SCHOOL OF COMPUTATION,
INFORMATION AND TECHNOLOGY

TECHNICAL UNIVERSITY OF MUNICH

Bachelor's Thesis in Informatics: Games Engineering

Comparing Information Visualization Modalities for 2D Diagrams on Handheld Devices for Industrial Augmented Reality Applications

Tim Brüggemann

SCHOOL OF COMPUTATION,
INFORMATION AND TECHNOLOGY

TECHNICAL UNIVERSITY OF MUNICH

Bachelor's Thesis in Informatics: Games Engineering

Comparing Information Visualization Modalities for 2D Diagrams on Handheld Devices for Industrial Augmented Reality Applications

Vergleich von Informationsvisualisierungs-Modalitäten für 2D-Diagramme auf tragbaren Geräten für industrielle Augmented-Reality-Anwendungen

Author: Tim Brüggemann
Supervisor: Prof. Gudrun Klinker, Ph.D.
Advisor: Linda Rudolph, M.Sc.
Submission Date: 15.01.2024

I confirm that this bachelor’s thesis in informatics: games engineering is my own work and I have documented all sources and material used.

Munich, 15.01.2024 Tim Brüggemann

Acknowledgments

I would like to express my sincere gratitude to my advisor, Linda Rudolph, for her invaluable guidance, support, and availability throughout the course of this research. Her expertise and insights have been instrumental in shaping this work.

I would also like to thank everyone who participated in my user study. Their feedback and input were extremely varied and helped me to gain new perspectives for my research.

Furthermore, I am grateful to my friend, Jack, for his valuable inspiration and support for my writing process.

Finally, I would like to thank my parents for their unwavering support and encouragement. Their feedback and guidance have been invaluable to me throughout my studies.

Abstract

The goal of this thesis is to examine and compare two data visualization methods for diagrams, namely a traditional 2D representation of a diagram on the screen of a mobile device on the one hand and on the other hand the placement of an image plane containing the diagram as an AR-element in a 3D immersive scene. Therefore, a prototype was developed to examine the feasibility of the two methods, both in regards to technical realization and user acceptance. As a result of the subsequent user study, a linked 2D and 3D approach is recommended. The AR visualization provides great overall orientation and navigation, while the 2D visualization offers more precise manipulation and spatial flexibility. The main hurdle in this comparison is the fact that the 2D visualization is already well-established and ever-present in our daily lives, which influenced the opinions of many participants. With the shift in demographics and children coming in contact with AR at much earlier ages, this remains an interesting topic in need of further investigation in the future.

Contents

1 Introduction

1.1 Use Case and Problem Statement

More than 50 years ago, Spence and Drew [1] first elaborated on the new opportunities for industrial information visualization of electronic circuit designs following the technical advances of increasing computational speed and display resolution. What they accentuated specifically was the interactive nature offered by such digital representations and the allowance of dynamic real-time data visualization [1] and manipulation [2].

Just a few years prior, the first head-mounted-display (HMD) had been invented by Ivan Sutherland [3], and as the technology evolved, information visualization researchers were eager to explore the many possibilities of 3D data representations and immersive environments such as Augmented Reality (AR) or Virtual Reality (VR)[1], believing 3D visualization to be superior over traditional 2D representations for abstract data visualization. This sentiment is what Munzner [5] called the "unbridled enthusiasm" for 3D representations during the late 1980s and early 1990s. This avid fascination, however, declined substantially after subsequent user studies were not able to confirm the assumed superiority of 3D representations over traditional 2D representations [6]. The pursuit of AR as a tool for abstract data visualization subsided primarily due to the limitations of technology at that time, especially insufficient tracking accuracy and uncomfortably heavy HMDs with a limited field of view [7]. Ever since, information visualization researchers have shifted their focus to flat 2D data visualization, primarily designed for desktop computers, and the general mood regarding spatially immersive technologies such as AR for abstract data visualization has turned into skepticism.

