more VR devices, or appropriately sized groups.

- Artistic Capabilities

  Our course groups were made up of artists and computer scientists at varying skill levels. Bringing the group to a basic common level of artistic capabilities required additional teaching effort. While it might be more efficient to use a small group of well-trained illustration artists, the mixed group had some distinct benefits. Learning the artists' terminology and critique style gave computer scientists the ability to articulate their questions and describe their sketches in the most constructive way. The teaching aspect of beginner critique sessions also familiarized our domain scientists and software experts to the artistic process and Scientific Sketching.

- Scientist-Artist moderation

  Keeping the dialogue between scientists and artists going throughout all stages of the design process is critical to its success. Scientists should be available to answer student questions even outside of critique sessions, to help them better understand the scientific problems and guide their visualization designs in relevant directions.

- VR Sketching Environment

  After completion of the course, several students commented that they would have preferred additional drawing features in the CavePainting VR environment, including the capability to load 3D models and animations from desktop applications. Giving the art students access to their usual design tools would most likely allow them to realize some of their designs in more detail and in shorter time. However, this would reduce their exposure to the VR environment and in turn their understanding of the specific interaction requirements for immersive visualizations.

- Importance of the Implementation stage
  While the original methodology paper of Scientific Sketching [Keefe et al., 2008] mainly focuses on the initial sketching and prototyping stages, we want to stress the importance of the transition from VR prototype to implementation. We suggest that the group of visualization experts extends their role during the VR prototype phase to provide warnings about potential implementation issues with problematic visualization concepts.

6.4.4 Limitations and Open Questions

The effectiveness of Scientific Sketching depends greatly on the provided data and the research questions to be answered. Since the methodology's goal is the inception and creation of novel immersive visualization techniques, it might fail to produce adequate solutions if the research problem does not benefit from the extended visual space provided by VR environments. A thorough study of existing techniques and desktop visualization applications should be performed before starting the project to ensure useful outcomes.

However, even if a non-immersive solution seems more appropriate for a project, Scientific Sketching could still be used to guide the design process. This would only require the replacement of the VR Sketch and VR Prototype stages with equivalent stages in 2D drawing applications. An evaluation of this adapted use of Scientific Sketching, would be an excellent target for future investigations.

Another research opportunity would be to evaluate whether immersive visualizations can be ”back-ported” into 2D applications. While our collaborators underline that the complex shapes of pathline clusters would be difficult to evaluate in non-immersive displays, aggregate visualization techniques like the (reverse) time surface could work with only minor losses in visual clarity in a desktop application.

6.5 Conclusion

We presented results of an iterative VR visualization development process to analyze unsteady flow in dinosaur track creation datasets of our paleontologist collaborators. The visualization concepts implemented as part of this work have become effective tools for our collaborators and continue to be the base for an ongoing research partnership. The process of developing these tools with the Scientific Sketching methodology was overall successful.

Iterative sketching and prototyping stages allowed us to examine a wide range of visualization ideas and choose the most viable ones for implementation. Ultimately, we created VR tools that help paleontologists to analyze their data through particle, pathline and time surface visualizations. By actively participating and guiding the sketching processes, our collaborators gained new insights into the intricacies of their simulated flow data and drew new connections between dinosaur foot movement and substrate deformation.

Assessing the Scientific Sketching method itself, we found that it was very effective in our scenario. Over the course of our project all three involved groups, artists, visualization experts and paleontologists gained new insights about visualization design and the effective use of VR as a medium for scientific exploration. Based on our results, we can highly recommend Scientific Sketching as design methodology for VR visualizations.

Chapter 7

Discussion and Conclusions

The goal of this dissertation was to generate insights benefiting the creation of effective and intuitively usable immersive visualizations. Over various chapters, we explored fidelity aspects of VR environments, how they influence user behavior in scientific date exploration tasks, and potential ways to design effective immersive visualization applications. Our collected outcomes cover newfound aspects of human perception and interaction paired with practical advice for the development of VR visualization applications. This chapter summarizes the insights derived from the performed experiments and discusses limitations as well as open questions for potential future research.

7.1 Summary of Contributions

Each of the four studies presented in this work provided key insights for effective VR application design, outlined in the following sections. The first two projects addressed basic research questions about human perception in varying VR hardware environments, while the later focused on practical evaluations of VR from a user and developer perspective.

7.1.1 Evaluating Field of Regard and Stereoscopy in Immersive Data Exploration

In our initial experiment we evaluated how VR display fidelity components affect user performance in immersive data exploration tasks on biological datasets (Chapter 3). We found that higher fidelity-levels, in particular the use of stereo displays, can improve task completion times in specific exploration scenarios of spatially complex visualizations. However, these benefits were not present in all tasks. We argue that the interaction strategies utilized by study participants, in particular their walking behavior, acted as a mediating factor in lower-fidelity conditions. This suggests that the use of environments that invite users to explore data by physically moving around it might mitigate shortcomings in display-fidelity and improve the effectiveness of VR applications. While physical motion is an intuitive way of interacting with VR visualizations, it may slow down the exploration process as a whole. This indicates that user interaction methods and exploration strategies should be considered as controlled variables in display fidelity experiments to ensure deeper insights into human perception.

Our outcomes have already inspired follow-up research by our collaborators at Virginia Tech, who analyzed differences in data exploration performance between walking and controller-based viewpoint selection [Lages and Bowman, 2018]. These extended findings show that the spatial cognition capabilities of users and their experience with non-walking interfaces like 3D games influence their exploration performance. Here, the intuitive walking controls improved task completion times of users without gaming background, while experienced users were able to reach shorter performance times using the efficient, but more mentally demanding wand interaction.

Our own follow-up research on text perception also benefited from our findings on the influence of user motion. As a result we incorporated restrictions on user motion into the experimental design, to avoid it as a confounding factor.

7.1.2 Text Perception in Immersive Environments

Our text perception experiment evaluated the readability of text panels under various viewing conditions in immersive particle visualizations. Our results reaffirmed the effect of text-resolution and angular size on reading speeds across multiple VR displays, leading to a recommendation of a minimum angular text size of approximately 30 arcmin (LogMAR 0.8) to accommodate for the resolution limitations of current HMD hardware (e.g., HTC Vive 2019). Smaller text sizes are feasible in higher-fidelity displays, opening up a wider design space in future applications. Yet even in the YURT, which operates close to the resolution-limit of the human eye, we observed reading speeds slower than reference speeds for physical text documents. This indicates that there are additional VR fidelity components influencing text perception at a lower level.

We also showed that several display concepts of 2D tooltip panels can be transferred from 2D desktop applications into immersive VR visualizations. In particular, removing occlusions in front of 3D text panels was very effective in increasing the text readability. In difficult reading conditions with high numbers of occluding objects, displays with higher visual fidelity offer improved reading speed, even without removing occlusions in front of the text panel. We could not find significant advantages of orienting tooltips to always face the user within our experimental framework (see Section 7.2.1). However, this last outcome was potentially influenced by the restrictions on user movement in the experiment and might have a meaningful impact if the application allows for changes of the viewpoint.

Overall we provided a list of guidelines regarding text size selection based on the display fidelity properties of VR devices and visualization-specific text occlusion factors.

7.1.3 Feature Perception in Immersive Medical Data Exploration

Our case study of immersive medical data exploration evaluated the feasibility of using VR MRI visualizations in the diagnosis of vessel-related diseases. Our results show that medical professionals were able to accurately identify minute diagnostically-relevant vessel features thanks to their increased scale and intuitive zooming interaction in the virtual data representation. Despite a lack of formal training in MRI data analysis, our study participants identified a majority of blood vessels marked by a vascular disease expert, highlighting the effectiveness of immersive data visualization for this exploration scenario. Study participants reported difficulties with the task of marking vessel regions at the correct 3D depth within the VR environment, underlining the difficulty of 3D tracing based on visual cues without haptic touch feedback and the need for interaction methods that support users in this task.

Interview feedback from study participants showed that the annotations generated in our experimental system can be helpful in analyzing and discussing individual vascular structures. Our application is a step towards a surgical planning VR environment for TTTS intervention. However, to become an integral part of the clinical workflow, VR technology needs to become more accessible to medical experts. This includes the available of immersive displays at medical facilities and interaction methods that are more intuitive to use for medical data analysis and procedure planning.