However, with the rise of Industry 4.0, the advancing digitization of data, and the progress of AR technologies – making AR accessible to the masses and easy to wear by introducing commodity immersive devices such as the HTC Vive or Microsoft HoloLens and even advancing onto handheld mobile devices – it is timely to reconsider the value of 3D representation for abstract data visualization.

1.2 Starting Point

This thesis is based on a research paper by Büchner et al. [8]. That paper explores the use of AR for information visualization in an industrial context, namely corrective

[1]While VR totally immerses the user in a purely virtual environment, AR overlays the physical environment with virtual information [4]

maintenance in chemical industries, and how or if this technology can be appropriately integrated into existing business processes.

The piping and instrumentation diagrams (P&ID) pose a primary visualization problem in this research. Those diagrams show the interconnection of mechanical parts and process equipment, and as the associated chemical plants can extend over several floors and even buildings, the P&IDs can quickly get overly complex and are not suited to be presented on a small mobile screen. Usually, this diagram format is printed in A0 and, unfortunately, needs better complexity management. Thus, users easily reach their perceptual limits – meaning with increasing diagram size it gets harder to discriminate between different diagram elements – and their cognitive limits as well – meaning the working memory capacity dictates how many diagram elements can be comprehended at a time [9]. Consequently, to ensure usability on a small handheld screen, an intuitive navigation method is essential.

Büchner et al. [8] propose two strategies to visualize the P&IDs: a traditional 2D representation on the flat screen of a mobile device (Figure 1.1c); or the placement of the diagram as an AR element in an immersive 3D scene (Figure 1.1B). No big-scale user studies have yet been conducted to analyze which of those visualization methods (if any) is better suited to navigate around complex diagrams.

Figure 1.1: "Example functionalities in augmented reality user interfaces for digital twins. From left to right: a) indicating the position of a not-visible component. B) positioning of virtual functional diagrams in the room c) placing functional diagrams as two-dimensional user interface elements on the hand-held screen and highlight components on them. D) highlighting of requested plant components directly with a visual overlay." [8]

1.3 Research Objectives

The purpose of this study is to further examine the research of Büchner et al. [8] and carefully analyze whether the visualization of abstract data diagrams (not just limited to P&IDs) can benefit from immersive 3D presentation or whether a traditional 2D visualization is just as or even more expedient.

The first objective in order to achieve this is the development of a prototypical application that offers the possibility to compare the two visualization methods proposed by Büchner et al. [8] as well as adding new features that facilitate the use and navigation of each method, respectively.

Subsequently, a small-scale pilot study and a larger-scale user study were designed and conducted to analyze whether 3D information visualization offers benefits over 2D presentations or is still met with skepticism. Furthermore, the user study aims to improve the prototype and get an impression of which features might be unnecessary, missing, or in need of improvement.

The code written for this thesis, the materials from the user study, and all original images used in this thesis, are provided on Gitlab [10].

2 Fundamentals

2.1 3D Data Visualization

In any discussion about 3D data visualization, the term 3D first needs to be disambiguated, for in the context of data visualization, it has had different meanings in different time periods. In the 1980s, 3D visualization usually indicated three-dimensional data (objects with three spatial dimensions) being presented on a flat electronic screen (with two spatial dimensions), while in more recent times, the term 3D visualization rather refers to data of any dimensions being presented in a three-dimensional immersive scene (e.g. AR, VR) [11]. Unless stated otherwise, in the following chapters of this thesis, the implied meaning of 3D is the latter definition.

2.2 Situated Analytics

2.2.1 Definition

In order to understand Situated Analytics, first, a complementary technique called Immersive Analytics must be defined in the words of a leading researcher in this field, Tim Dwyer: "Immersive Analytics is the use of engaging, embodied analysis tools to support data understanding and decision making." [12] Immersive Analytics is a visualization technique that draws from a broad range of display and interaction technologies and explores the appropriateness of human-computer interfaces.

Situated Analytics [13], more specifically, refers to 3D visualization of data in relation to a referent in the physical world (e.g., objects, places, or persons). For example, the physical referent could be products in a supermarket, and the visualized data could be an ingredient list that is virtually overlaid to appear next to the physical products. Thus, Situated Analytics combines elements of Visual Analytics and Augmented Reality to offer "a new form of in-situ interactive visual analysis" [14].