7.1.4 Methodology for VR Application Design

Alongside the other experiments, we evaluated the *Scientific Sketching* design methodology in a multi-year study, developing an iterative VR visualization to analyze unsteady flow in dinosaur track creation datasets. The iterative sketching and prototyping stages allowed us to examine a wide range of visualization ideas and choose the most viable ones for implementation. Ultimately, we created VR tools that help paleontologists to analyze their data through particle, pathline and time surface visualizations.

By actively participating and guiding the sketching processes, our scientific collaborators gained new insights into the intricacies of their simulated flow data and drew new connections between dinosaur foot movement and substrate deformation. Over the course of our project the involved artists, visualization experts, and paleontologists gained new insights about visualization design and the effective use of VR as a medium for scientific exploration. The results underline the effectiveness of

Scientific Sketching as a design methodology for VR visualizations. In particular, we provide practical recommendations on group sizes, actor composition, initial presentation of the visualization task and VR hardware selection.

7.2 Discussion and Open Research Questions

In our experiments we evaluated general visualization and interaction concepts and studied detailed domain-specific use cases. As a result, we generated insights for the field of perception research and the practical use of VR for immersive visualizations. However, each project also opened up compelling research questions, leaving several avenues for future work.

7.2.1 Utilizing VR Display Resolution

The current popularity of VR devices has led to rapid advancements in immersive display technology creating two diverging challenges for immersive perception studies. One is to design experiments that can provide insights applicable to future device generations, the other is to ensure a certain level of backward-compatibility to existing devices in order to value for the increasing fidelity spread of VR displays. Over the course of this thesis we have upgraded our experimental equipment multiple times to keep up with technological advances. HMDs were a particular focus of this, due to their high availability and their potential for hardware improvements.

Based on the almost yearly HMD releases, it is obvious that hardware providers focus on increasing the display resolution with every new generation. Our experimental results on text and feature detection suggest that this development will increase the effectiveness of new devices for scientific applications. However, our work focused on using current devices at their highest resolution settings, which caused a wide gap in fidelity between evaluated HMDs and our high-resolution YURT system. While HMDs may at some point reach retina-level resolution, investigating intermediate resolution steps at potential resolution levels provided by future HMD generations could provide valuable insights for optimizing viewing parameters in immersive visualizations. Even in retina-resolution VR displays, some viewing Parameters that could potentially benefit from a more detailed evaluation include the scale of geometric primitives and text representation used within VR visualization applications. Due to its orthogonal condition design, our experiment on tooltip reading evaluated only two density setting for occluding objects. Further study of this topic could lead to more insights into the interaction between occluder geometry, size, and density and embedded text perception.

In a similar way, further analysis of occlusion reduction is needed to evaluate techniques that preserve the context of given host-visualizations, e.g., through the use of object transparency. Here the additional depth cues provided by VR displays have the potential help distinguish between semi-transparent 3D objects and maintain text readability, without a loss of data clarity.

Finally, the orientation property of text embedded within immersive visualizations warrants further investigation. Due to the limitations of our study design we did not evaluate viewing angles with a large offset from the orthogonal view. While one can expect an increased reading difficulty in more slanted views, it remains an open question if panels that rotate towards the user, and thereby potentially intersect with other virtual primitives of a scene, have effects on their sense of immersion within the VR environment. Evaluating the trade-off between readability and immersion could lead to distinct guidelines to build effective immersive applications.

7.2.2 Developing Interactions in Immersive Visualizations

The development process of our particle visualization application with *Scientific Sketching* led to novel immersive visualization techniques that greatly contributed to the data exploration work of our collaborators. However, creating immersive mockup drawings of new interaction techniques was one of the most challenging aspects for our artist collaborators. This was in part caused by the use of our VR drawing application, CavePainting, which lacked sophisticated animation features.

With the emergence of more powerful immersive drawing applications and design tools, it would be of value to revisit *Scientific Sketching* and evaluate its effectiveness with respect to the development of new interaction techniques. An example would be the interaction with time surfaces as shown in Figure 7.1. The tight and sometimes overlapping placement of 3D pathlines can make it difficult to select specific lines with tracked wand devices. Developing new interaction techniques to aid a user's selection accuracy would be a relevant evaluation scenario for upcoming VR application design studies.

7.2.3 Improving Particle Flow Data Exploration

Our design study on immersive particle flow visualization indicated advantages of pathline clustering to derive new insights into fluid motion. The clustering methods used in our prototype were based on distance functions on well-established line shape and orientation metrics. While this provided our collaborators with a valuable new data analysis tool, the use of immersive environments opens up new research directions to improve clustering results.

The 3D selection capabilities available in immersive environments, especially spatial brushing techniques, could be employed to give users the ability to directly improve clustering outcomes by intuitively marking pathline bundles for inclusion or outliers for exclusion. Using these user selec-

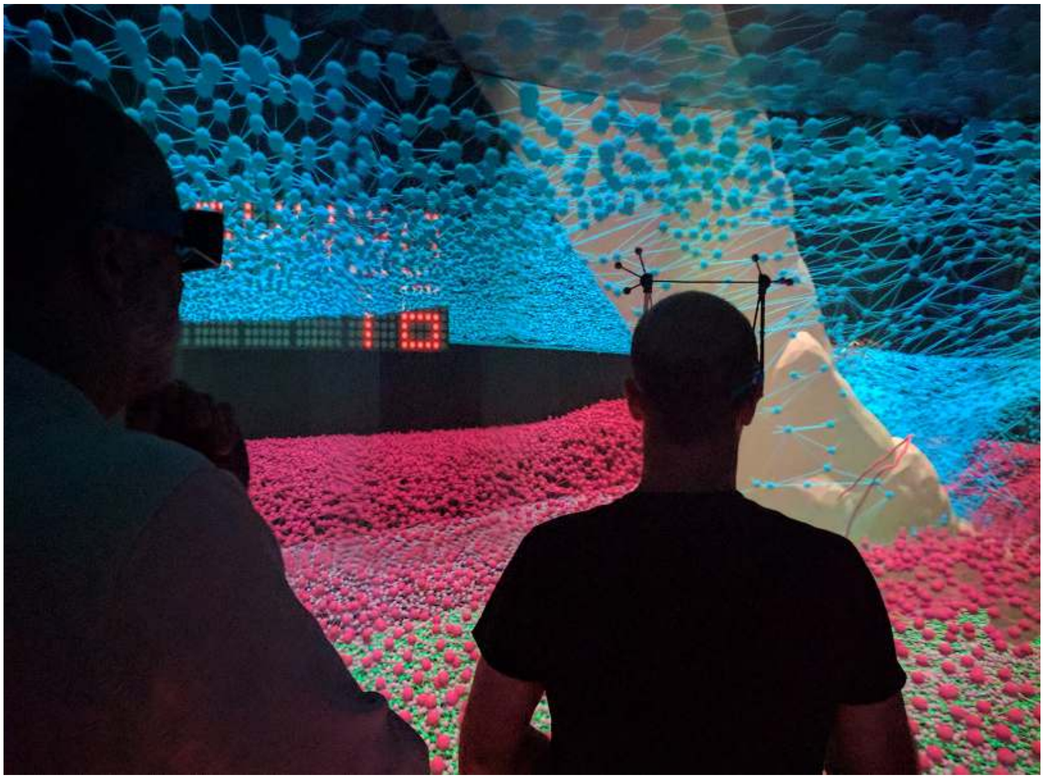


Figure 7.1: A user interacting with an particle immersive particle visualization to analyze time surfaces.

tions as additional inputs in Machine Learning methods like the FlowNet Deep Learning Framework presented by Han et al. [2020], could lead to novel ways of exploring particle movement in spatially dense datasets.

7.3 The Future of VR

Over the multiple years it took to complete the projects presented in this thesis VR technology saw rapid development. Devices used in our first perception experiment would be considered out-dated by the of the submission of this work. In this section I would like to summarize personal insights in the development for and use of VR displays, collected over the course of my research.

One of the main problems encountered when starting inter-disciplinary collaboration projects for VR visualizations was the perceived entry cost for domain scientists. The development time of domain-specific VR software to a level matching that of then existing desktop solutions and the price of VR hardware were a prohibiting factor. But with every year, developing applications for VR displays has been becoming easier, especially when considering head-worn devices. As more

HMDs found their way into the homes and labs of users, extending the APIs of graphics frameworks (like Unity3D, Blender and Unreal Engine) and visualization toolkits (like VTK and ParaView) to support VR displays became a priority. Consequently, the bar to develop novel visualization methods tailored to the capabilities of immersive displays has been lowered significantly.

It seems like this trend will continue to a point where domain-specific applications will provide VR support directly. This will be the milestone after which VR visualization techniques could see significant improvements. As shown in our work on *Scientific Sketching*, domain scientists are in some cases unaware of the possibilities provided by VR displays. Once they are provided with means to freely explore the design space of immersive visualizations on their own terms, they may be able to formulate problems with existing visualization methods and collaborate with computer scientists to find novel solutions. The ease of entry makes HMDs the prime choice for single user VR applications.

CAVE displays have not benefited as strongly from API improvements, due to the focus on HMD devices. The relative rarity of CAVE setups combined with the fact that most of them use unique combinations of projectors and screens make them a difficult target to develop for. However, even here new frameworks, like MinVR, have reduced the entry requirements for developers.

Despite the success of HMD devices, I would disagree that they make CAVE systems obsolete at this point in time. As the YURT system at Brown University exemplifies, detaching the physical screens from the headset removes limitations on their size and allows for retina resolution rendering across the entire screen area. At the same time this change greatly reduces the weight of the headset, increasing the wearing comfort. Additionally, the ability to view your own body and those of other users within a CAVE display is a feature that improved the effectiveness of collaborative data exploration within our own experiments. In HMDs, mimicking this personal interaction with virtual avatars is still an active research topic and it will take more time until the fidelity of these avatars reaches a realistic representation. The lighter and easy to handle stereo glasses typically used in CAVEs are also a benefit when demonstrating VR environments to multiple groups, as they can be changed in quick succession. If a VR systems aim is focused on teaching scenarios or if local collaboration of up to three users is a concern, CAVEs are still an architecture that should be considered.

Despite their advantages, HMD and CAVE VR displays incur a “cost” on the user that might limit their inclusion into small scale data exploration workflows. Switching from actual to virtual reality takes time and once inside the immersive environment users often lose access to tools and information available in their real world workplace. Until a majority of the working environment can be adequately represented within the environment, users will frequently have to swap out of VR for tasks like reference reading and note taking.

The devices that bridge this gap will hopefully be future AR displays. Blending virtual objects and scenes into a users perception of the real world allows the seamless use of immersive visualizations in general research workflows, even for short exploration tasks. Since users don't lose access to any of their other tools, developers and scientists can focus on directly creating visualizations, without having to replicate entire work environments for the user. Instead, AR visualizations could be integrated into desktop software as immersive window that can be accessed at a moments notice. AR devices have yet to reach the Field of View and resolution of current VR HMDs like the HTC Vive Pro, yet I am convinced that they will see their use scientific visualization once they catch up.

7.4 General Conclusions

This research thesis investigated effective ways of using VR visualizations in scientific data exploration and their specific benefits for perception and interaction. In a set of multiple perception experiments and case studies we have identified effective ways of using immersive visualizations to aid domain scientists. Our results show that VR tools allowed them to view and analyze their data in ways currently not provided by 2D displays, by providing them with intuitive movement and interaction methods as well as screen-space encompassing almost their entire field of vision. The results of our experiments advance our understanding of effective data representation in VR displays in relation to human perception.

As VR displays become more prevalent in the future, opportunities for novel scientific visualization techniques and interaction methods will open up. Yet to make full use of all aspects provided by immersive displays, further studies of human perception is required. We believe that the results presented in this thesis will form a solid basis for these future experiments, and will provide immediate insights for software and hardware developers to build effective VR applications pushing domain and visualization research to greater heights.

Bibliography

Daniel Acevedo-Feliz, Bret Jackson, Daniel Keefe, Ryan Pavlik, Jurgen Schulze, Bill Sherman, and Alexis Vartanian. VR Toolkits: Why do we keep reinventing the wheel? In *Proceedings of the IEEE Virtual Reality Conference 2014 (VR '14) Panels*, pages 1–3, 3 2014. doi: 10.1109/VR.2014.6802114.

Amir Alansary, Konstantinos Kamnitsas, Alice Davidson, Rostislav Khlebnikov, Martin Rajchl, Christina Malamateniou, Mary Rutherford, Joseph V. Hajnal, Ben Glocker, Daniel Rueckert, and Bernhard Kainz. Fast Fully Automatic Segmentation of the Human Placenta from Motion Corrupted MRI. In Sebastien Ourselin, Leo Joskowicz, Mert R Sabuncu, Gozde Unal, and William Wells, editors, *MICCAI 2016: Medical Image Computing and Computer-Assisted Intervention*, pages 589–597. Springer International Publishing, 2016. ISBN 978-3-319-46723-8. doi: 10.1007/978-3-319-46723-8{_}68. URL http://link.springer.com/10.1007/978-3-319-46723-8_68.

Kamran Ali, Knut Hartmann, and Thomas Strothotte. Label Layout for Interactive 3D Illustrations. *Journal of the WSCG*, 13(1):1–8, 2005. URL http://wscg.zcu.cz/wscg2005/Papers_2005/Journal/L07-full.pdf.

John R. L. Allen. Short Paper: Fossil vertebrate tracks and indenter mechanics. *Journal of the Geological Society*, 146(4):600–602, 7 1989. ISSN 0016-7649. doi: 10.1144/gsjgs.146.4.0600. URL http://jgs.lyellcollection.org/lookup/doi/10.1144/gsjgs.146.4.0600.

Autodesk. Maya Computer Animation and Modeling Software [Accessed: 2018-03-27], 2018. URL https://www.autodesk.com/products/maya/overview.

Blaine Bell, Steven Feiner, and Tobias Höllerer. View Management for Virtual and Augmented Reality. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology*, UIST 01, page 101110, New York, NY, USA, 2001. Association for Computing

Machinery. ISBN 158113438X. doi: 10.1145/502348.502363. URL https://doi.org/10.1145/502348.502363.

Doug A. Bowman and Ryan P. McMahan. Virtual Reality: How Much Immersion Is Enough? *Computer*, 40(7):36–43, 2007. ISSN 0018-9162. doi: 10.1109/MC.2007.257. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4287241.

Doug A. Bowman and Dheva Raja. A method for quantifying the benefits of immersion using the cave. *Presence-Connect*, 4(4), 2004.

Doug A. Bowman, Chris North, Jian Chen, Nicholas F. Polys, Pardha S. Pyla, and Umur Yilmaz. Information-Rich Virtual Environments: Theory, Tools, and Research Agenda. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST 03, page 8190, New York, NY, USA, 2003. Association for Computing Machinery. ISBN 1581135696. doi: 10.1145/1008653.1008669. URL https://doi.org/10.1145/1008653.1008669.

Doug A. Bowman, Cheryl Stinson, Eric D. Ragan, Siroberto Scerbo, Tobias Höllerer, Cha Lee, Ryan P. McMahan, and Rgis Kopper. Evaluating effectiveness in virtual environments with MR simulation. In *Interservice/Industry Training, Simulation, and Education Conference*, Orlando, FL, 2012. National Training and Simulation Association, National Training and Simulation Association. URL http://ntsa.metapress.com/link.asp?id=9h61028206h3503w.

Elizabeth L. Brainerd, David B. Baier, Stephen M. Gatesy, Tyson L. Hedrick, Keith A. Metzger, Susannah L. Gilbert, and Joseph J. Crisco. X-ray reconstruction of moving morphology (XROMM): precision, accuracy and applications in comparative biomechanics research. *Journal of experimental zoology. Part A, Ecological genetics and physiology*, 313(5):262–79, 6 2010. ISSN 1932-5231. doi: 10.1002/jez.589. URL http://doi.wiley.com/10.1002/jez.589.

Andrea Brambilla, Robert Carnecky, Ronald Peikert, Ivan Viola, and Helwig Hauser. Illustrative Flow Visualization: State of the Art, Trends and Challenges. In Marie-Paule Cani and Fabio Ganovelli, editors, *Eurographics 2012 - State of the Art Reports*. The Eurographics Association, 2012. doi: 10.2312/conf/EG2012/stars/075-094.