2.2.2 Further Characterization

What distinguishes Situated Analytics from Visual Analytics is the situatedness of information, which describes the degree to which the information is connected to the task or location. More specifically, a visualization of data is spatially situated if it is close to its physical referent object. The definition of what is close remains vague on purpose. [14] For example, while seeing a virtual ingredient list next to a physical product in a supermarket is spatially situated, seeing this same ingredient list in an online shop is not.

A distinction must be made between physically situated visualizations and perceptually situated visualizations. While the former requires the data visualization and the physical referent to be physically close, the latter indicates they just appear to be close. [14] For example, the height of a mountain being displayed on a digital screen appears to be near the mountain while actually the data visualization is just pixels on a device's screen, which is situated possibly multiple kilometers away from the physical referent, the mountain. This data is merely perceptually situated in space.

Another variant of situatedness is temporally situated visualization, meaning "the data's temporal referent is close to the moment in time the physical presentation is observed" [14]. For example, a car's speedometer is temporally (and spatially) situated because it shows real-time data. A car's odometer, however, is not strongly temporally situated as it mainly shows data from the past [14].

2.2.3 Advantages

Traditional non-situated interactive visualization systems offer several interaction operations on the data, such as zooming, highlighting, isolating, or changing the visual style. An extensive discussion of those operations has already been done by Jansen and Dragicevic [15]. Suppose, additionally, the visualization system is spatially and temporally situated. In that case, the data can also be interacted with via the physical referent, allowing analytical data analysis and physical action to intertwine in real-time [16], for example, fixing a thermal leak in a room while the airflow is visualized on a situated screen (e.g. a HMD); the visualized data interactively adapts to the physical actions on the referent, a possibility of interaction that traditional non-situated visualization systems usually do not offer. [14]

Another advantage is the sensorimotor contingencies of AR systems, which, according to Slater and Wilbur, further the user's immersion and provide a more realistic and natural interaction with immersive virtual scenes [17]. Sensorimotor contingencies include, for example, moving the body or head to change the visual perspective, as is the case with HMDs. [18]

Finally, Situated Analytics aims to steer the user away from their workspace and into the everyday physical environment [14], which is supported by wearable and mobile computing [19] of immersive AR technologies and fits right into Mark Weiser's vision of ubiquitous computing which strives to provide the user with several computers at their disposal, however, mostly without being aware of them [20]. Since Situated Analytics is predestined to be used in immersive AR systems, it is of utmost importance to develop appropriate strategies and guidelines on how to best implement Situated Analytics so that one day, this method might become a seamless part of our everyday life.

2.3 Related Work

Twenty years ago, a research paper by Appel and Navab [21] proposed a strategy to combine AR and chemical industry, even though the AR technology had not been developed as thoroughly back then, and the term AR in the paper rather refers to still images which overlay virtual information over the physical environment. The paper presents a method of aligning and merging technical drawings with a database of calibrated images of chemical plants (Figure 2.1). Sparse pairs of matching points between the technical drawing and multiple photogrammetric images of the plant are used to project the drawing into the world coordinate system of any image in the database by calculating matrix transformations for the drawing, which adjust the size, orientation, and position accordingly. To obtain these point correspondences different marker-based and marker-less methods are proposed.

Figure 2.1: "Visual combination of a photogrammetric image and a technical drawing" [21]

This method of visualizing additional data for the chemical plant is – just like the visualization method for P&IDs proposed by Büchner et al. [8] – perceptually situated in space, only that the degree of situatedness is much higher as the technical drawing spatially coincides with the actual plant. This stronger form of spatial situatedness is called spatial embeddedness. Data visualization is spatially embedded if each of its sub-presentations (e.g., each part of the technical drawing) is spatially aligned with their corresponding physical sub-referent (e.g., each part of the physical plant) [16].