Frederick P. Brooks. What's real about virtual reality? *IEEE Computer Graphics and Applications*, 19(6):16–27, 1999. ISSN 02721716. doi: 10.1109/38.799723.

Steve Bryson. Virtual reality in scientific visualization. *Communications of the ACM*, 39(5):62–71, 1996. ISSN 00010782. doi: 10.1145/229459.229467.

Bill Buxton. *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2007. ISBN 0123740371, 9780123740373.

Gwnal Caravaca, Stphane Le Mouélic, Nicolas Mangold, Jonas LHaridon, Laetitia Le Deit, and Marion Massé. 3D digital outcrop model reconstruction of the Kimberley outcrop (Gale crater, Mars) and its integration into Virtual Reality for simulated geological analysis. *Planetary and Space Science*, 182:104808, 2020. ISSN 0032-0633. doi: https://doi.org/10.1016/j.pss.2019.104808. URL `https://www.sciencedirect.com/science/article/pii/S0032063319302600`.

Jian Chen, Haipeng Cai, Alexander P. Auchus, and David H. Laidlaw. Effects of stereo and screen size on the legibility of three-dimensional streamtube visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2130–2139, 2012. ISSN 10772626. doi: 10.1109/TVCG.2012.216.

Norman Cliff. *Ordinal Methods for Behavioral Data Analysis*. Psychology Press, 3 2014. ISBN 9781315806730. doi: 10.4324/9781315806730. URL `https://www.taylorfrancis.com/books/9781315806730`.

Carolina Cruz-Neira, Daniel J. Sandin, and Thomas A. DeFanti. Surround-screen projection-based virtual reality. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques (SIGGRAPH '93)*, pages 135–142, New York, New York, USA, 1993. ACM Press. ISBN 0897916018. doi: 10.1145/166117.166134. URL `http://portal.acm.org/citation.cfm?doid=166117.166134`.

James J. Cummings and Jeremy N. Bailenson. How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology*, 19(2):272–309, 2016. doi: 10.1080/15213269.2015.1015740. URL `https://doi.org/10.1080/15213269.2015.1015740`.

Saverio Debernardis, Michele Fiorentino, Michele Gattullo, Giuseppe Monno, and Antonio Emmanuele Uva. Text Readability in Head-Worn Displays: Color and Style Optimization in Video versus Optical See-Through Devices. *IEEE Transactions on Visualization and Computer Graphics*, 20(1):125–139, 1 2014. ISSN 1077-2626. doi: 10.1109/TVCG.2013.86. URL `http://ieeexplore.ieee.org/document/6520861/`.

Caatay Demiralp, Cullen D. Jackson, David B. Karelitz, Song Zhang, and David H. Laidlaw. Cave and fishtank virtual-reality displays: A qualitative and quantitative comparison. *IEEE Transac-*

tions on Visualization and Computer Graphics, 12(3):323–330, 2006. ISSN 1077-2626. doi: 10.1109/TVCG.2006.42. URL http://www.ncbi.nlm.nih.gov/pubmed/16640246.

Tilman Dingler, Kai Kunze, and Benjamin Outram. VR Reading UIs. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, pages 1–6, New York, NY, USA, 4 2018. ACM. ISBN 9781450356213. doi: 10.1145/3170427.3188695. URL http://doi.acm.org/10.1145/3170427.3188695.

Elisabeth Dittrich, Stefan Brandenburg, and Boris Beckmann-Dobrev. Legibility of Letters in Reality, 2D and 3D Projection. In Randall Shumaker, editor, *Virtual Augmented and Mixed Reality. Designing and Developing Augmented and Virtual Environments*, pages 149–158. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013. ISBN 978-3-642-39405-8. doi: 10.1007/978-3-642-39405-8{_}18. URL http://link.springer.com/10.1007/978-3-642-39405-8_18.

Jonathan Dobres, Benjamin Wolfe, Nadine Chahine, and Bryan Reimer. The effects of visual crowding, text size, and positional uncertainty on text legibility at a glance. *Applied Ergonomics*, 70:240–246, 2018. ISSN 0003-6870. doi: https://doi.org/10.1016/j.apergo.2018.03.007. URL http://www.sciencedirect.com/science/article/pii/S0003687018300681.

Ciro Donalek, S. George Djorgovski, Alex Cioc, Anwell Wang, Jerry Zhang, Elizabeth Lawler, Stacy Yeh, Ashish Mahabal, Matthew Graham, Andrew Drake, Scott Davidoff, Jeffrey S. Norris, and Giuseppe Longo. Immersive and Collaborative Data Visualization Using Virtual Reality Platforms. In *Proceedings of the IEEE International Conference on Big Data 2014*, pages 609–614, 10 2014. ISBN 9781479956661. doi: 10.1109/BigData.2014.7004282.

Ruth B. Ekstrom, John W. French, and Harry H. Harman. Cognitive factors: Their identification and replication. *Multivariate Behavioral Research Monographs*, 79(2):84, 1979.

David B. Elliott. The good (logMAR), the bad (Snellen) and the ugly (BCVA, number of letters read) of visual acuity measurement. *Ophthalmic and Physiological Optics*, 36(4):355–358, 7 2016. ISSN 02755408. doi: 10.1111/opo.12310. URL http://doi.wiley.com/10.1111/opo.12310.

Khaled M. Elsayes, Andrew T. Trout, Aaron M. Friedkin, Peter S. Liu, Ronald O. Bude, Joel F. Platt, and Christine O. Menias. Imaging of the Placenta: A Multimodality Pictorial Review. *RadioGraphics*, 29(5):1371–1391, 9 2009. ISSN 0271-5333. doi: 10.1148/rg.295085242. URL http://pubs.rsna.org/doi/10.1148/rg.295085242.

Peter L. Falkingham and Stephen M. Gatesy. The birth of a dinosaur footprint: Subsurface 3D motion reconstruction and discrete element simulation reveal track ontogeny. *Proceedings of the National Academy of Sciences*, 111(51):18279–18284, 12 2014. ISSN 0027-8424. doi: 10.1073/pnas.1416252111. URL `http://www.ncbi.nlm.nih.gov/pubmed/25489092`.

Peter L. Falkingham, Morgan L. Turner, and Stephen M. Gatesy. Constructing and testing hypotheses of dinosaur foot motions from fossil tracks using digitization and simulation. *Palaeontology*, 63(6):865–880, 2020. doi: https://doi.org/10.1111/pala.12502. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/pala.12502`.

Alessandro Febretti, Arthur Nishimoto, Terrance Thigpen, Jonas Talandis, Lance Long, J. D. Pirtle, Tom Peterka, Alan Verlo, Maxine D. Brown, Dana Plepys, Dan Sandin, Luc Renambot, Andrew Johnson, and Jason Leigh. CAVE2: a hybrid reality environment for immersive simulation and information analysis. In Margaret Dolinsky and Ian E. McDowall, editors, *Proceedings of SPIE - The Engineering Reality of Virtual Reality 2013*, volume 8649, pages 9 – 20, 3 2013. ISBN 9781479928712. doi: 10.1117/12.2005484. URL `http://dx.doi.org/10.1117/12.2005484http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2005484`.

Florian Ferstl, Kai Burger, and Rdiger Westermann. Streamline Variability Plots for Characterizing the Uncertainty in Vector Field Ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):767–776, 1 2016. ISSN 1077-2626. doi: 10.1109/TVCG.2015.2467204. URL `http://ieeexplore.ieee.org/document/7192675/`.

Andrew S. Forsberg, Michael Katzourin, Kristi Wharton, Mel Slater, and Prabhat. A comparative study of Desktop, fishtank, and cave systems for the exploration of volume rendered confocal data sets. *IEEE Transactions on Visualization and Computer Graphics*, 14(3):551–563, 2008. ISSN 10772626. doi: 10.1109/TVCG.2007.70433. URL `http://www.ncbi.nlm.nih.gov/pubmed/18461742`.

Stephen M. Gatesy, Kevin M. Middleton, Farish A. Jenkins, and Neil H. Shubin. Three-dimensional preservation of foot movements in Triassic theropod dinosaurs. *Nature*, 399(6732):141–144, 5 1999. ISSN 0028-0836. doi: 10.1038/20167. URL `http://www.nature.com/articles/20167`.

Stephen M. Gatesy, David B. Baier, Farish A. Jenkins, and Kenneth P. Dial. Scientific rotoscoping: a morphology-based method of 3-D motion analysis and visualization. *Journal of experimental* *zoology. Part A, Ecological genetics and physiology*, 313(5):244–61, 6 2010. ISSN 1932-5231. doi: 10.1002/jez.588. URL `http://doi.wiley.com/10.1002/jez.588`.

Cameron Grout, William Rogers, Mark Apperley, and Steve Jones. Reading text in an immersive head-mounted display. In *Proceedings of the 15th New Zealand Conference on Human-Computer Interaction (CHINZ '15)*, CHINZ 2015, pages 9–16, New York, NY, USA, 9 2015. ACM. ISBN 9781450336703. doi: 10.1145/2808047.2808055. URL `http://doi.acm.org/10.1145/2808047.2808055`.

Jun Han, Jun Tao, and Chaoli Wang. FlowNet: A Deep Learning Framework for Clustering and Selection of Streamlines and Stream Surfaces. *IEEE Transactions on Visualization and Computer Graphics*, 26(4):1732 – 1744, 2020. ISSN 1077-2626. doi: 10.1109/TVCG.2018.2880207. URL `https://ieeexplore.ieee.org/document/8532319/`.

Claudia Hänel, Benjamin Weyers, Bernd Hentschel, and Torsten W. Kuhlen. Visual Quality Adjustment for Volume Rendering in a Head-Tracked Virtual Environment. *IEEE Transactions on Visualization and Computer Graphics*, 22(4):1472–1481, 4 2016. ISSN 10772626. doi: 10.1109/TVCG.2016.2518338.

Peter Hanula, Kamil Piekutowski, Carlos Uribe, Kyle Almryde, Arthur Nishimoto, Julieta Aguilera, and G. Elisabeta Marai. Cavern Halos: Exploring spatial and nonspatial cosmological data in an immersive virtual environment. In *Proceedings of the IEEE Scientific Visualization Conference 2015 (SciVis '15)*, pages 87–99. IEEE, 10 2015. ISBN 978-1-4673-9785-8. doi: 10.1109/SciVis.2015.7429497. URL `http://ieeexplore.ieee.org/document/7429497/`.

Sandra G. Hart. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):904–908, 10 2006. ISSN 1541-9312. doi: 10.1177/154193120605000909. URL `https://doi.org/10.1177/154193120605000909`.

Anders Helgeland and Thomas Elboth. High-Quality and Interactive Animations of 3D Time-Varying Vector Fields. *IEEE Transactions on Visualization and Computer Graphics*, 12(6):1535–1546, 11 2006. ISSN 1077-2626. doi: 10.1109/TVCG.2006.95. URL `http://ieeexplore.ieee.org/document/1703373/`.

Edward Hitchcock. *Ichnology of New England: A Report on the Sandstone of the Connecticut Valley Especially Its Fossil Footmarks, Made to the Government of the Commonwealth of Massachusetts*. William White, printer, 1858.

David M. Hoffman, Ahna R. Girshick, Kurt Akeley, and Martin S. Banks. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of vision*, 8(3):1–30, 2008. ISSN 1534-7362. doi: 10.1167/8.3.33. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2879326&tool=pmcentrez&rendertype=abstract%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/18484839%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2879326%5Cn/pmc/articles/PMC2879326/?report.

Jagathshree Iyer, Nicholas F. Polys, and Lance Arsenault. Text density and display bandwidth. In *Proceedings of the 22nd International Conference on 3D Web Technology*, Web3D '17, pages 1–7, New York, NY, USA, 6 2017. ACM. ISBN 9781450349550. doi: 10.1145/3055624.3075958. URL http://doi.acm.org/10.1145/3055624.3075958.

Jacek Jankowski, Krystian Samp, Izabela Irzynska, Marek Jozwowicz, and Stefan Decker. Integrating text with video and 3D graphics: The Effects of Text Drawing Styles on Text Readability. In *Proceedings of the 28th international conference on Human factors in computing systems (CHI '10)*, pages 1321–1330, New York, New York, USA, 2010. ACM, ACM Press. ISBN 9781605589299. doi: 10.1145/1753326.1753524. URL http://portal.acm.org/citation.cfm?doid=1753326.1753524.

David Karelitz, Daniel Keefe, and David H. Laidlaw. Using CavePainting to Create Scientific Visualizations. In *Proceedings of the IEEE Visualization Conference 2003, Posters Compendium*, page 109, 10 2003.

Daniel F. Keefe. Official MinVR GIT Repository [Accessed: 2018-03-27], 2018. URL https://github.com/MinVR/MinVR.

Daniel F. Keefe, Robert C. Zeleznik, and David H. Laidlaw. Drawing on Air: Input Techniques for Controlled 3D Line Illustration. *IEEE Transactions on Visualization and Computer Graphics*, 13(5):1067–1081, 9 2007. ISSN 1077-2626. doi: 10.1109/TVCG.2007.1060. URL http://ieeexplore.ieee.org/document/4135646/.

Daniel F. Keefe, Daniel Acevedo, Jadrian Miles, Fritz Drury, Sharon M. Swartz, and David H. Laidlaw. Scientific Sketching for Collaborative VR Visualization Design. *IEEE Transactions on Visualization and Computer Graphics*, 14(4):835–847, 7 2008. ISSN 1077-2626. doi: 10.1109/TVCG.2008.31. URL http://ieeexplore.ieee.org/document/4447665/.

Anne Kenyon, John van Rosendale, Samuel Fulcomer, and David H. Laidlaw. The design of a retinal resolution fully immersive VR display. In *Proceedings of the IEEE Virtual Reality Conference*

2014 (VR '14), pages 89–90. IEEE, 3 2014. ISBN 978-1-4799-2871-2. doi: 10.1109/VR.2014.6802065. URL http://ieeexplore.ieee.org/document/6802065/.

Ron Kikinis, Steve D. Pieper, and Kirby G. Vosburgh. 3D Slicer: A Platform for Subject-Specific Image Analysis, Visualization, and Clinical Support. In Ferenc A Jolesz, editor, *Intraoperative Imaging and Image-Guided Therapy*, pages 277–289. Springer New York, New York, NY, 2014. ISBN 978-1-4614-7657-3. doi: 10.1007/978-1-4614-7657-3{_}19. URL http://link.springer.com/10.1007/978-1-4614-7657-3_19.

Kyungyoon Kim, Bret Jackson, Ioannis Karamouzas, Moses Adeagbo, Stephen J. Guy, Richard Graff, and Daniel F. Keefe. Bema: A Multimodal Interface for Expert Experiential Analysis of Political Assemblies at the Pnyx in Ancient Greece. In *Proceedings of the IEEE Symposium on 3D User Interfaces 2015*, pages 19–26, 2015.

Christoph Kloss and Christoph Goniva. LIGGGHTS - Open Source Discrete Element Simulations of Granular Materials Based on Lammps. In *Supplemental Proceedings*, pages 781–788. John Wiley & Sons, Inc., Hoboken, NJ, USA, 4 2011. doi: 10.1002/9781118062142.ch94. URL http://dx.doi.org/10.1002/9781118062142.ch94.

Oliver Kreylos. Environment-Independent VR Development. In G. et al. Bebis, editor, *Advances in Visual Computing*, pages 901–912. Springer, Berlin, Heidelberg, 2008. doi: 10.1007/978-3-540-89639-5{_}86. URL http://link.springer.com/10.1007/978-3-540-89639-5_86.

Oliver Kreylos. VRUI Applications [Accessed: 2017-12-16], 2017. URL http://idav.ucdavis.edu/~okreylos/ResDev/KeckCAVES/Applications.html.

Hari Krishnan, Christoph Garth, and Kenneth I. Joy. Time and Streak Surfaces for Flow Visualization in Large Time-Varying Data Sets. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1267–1274, 11 2009. ISSN 1077-2626. doi: 10.1109/TVCG.2009.190. URL http://ieeexplore.ieee.org/document/5290738/.