3 The Application

3.1 Introduction

The application "Comparing Diagrams" was created as a prototype to implement the two visualization methods for P&IDs described by Büchner et al. [8]. The goal of the prototype was to examine the feasibility of the two methods, both in regards to technical realization and user acceptance. Therefore, the initial ideas of Büchner et al. [8] were integrated, improved with new features, and tested in a usability study. Instead of P&IDs, more simple decision tree diagrams were utilized for the prototype as they do not need any specialized knowledge to understand and allow the results of this thesis to be more generally applicable.

The application was developed with Unity Version 2021.3.3f1 using C# and the ARFoundation Package Version 4.2.3. It is available on Android and has been extensively tested on a Google Pixel 2 5" (Android 11) and a Samsung Galaxy Tab S8 11" (Android 13).

3.2 Requirements

The two visualization methods were assigned the following designations:

Screen Space refers to the traditional 2D visualization on the flat screen.

World Space refers to visualizing an image-plane as an object in a 3D immersive AR scene.

The following requirements should be met by the application:

R1 move and scale the diagram in Screen Space

R2 allow an easy reset of the diagram's position and scale in Screen Space

R3 do not allow the diagram to be moved beyond the boundaries of the screen in Screen Space

R4 scan a marker and project the diagram onto an image plane in World Space

R5 pitch rotate the diagram in World Space

R6 allow an easy realignment of the diagram to face the camera in World Space

R7 provide support to find the diagram if it leaves the visual area of the scene in World Space

R8 change the visualization method

R9 no display timeout while working with the application

R10 track the user's time needed for completing tasks, the number of button presses, and average values of scale factor in Screen Space as well as the distance to the diagram, pitch rotation of the diagram, and rotation of the diagram relative to the camera in World Space

R11 clear and intuitive UI design

3.3 Design Decisions

The requirements above are implemented as follows:

D1 There already exist well-established design conventions regarding the manipulation of a Screen Space UI element. Therefore, a finger-swipe motion moves the diagram, and a two-finger pinch scales it. While performing both tasks, the fingers are anchored to the diagram on their touch position to avoid discrepancies between the fingers and the corresponding motion of the diagram. Instead of abruptly removing any velocity when releasing the finger after moving, the diagram retains some momentum and smoothly comes to a halt, which users perceive to be more natural. Additionally, a double tap results in a quick zoom-in and zoom-reset, respectively.

D2 Even though the double tap results in a reset of position and scale, this behavior might not be intuitive or known to all. Thus, a delegated `Center` button is provided, which implements the same resetting functionality as the double tap – if the diagram has been zoomed in (Figure 3.1). Both the double tap and the `Center` button adjust the diagram's position and scale with a smooth animation instead of an abrupt change to aid the spatial understanding of the large diagram. If the diagram has been reset with either the double tap or the `Center` button, the button's functionality transforms and allows the user to go back to the last position and scale before reset (Figure 3.2). This reversibility is recommended as a means for error robustness in ISO 9241-110 [22]. Once this functionality has been used or the diagram has been moved or scaled in any other way, the `Center` button resumes its original reset functionality. When initially loading the Screen Space scene, the `Center` button is unavailable and greyed out as the diagram is initialized in its reset state.

D3 The user cannot, by any operation, transform the diagram in such a way that it leaves the visible screen (error prevention as part of error robustness in ISO

9241-110 [22]). The threshold that the farthest diagram edge cannot travel beyond is 20 % of either the screen's width or height, depending on which value is smaller. This way, the minimum distance from the diagram to the screen edge is uniform on both screen axes (Figure 3.3). The `KeepOnScreen` component works independently from any other script, diagram, and screen dimensions. Thus, it can be easily reused and applied to any UI element in Unity. The percentage threshold can easily be adjusted via a serialized field in the editor.

D4 For the image marker (Figure 3.4), the marker generator website BrosVision[1] is used, as their randomly generated markers are unique and optimized for use as AR image targets. The printed marker is glued on an acrylic plate to smoothly flatten the paper. If no marker has yet been scanned in the application, the visual prompt "Scan Marker" is shown on the screen to provide the user with a call to action in an otherwise empty scene (Figure 3.5). After being scanned successfully, the marker acts as a parent to the instantiated diagram in World Space, enabling the repositioning of the diagram by moving the marker accordingly, which, however, is not recommended. Instead, it is advised to fixate the marker on the ground and have it act as just a tracking stabilizer. As the diagrams are intended to be quite large, inspecting the outer regions can easily lead to the camera losing track of the marker, so extended tracking is used to accommodate for that. The extended tracking works better if there are recognizable landmarks in the environment and the marker is not moved.