Wallace S. Lages and Doug A. Bowman. Move the Object or Move Myself? Walking vs. Manipulation for the Examination of 3D Scientific Data. *Frontiers in ICT*, 5:15, 2018. ISSN 2297-198X. doi: 10.3389/fict.2018.00015. URL https://www.frontiersin.org/article/10.3389/fict.2018.00015.

Wallace S. Lages, Bireswar Laha, Wesley R. Miller, Johannes Novotny, David H. Laidlaw, John J. Socha, and Doug A. Bowman. Effects of field of regard and stereoscopy and the validity of MR simulation for visual analysis of scientific data. In *Proceedings of the IEEE Virtual Reality Conference 2016 (VR 2016)*, pages 215–216. IEEE, 3 2016. ISBN 978-1-5090-0836-0. doi: 10.1109/VR.2016.7504730. URL `http://ieeexplore.ieee.org/document/7504730/`.

Bireswar Laha and Doug A. Bowman. Identifying the benefits of immersion in virtual reality for volume data visualization. In *Immersive visualization revisited workshop of the IEEE Virtual Reality Conference*, pages 1–2, 2012.

Bireswar Laha, Kriti Sensharma, James D. Schiffbauer, and Doug A. Bowman. Effects of immersion on visual analysis of volume data. *IEEE Transactions on Visualization and Computer Graphics*, 18(4):597–606, 2012.

Bireswar Laha, Doug A. Bowman, and James D. Schiffbauer. Validation of the MR simulation approach for evaluating the effects of immersion on visual analysis of volume data. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):529–38, 2013. ISSN 1941-0506. doi: 10.1109/TVCG.2013.43. URL `http://www.ncbi.nlm.nih.gov/pubmed/23428436`.

Bireswar Laha, Doug A. Bowman, and John J. Socha. Effects of VR system fidelity on analyzing isosurface visualization of volume datasets. *IEEE Transactions on Visualization and Computer Graphics*, 20(4):513–522, 4 2014. ISSN 1941-0506. doi: 10.1109/TVCG.2014.20. URL `http://www.ncbi.nlm.nih.gov/pubmed/24650978`.

Bireswar Laha, Doug A. Bowman, David H. Laidlaw, and John J. Socha. A classification of user tasks in visual analysis of volume data. In *Proceedings of the IEEE Scientific Visualization Conference 2015 (SciVis 2015)*, pages 1–8, 10 2015. doi: 10.1109/SciVis.2015.7429485.

Cha Lee, Gustavo A. Rincon, Greg Meyer, Tobias Höllerer, and Doug A. Bowman. The effects of visual realism on search tasks in mixed reality simulation. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):547–556, 2013. ISSN 10772626. doi: 10.1109/TVCG.2013.41.

Ajey Lele. Virtual reality and its military utility. *Journal of Ambient Intelligence and Humanized Computing*, 4(1):17–26, 2013.

Michele L. Lombardo, Debra J. Watson-Smith, Christopher S. Muratore, Stephen R. Carr, Barbara M. O'Brien, and Franois I. Luks. Laser Ablation of Placental Vessels in Twin-to-Twin Transfusion Syndrome: A Paradigm for Endoscopic Fetal Surgery. *Journal of Laparoendoscopic & Advanced Surgical Techniques*, 21(9):869–872, 11 2011. ISSN 1092-6429. doi: 10.1089/lap.2011.0119. URL `http://www.liebertpub.com/doi/10.1089/lap.2011.0119`.

Franois I. Luks, Stephen R. Carr, Beverly Ponte, Jeffrey M. Rogg, and Thomas F. Tracy. Preoperative planning with magnetic resonance imaging and computerized volume rendering in twin-to-twin transfusion syndrome. *American Journal of Obstetrics and Gynecology*, 185(1):216–219, 7 2001. ISSN 00029378. doi: 10.1067/mob.2001.115111. URL `https://linkinghub.elsevier.com/retrieve/pii/S0002937801911871`.

Franois I. Luks, Stephen R. Carr, Monique E. De Paepe, and Thomas F. Tracy. Whatand whythe pediatric surgeon should know about twin-to-twin transfusion syndrome. *Journal of Pediatric Surgery*, 40(7):1063–1069, 7 2005. ISSN 00223468. doi: 10.1016/j.jpedsurg.2005.03.056. URL `https://linkinghub.elsevier.com/retrieve/pii/S0022346805002411`.

Vito Modesto Manghisi, Michele Gattullo, Michele Fiorentino, Antonio Emmanuele Uva, Francescomaria Marino, Vitoantonio Bevilacqua, and Giuseppe Monno. Predicting Text Legibility over Textured Digital Backgrounds for a Monocular Optical See-Through Display. *Presence: Teleoperators & Virtual Environments*, 26(1):1–15, 2017. ISSN 10547460. URL `http://search.ebscohost.com.revproxy.brown.edu/login.aspx?direct=true&db=aph&AN=125737645&site=eds-live&scope=site`.

Susanne Mayr, Maja Köpper, and Axel Buchner. Effects of high pixel density on reading comprehension, proofreading performance, mood state, and physical discomfort. *Displays*, 48:41–49, 2017. ISSN 0141-9382. doi: https://doi.org/10.1016/j.displa.2017.03.002. URL `http://www.sciencedirect.com/science/article/pii/S0141938216301494`.

Rory McCloy and Robert Stone. Science, medicine, and the future: Virtual reality in surgery. *British Medical Journal*, 323(7318):912–915, 10 2001. ISSN 0959-8138. doi: 10.1136/bmj.323.7318.912. URL `https://www.bmj.com/lookup/doi/10.1136/bmj.323.7318.912`.

Tony McLoughlin, Robert S. Laramee, Ronald Peikert, Frits H. Post, and Min Chen. Over two decades of integration-based, geometric flow visualization. *Computer Graphics Forum*, 29(6):1807–1829, 2010. ISSN 01677055. doi: 10.1111/j.1467-8659.2010.01650.x.

Tony McLoughlin, Mark W. Jones, Robert S. Laramee, Rami Malki, Ian Masters, and Charles D. Hansen. Similarity Measures for Enhancing Interactive Streamline Seeding. *IEEE Transactions on Visualization and Computer Graphics*, 19(8):1342–1353, 8 2013. ISSN 1077-2626. doi: 10.1109/TVCG.2012.150. URL `http://ieeexplore.ieee.org/document/6231627/`.

Jesper Milàn and Richard G. Bromley. True tracks, undertracks and eroded tracks, experimental work with tetrapod tracks in laboratory and field. *Palaeogeography, Palaeoclimatology, Palaeoe-*

cology, 231(3-4):253–264, 2 2006. ISSN 00310182. doi: 10.1016/j.palaeo.2004.12.022. URL https://linkinghub.elsevier.com/retrieve/pii/S0031018204006728.

Carol Bergfeld Mills and Linda J. Weldon. Reading Text from Computer Screens. *ACM Computing Surveys*, 19(4):329–357, 12 1987. ISSN 0360-0300. doi: 10.1145/45075.46162. URL http://doi.acm.org/10.1145/45075.46162.

Johannes Novotny, Joshua Tveite, Morgan L. Turner, Stephen Gatesy, Fritz Drury, Peter Falkingham, and David H. Laidlaw. Developing Virtual Reality Visualizations for Unsteady Flow Analysis of Dinosaur Track Formation using Scientific Sketching. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):2145–2154, 5 2019. ISSN 1077-2626. doi: 10.1109/TVCG.2019.2898796. URL https://ieeexplore.ieee.org/document/8672601/.

Johannes Novotny, Wesley R. Miller, Franois I. Luks, Derek Merck, Scott Collins, and David H. Laidlaw. Towards Placental Surface Vasculature Exploration in Virtual Reality. *IEEE Computer Graphics and Applications*, 40(1):28–39, 1 2020. ISSN 0272-1716. doi: 10.1109/MCG.2018.2881985. URL https://ieeexplore.ieee.org/document/8586936/.

Steffen Oeltze-Jafra and Bernhard Preim. Survey of Labeling Techniques in Medical Visualizations. In *Proceedings of the 4th Eurographics Workshop on Visual Computing for Biology and Medicine*, VCBM 14, page 199208, Goslar, DEU, 2014. Eurographics Association. ISBN 9783905674620. doi: 10.2312/vcbm.20141192.

Jaisa Olasky, Ganesh Sankaranarayanan, Neal E. Seymour, J. Harvey Magee, Andinet Enquobahrie, Ming C Lin, Rajesh Aggarwal, L. Michael Brunt, Steven D. Schwaitzberg, Caroline G. L. Cao, Suvranu De, and Daniel B. Jones. Identifying Opportunities for Virtual Reality Simulation in Surgical Education. *Surgical Innovation*, 22(5):514–521, 10 2015. ISSN 1553-3506. doi: 10.1177/1553350615583559. URL http://journals.sagepub.com/doi/10.1177/1553350615583559.

Alex Olwal and Steven Feiner. The flexible pointer: An interaction technique for selection in augmented and virtual reality. In *Proceedings of the ACM Symposium on User Interface Software and Technology 2003 (UIST '03)*, pages 81–82, 2003.

Charilaos Papadopoulos, Kaloian Petkov, Arie E. Kaufman, and Klaus Mueller. The Reality Deck–An Immersive Gigapixel Display. *IEEE Computer Graphics and Applications*, 35:33–45, 2015. ISSN 0272-1716. doi: 10.1109/MCG.2014.80. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6866849.

Randy Pausch, Dennis Proffitt, and George Williams. Quantifying immersion in virtual reality. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques (SIGGRAPH '97)*, pages 13–18, New York, New York, USA, 1997. ACM Press/Addison-Wesley Publishing Co., ACM Press. ISBN 0897918967. doi: 10.1145/258734.258744. URL `http://portal.acm.org/citation.cfm?doid=258734.258744`.

Claudio Pensieri and Maddalena Pennacchini. Overview: Virtual Reality in Medicine. *Journal For Virtual Worlds Research*, 7(1), 2014. ISSN 1941-8477. doi: 10.4101/jvwr.v7i1.6364.

Wen Qi, Russell M Taylor II, Christopher G Healey, and Jean-Bernard Martens. A Comparison of Immersive HMD, Fish Tank VR and Fish Tank with Haptics Displays for Volume Visualization. In *Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization*, APGV '06, pages 51–58, New York, NY, USA, 2006. ACM. ISBN 1-59593-429-4. doi: 10.1145/1140491.1140502. URL `http://doi.acm.org/10.1145/1140491.1140502`.

Wolfgang Radner. Reading charts in ophthalmology. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 255(8):1465–1482, 8 2017. ISSN 0721-832X. doi: 10.1007/s00417-017-3659-0. URL `http://link.springer.com/10.1007/s00417-017-3659-0`.

Wolfgang Radner and Gabriela Diendorfer. English sentence optotypes for measuring reading acuity and speedthe English version of the Radner Reading Charts. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 252(8):1297–1303, 8 2014. ISSN 0721-832X. doi: 10.1007/s00417-014-2646-y. URL `https://doi.org/10.1007/s00417-014-2646-y`.

Eric D. Ragan, Regis Kopper, Philip Schuchardt, and Doug A. Bowman. Studying the effects of stereo, head tracking, and field of regard on a small-scale spatial judgment task. *IEEE Transactions on Visualization and Computer Graphics*, 19(5):886–896, 5 2013. ISSN 1941-0506. doi: 10.1109/TVCG.2012.163. URL `http://www.ncbi.nlm.nih.gov/pubmed/22868674`.

Jonathan C. Roberts, Panagiotis D. Ritsos, Sriram Karthik Badam, Dominique Brodbeck, Jessie Kennedy, and Niklas Elmqvist. Visualization beyond the desktop–the next big thing. *IEEE Computer Graphics and Applications*, 34(6):26–34, 2014. ISSN 1558-1756. doi: 10.1109/MCG.2014.82. URL `http://www.ncbi.nlm.nih.gov/pubmed/25137723`.

Ramik Sadana, Vidya Setlur, and John Stasko. Redefining a Contribution for Immersive Visualization Research. In *Proceedings of the 2016 ACM Companion on Interactive Surfaces and*

Spaces, ISS 16 Companion, page 4145, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450345309. doi: 10.1145/3009939.3009946. URL https://doi.org/10.1145/3009939.3009946.

Tobias Schafhitzel, Eduardo Tejada, Daniel Weiskopf, and Thomas Ertl. Point-based stream surfaces and path surfaces. In *Proceedings of Graphics Interface 2007 (GI '07)*, GI '07, page 289, New York, New York, USA, 2007. ACM Press. ISBN 9781568813370. doi: 10.1145/1268517.1268564. URL http://doi.acm.org/10.1145/1268517.1268564.

Philip Schuchardt and Doug A. Bowman. The benefits of immersion for spatial understanding of complex underground cave systems. In *Proceedings of the 2007 ACM symposium on Virtual reality software and technology (VRST '07)*, page 121, New York, New York, USA, 2007. ACM, ACM Press. ISBN 9781595938633. doi: 10.1145/1315184.1315205. URL http://portal.acm.org/citation.cfm?doid=1315184.1315205.

Mel Slater. A Note on Presence Terminology. *Emotion*, 3(3):1–5, 2003. ISSN 12938505.

John J. Socha and Francesco De Carlo. Use of synchrotron tomography to image naturalistic anatomy in insects. In *Proceedings of SPIE, Developments in X-Ray Tomography VI*, page 70780A, San Diego, CA, USA, 2008. SPIE. ISBN 9780819472984. doi: 10.1117/12.795210. URL http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=794866%5Cnpapers2://publication/uuid/BC08B6AF-52AC-493C-B812-3458735AEDCB.

John J. Socha, Wah-Keat Lee, Jon F. Harrison, James S. Waters, Kamel Fezzaa, and Mark W. Westneat. Correlated patterns of tracheal compression and convective gas exchange in a carabid beetle. *The Journal of experimental biology*, 211(21):3409–20, 2008. ISSN 0022-0949. doi: 10.1242/jeb.019877. URL http://www.ncbi.nlm.nih.gov/pubmed/18931314.

Olga Sorkine-Hornung and Michael Rabinovich. Least-Squares Rigid Motion Using SVD. Technical report, ETH Zürich, Switzerland, 1 2017. URL https://igl.ethz.ch/projects/ARAP/svd_rot.pdf.

Alexander Stukowski. Visualization and analysis of atomistic simulation data with OVITO - the Open Visualization Tool. *Modelling and Simulation in Materials Science and Engineering*, 18(1):015012, 2009. doi: 10.1088/0965-0393/18/1/015012. URL https://doi.org/10.1088/0965-0393/18/1/015012.

Ahalya Subramanian and Shahina Pardhan. Repeatability of Reading Ability Indices in Subjects with Impaired Vision. *Investigative Opthalmology & Visual Science*, 50(8):3643, 8 2009. ISSN 1552-5783. doi: 10.1167/iovs.08-2823. URL `http://iovs.arvojournals.org/article.aspx?doi=10.1167/iovs.08-2823`.

Ivan E. Sutherland. A head-mounted three dimensional display. In *Proceedings of the Fall Joint Computer Conference 1968 (AFIPS '68)*, AFIPS 68 (Fall, part I), page 757, New York, New York, USA, 1968. ACM Press. ISBN 9781450378994. doi: 10.1145/1476589.1476686. URL `https://doi.org/10.1145/1476589.1476686http://portal.acm.org/citation.cfm?doid=1476589.1476686`.

J. Edward Swan, Joseph L. Gabbard, Deborah Hix, Robert S. Schulman, and Keun Pyo Kim. A comparative study of user performance in a map-based virtual environment. In *Proceedings of the IEEE Virtual Reality Conference 2003 (VR '03)*, pages 259–266. IEEE, IEEE Comput. Soc, 2003. ISBN 0-7695-1882-6. doi: 10.1109/VR.2003.1191149. URL `http://ieeexplore.ieee.org/document/1191149/`.

Jean Trumbo. The Process of Critique in Visual Communication. *Journalism & Mass Communication Educator*, 52(2):15–23, 6 1997. ISSN 1077-6958. doi: 10.1177/107769589705200202. URL `https://doi.org/10.1177/107769589705200202`.

Morgan L. Turner, Peter L. Falkingham, and Stephen M. Gatesy. It's in the loop: shared subsurface foot kinematics in birds and other dinosaurs shed light on a new dimension of fossil track diversity. *Biology Letters*, 16(7):20200309, 2020. doi: 10.1098/rsbl.2020.0309. URL `https://royalsocietypublishing.org/doi/abs/10.1098/rsbl.2020.0309`.