D5 The image plane can be pitch rotated by swiping vertically with one finger (Figure 3.6). Yaw rotation is not supported as this swiping method of manipulation is not intended for dynamic interaction purposes in World Space; instead of moving and scaling the diagram, the user is supposed to move their screen device around the diagram in order to get the desired view perspectives. This manual pitch manipulation is rather intended for the initial setup right after the marker has been scanned to accommodate for suboptimal marker placements; for example, the marker itself is pitch rotated or positioned relatively high to make an orthogonal alignment of the diagram with respect to the marker more comfortable. As the diagram is rendered on a 3D plane in a 3D immersive scene, it is possible to view the back of the plane by walking around the diagram or by excessively pitch rotating. It was decided against rendering the diagram on the back as well since this feels unnatural and would likely confuse users. Instead, the back is rendered with a uniform stripe pattern to clearly distinguish it from any real-life background in the scene. The equally spaced stripes furthermore give a better understanding of the perspective in case the user looks at the diagram from very sharp angles (Figure 3.7).

D6 A disadvantage the visualization in World Space poses is the stationary anchoring of the diagram in the scene, discouraging the user from moving dynamically

[1] `https://www.brosvision.com/ar-marker-generator/`

through the room. To circumvent this fixation, the `LookAtMe` button allows re-orienting the diagram to face the camera directly. As this is not an established feature the users would be familiar with, particular emphasis is placed on a smooth readjustment of the rotation instead of an abrupt one to not confuse users about what action exactly has been performed on the diagram. Additionally, if, through various manipulations, the striped back of the diagram becomes visible, the `LookAtMe` button is highlighted to offer users a call to action (Figure 3.7). If, however, the diagram is not visible at all – neither the front nor the back – the `LookAtMe` button is greyed out because using it in this situation would not produce any observable effects, possibly confusing the user (Figure 3.5, Figure 3.8).

D7 If the camera turns away from the image plane, a tiny arrow is displayed on the screen edge (Figure 3.8) and shows the direction back to the image plane (error prevention is not possible; instead, error management as part of error robustness in ISO 9241-110 [22]). As the arrow does not exist in the scene but rests on the screen, it can only show directions in 2D. This two-dimensionality is complicated to interpret if the image plane is hidden directly behind the user; however, in most scenarios, the image plane would be positioned somewhere next to the user, in which cases the 2D arrow can be much more readable and clear. It was decided against a 3D arrow because, for maximum readability, this arrow would have to be positioned directly at the center of the screen, making it too intrusive, while the 2D arrow can be subtly delegated to the edges of the screen.

D8 There is a special sandbox mode where users can freely explore and compare the two methods. The method can be switched with a button press next to which the current method is displayed (Figure 3.5). The sandbox mode visualizes a sample picture of the Mona Lisa instead of a diagram and should support the user in getting familiar with the functionality of the prototype before working on the user study tasks. No data is logged in the sandbox mode.

D9 While starting the application, a script sets the sleep timeout of the screen to `NeverSleep`. This is especially useful for World Space, as this method uses the camera and accelerometer as the primary input source.

D10 There is a hidden developer menu accessible with a double tap of two fingers. The menu shows additional information to the user test supervisor and allows them to start and stop a test iteration (Figure 3.9). After the supervisor has started the test, the necessary information is logged automatically and written to a text file once the supervisor has stopped the test. The files are associated with a numerical pseudonym identifier that can be tweaked manually in the developer menu or, alternatively, increments by one automatically each time the application is started. The language of the User Experience Questionnaire can be changed to a certified German translation [23] if participants prefer this.

D11 All buttons adhere to the recommended contrast ratio of at least 4.5:1 between text and background color, according to the latest Web Content Accessibility Guidelines (WCAG) 2.1 AA [24]. Each button features a descriptive text coupled with a simplistic monochrome icon. Buttons that cannot be used in the current circumstances are greyed out. If appropriate in certain scenarios, useful buttons are highlighted with a blinking color.

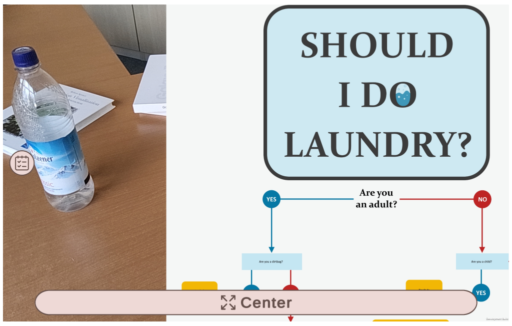


Figure 3.1: The `Center` button can be used to reset the diagram's position and scale

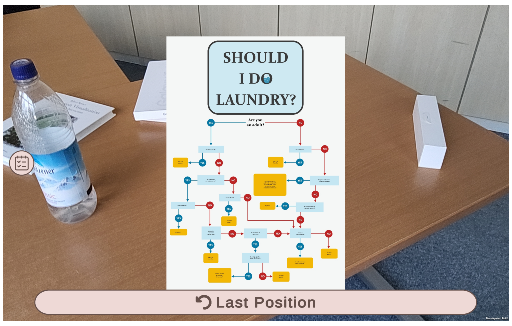


Figure 3.2: Press the `Center` button again to revert to the previous position and scale

Figure 3.3: The diagram cannot be moved beyond the red area (20 % screen height)

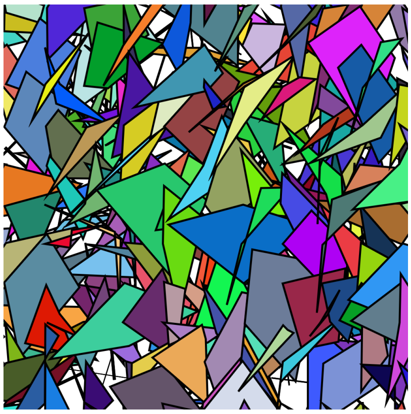


Figure 3.4: Randomly generated markers from BrosVision are unique and optimized for use as AR image targets

Figure 3.5: If no marker has yet been scanned, a visual prompt is shown on the screen and the `LookAtMe` button is greyed out; the sandbox mode offers a button to switch between visualization methods

Figure 3.6: The image plane is positioned in 3D space and can be pitch rotated by swiping up and down

Figure 3.7: The back of the image plane is striped for better visibility and the `LookAtMe` button is highlighted

Figure 3.8: If the camera turns away from the image plane, a tiny arrow is displayed on the screen edge and the `LookAtMe` button is greyed out