William N. Venables and Brian D. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002. ISBN 978-0-387-21706-2. URL `http://www.stats.ox.ac.uk/pub/MASS4/`.

Guotai Wang, Maria A. Zuluaga, Rosalind Pratt, Michael Aertsen, Anna L. David, Jan Deprest, Tom Vercauteren, and Sebastien Ourselin. Slic-Seg: Slice-by-Slice Segmentation Propagation of the Placenta in Fetal MRI Using One-Plane Scribbles and Online Learning. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*, pages 29–37. Springer, 2015. doi: 10.1007/978-3-319-24574-4{_}4. URL `http://link.springer.com/10.1007/978-3-319-24574-4_4`.

Z J Wang, X W Zhu, Q T Huang, Z Q Yun, Y W Cao, Y Y Chen, and M Zhong. Three-dimensional reconstruction of human placental vascular network using in-vitro MRI data. *Ultrasound in Ob-*

stetrics & Gynecology, 47(6):790–792, 2016. ISSN 1469-0705. doi: 10.1002/uog.15809. URL http://dx.doi.org/10.1002/uog.15809.

Alexander Wiebel, Frans M. Vos, David Foerster, and Hans-Christian Hege. WYSIWYP: What You See Is What You Pick. *IEEE Transactions on Visualization and Computer Graphics*, 18 (12):2236–2244, 12 2012. ISSN 1077-2626. doi: 10.1109/TVCG.2012.292. URL http://ieeexplore.ieee.org/document/6327228/.

David R. Williams, Geunyoung Yoon, Antonio Guirao, Heidi Hofer, Jason Porter, Scott M. Macrae, Ronald R. Krueger, and Raymond A. Applegate. How far can we extend the limits of human vision. In Ronald R. Krueger, Raymond A. Applegate, and Scott M. MacRae, editors, *Wavefront Customized Visual Correction: A Quest for Super Vision II*, pages 19–38. Slack Inc, Thorofare, NJ, 2004.

Team YouGoWords. You Go Words - Word finder [Accessed: 2018-08-22], 2018. URL http://www.yougowords.com/.

Song Zhang, Caatay Demiralp, Daniel F. Keefe, Marco DaSilva, David H. Laidlaw, Benjamin D. Greenberg, Peter J. Basser, Carlo Pierpaoli, E. A. Chiocca, and Thomas S. Deisboeck. An immersive virtual environment for DT-MRI volume visualization applications: a case study. In *Proceedings of the IEEE Visualization Conference 2001 (VIS '01)*, pages 437–584. IEEE, IEEE, 2001. ISBN 0-7803-7200-X. doi: 10.1109/VISUAL.2001.964545. URL http://ieeexplore.ieee.org/document/964545/.

Malte Zöckler, Detlev Stalling, and Hans-Christian Hege. Interactive Visualization of 3D-vector Fields Using Illuminated Stream Lines. In *Proceedings of the 7th Conference on Visualization (Vis '97)*, VIS '96, pages 107–ff., Los Alamitos, CA, USA, 1996. IEEE Computer Society Press. ISBN 0-89791-864-9. URL http://dl.acm.org/citation.cfm?id=244979.245023.

Appendix A

Evaluating the Effects of VR Fidelity on Data Exploration Performance

A.1 Detailed results of the statistical analysis

Table A.1: Anova Results - Significant effects on time (all tasks)

Source	Nparm	DFNum	DFDen	F Ratio	Prob >F
Platform	2	2	48.0	1.0278577	0.3655
FOR	1	1	48.0	2.784237	0.1017
Platform*FOR	2	2	48.0	0.5482773	0.5815
Stereo	1	1	48.0	20.754133	<.0001
Platform*Stereo	2	2	48.0	0.2768874	0.7593
FOR*Stereo	1	1	48.0	1.5533848	0.2187
Platform*FOR*Stereo	2	2	48.0	0.8287739	0.4427

Table A.2: Group numbers of our twelve experimental conditions

FOR	Stereoscopy	CAVE A	CAVE B	HMD
90	Stereo	1	5	9
90	Mono	2	6	10
270	Stereo	3	7	11
270	Mono	4	8	12

Table A.3: Significant differences between CAVE A groups. Conditions by group number are (1) 90 stereo, (2) 90 mono, (3) 270 stereo, (4) 270 mono.

Metric	Task	Groups	$\hat{p}$	CI_{lower}	CI_{upper}	p
Time	10	2,4	0.00	0.0000	0.6161	0.032
Time	6	2,4	0.10	0.0198	0.5543	0.025
Time	4	2,4	0.12	0.0133	0.5791	0.042
Time	2	2,4	0.10	0.0188	0.5816	0.035
Time	6	1,2	0.96	0.5364	0.9980	0.006
Time	1	1,2	1.00	0.3838	1.0000	0.032
Time	6	2,3	0.00	0.0000	0.5837	0.032
Time	4	2,3	0.00	0.0000	0.6018	0.032
Time	2	2,3	0.08	0.0157	0.5647	0.024
Time	1	2,3	0.12	0.0107	0.6308	0.065
Time	4	1,3	0.02	0.0009	0.2979	$<$0.01
Accuracy	13	1,3	0.14	0.0175	0.5980	0.042
Accuracy	1	1,2	0.2	0.0415	0.5905	0.047

Table A.4: Significant differences between CAVE B groups. Conditions by group number are (5) 90 stereo, (6) 90 mono, (7) 270 stereo, (8) 270 mono.

Metric	Task	Groups	$\hat{p}$	CI_{lower}	CI_{upper}	p
Time	3	5,6	0.04	0.0019	0.4736	0.006
Time	5	5,8	0.08	0.0058	0.5647	0.024
Time	8	6,7	0.00	0.0000	0.6019	0.032
Time	8	6,7	0.00	0.0000	0.6161	0.032
Grade	6	5,8	0.16	0.0230	0.6057	0.05
Grade	13	5,8	0.14	0.0191	0.5764	0.032

Table A.5: Significant differences between HMD groups. Conditions by group number are (9) 90 stereo, (10) 90 mono, (11) 270 stereo, (12) 270 mono.

Metric	Task	Groups	$\hat{p}$	CI_{lower}	CI_{upper}	p
Time	12	9,10	0.08	0.0060	0.5544	0.024
Time	03	9,10	0.88	0.3990	0.9878	0.042
Time	12	9,11	0.08	0.0058	0.5647	0.024
Time	06	9,11	0.10	0.0094	0.5642	0.025
Time	05	9,11	0.12	0.0127	0.5914	0.042
Time	11	9,11	0.08	0.0060	0.5544	0.024
Time	11	10,11	0.04	0.0019	0.4736	0.006
Time	14	10,11	0.00	0.0000	0.6161	0.032
Time	05	9,12	0.08	0.0073	0.5059	0.010
Accuracy	14	9,12	0.90	0.4685	0.9892	0.015
Accuracy	10	9,11	0.88	0.5139	0.9807	0.007

Table A.6: Significant differences found between same conditions on the three platforms.

Metric	Condition	Best			Tasks
Time	Mono 270	CAVE A	$<$	HMD	1
Time	Stereo 270	CAVE A	$<$	HMD	4
Time	Mono 90	CAVE B	$<$	HMD	3
Time	Mono 90	CAVE A	$<$	HMD	3
Time	Mono 90	HMD	$<$	CAVE B	13
Time	Mono 90	HMD	$<$	CAVE A	13
Accuracy	Mono 270	CAVE B	$>$	HMD	13
Accuracy	Mono 270	CAVE B	$>$	CAVE A	15
Accuracy	Stereo 270	HMD	$>$	CAVE A	7,10
Accuracy	Stereo 270	HMD	$>$	CAVE B	10
Accuracy	Stereo 90	HMD	$>$	CAVE B	6,14
Accuracy	Stereo 90	CAVE B	$>$	CAVE A	6,8
Accuracy	Stereo 90	HMD	$>$	CAVE A	14
Accuracy	Mono 90	HMD	$>$	CAVE A	1
Accuracy	Mono 90	CAVE B	$>$	HMD	11
Accuracy	Mono 90	CAVE B	$>$	CAVE A	1,15