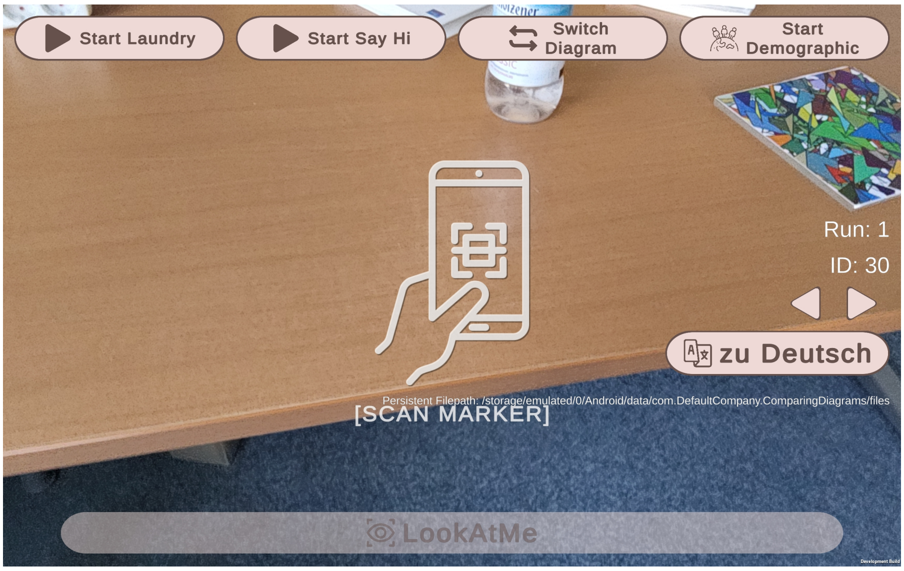


Figure 3.9: The developer menu contains buttons to start the test (+logging) with either diagram, to manually switch the diagram, to start the demographic questionnaire, to adjust the numerical pseudonym identifier and to switch the language of the User Experience Questionnaire to a certified German translation [23]

3.4 Technical Implementation

3.4.1 Screen Space

Movement

The Unity events `OnPointerDown` and `OnPointerUp` are used to track the touches on the screen. Touches in the TouchPhases `Ended` or `Cancelled` are ignored, as well as touches whose position is tracked outside of the image UI element. If the number of active touches on the screen increases from zero to one or decreases from two to one, the movement phase is initialized by saving the current offset between the touch position and the image position. This offset is then used to calculate the new position of the image relative to the touch position in each frame, resulting in a direct translation of the finger movement instead of a relative one.

The movement phase ends if the number of active touches on the screen reaches zero. In this instant, the current velocity of the moving image is calculated by dividing the delta position by the delta time. This velocity is gradually decreased by multiplication with a damping factor in each frame, and until the velocity approximates zero, it is applied to the position of the image in order to retain momentum after release and create a smooth deceleration effect.

Scaling

If the number of active touches on the screen reaches two, the scaling phase is initialized by saving the initial distance between the two touch locations as well as the initial scale of the image. Then, in each frame, the scale is equal to the initial scale multiplied with a scale factor, which is computed by dividing the current distance between the two touch locations by the initial distance.

The issue that remains is finding the correct pivot point around which to scale the image. This is achieved by interpolating between the two touches' positions according to their respective delta movements. For example, if both touches have moved the same amount in perfectly opposite directions, the pivot point is the exact midpoint between the two touches. If, however, one touch has not moved at all, this stationary touch is used as the pivot point. All constellations in between need to be correctly interpolated.

If both touches have moved the same amount in the same direction, no scaling is applied, and instead, the image is moved with two fingers. This can be realized by interpolating between the two touches' delta movements according to their respective delta movements and adding the resulting vector to the image's position at the end of the scaling algorithm.

Lastly, the standard `UnityEngine` library does not provide a method for scaling an object around a pivot other than the object's center. In order to achieve this, next to scaling the object, the offset between the pivot point and the image position is computed. The offset is then scaled with the same scale factor as the image and added to the original pivot position. The resulting vector equals the new position of the image.

Double Tap and `Center` button

The double tap, as well as the `Center` button (R2), both either zoom into the image or out, depending on the current interaction state. The double tap zooms in if the image has not been scaled in any way (but possibly moved). The double tap zooms out (ergo resetting position and scale) if the image has been scaled in any way (and possibly moved). The `Center` button resets position and scale if the image has been scaled or moved in any way. Directly after the `Center` button has been used in this manner or the position and scale have been reset with a double tap, the `Center` button switches functionality to revert back to the last position and scale values before reset. The `Center` button resumes its original functionality if the image is moved or scaled in any way. The zoom is animated gradually within an `IEnumerator`, cannot be canceled – except with the Center-button itself – and precludes any other simultaneous actions such as moving or scaling.

To realize this functionality more elegantly, a state machine was implemented (Figure 3.10). The possible actions are:

IDLE there is no interaction

MOVING the user swipes with one finger

SCALING the user pinches with two fingers

SLIDING the movement phase has just ended

CENTER the `Center` button is pressed

DOUBLETAP the user quickly taps two times

The possible states are:

RESET the initial state when starting the application; `Center` button is unavailable and greyed out

RESET_MOVED the `Center` button has been pressed while the image had been moved but not scaled

RESET_ZOOMED either a double tap has occurred while the image had been scaled and possibly moved or the `Center` button has been pressed while the image had been scaled and possibly moved

MOVED the image has been moved but not scaled

ZOOMED the image has been scaled and possibly moved

This state machine replaced numerous boolean variables and nested if-statements, thus making the code a lot cleaner and less prone to errors.

Keep on Screen and Aspect Ratio

To keep the image element bound inside the screen according to R3, the edge coordinates of the image are computed using the `RectTransform.anchoredPosition`, the `transform.localScale`, and the `RectTransform.rect.width` and `height`, respectively, to clamp the image's position to the edges of the canvas including the buffer zone that can be specified in percentage.

The Unity `image` component offers a native preserve aspect option; this option, however, is not feasible as it only adjusts the visual presentation of the image without changing the underlying width and height values of the `rectTransform` component, thus making the correct size values inaccessible via code. Consequently, the correct aspect ratio is achieved by directly multiplying the x and y values of the scale component with the aspect ratio of the currently used image texture. As the image object is represented in 2D, the z value of the scale component can be repurposed to act as a reference value of the actual scale, which can be changed by the various aforementioned methods of manipulation. Additionally, by dividing every scale value by the z value, the scaling can be reset without losing the aspect ratio. Implementing the aspect ratio this way, the script can also be reused for the image plane in World Space.

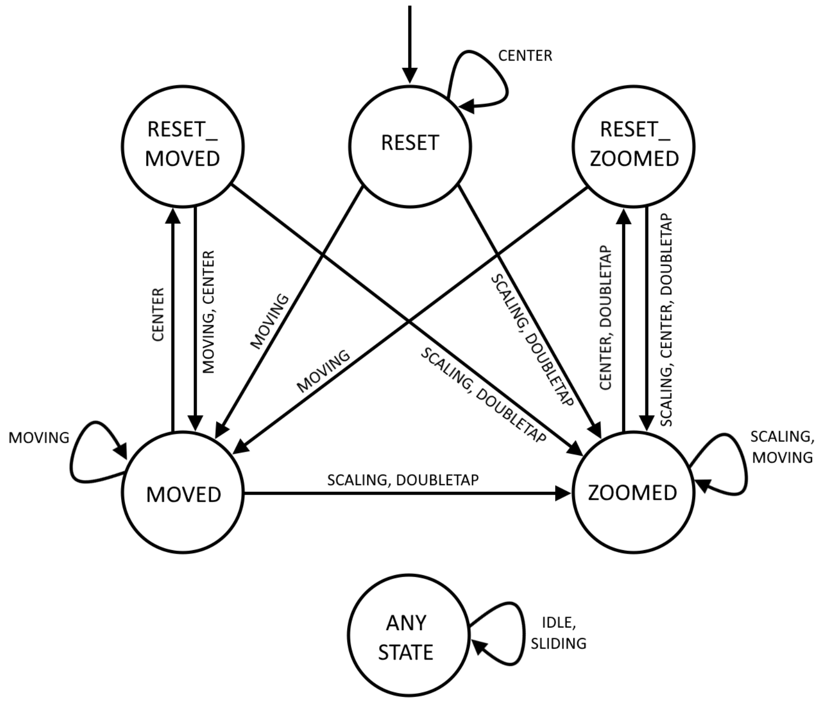


Figure 3.10: The state machine for diagram manipulation in Screen Space

3.4.2 World Space

Image Tracker

The marker tracking (R4) is realized with a `ARTrackedImageManager` component from the ARFoundation package, which exposes an event `trackedImagesChanged`. A script subscribes to this event and observes the tracking state of the marker that is saved in a corresponding `ReferenceImageLibrary`.

Rotation and `LookAtMe` button

The image plane is rotated (R5) by applying the delta position of a swiping touch, and the `LookAtMe` button (R6) calculates the quaternion, which rotates the plane so its normal faces the camera with a gradual animation using an `IEnumerator`.

To highlight the `LookAtMe` button, if the back of the image plane is visible, the dot product between the back's normal, and the vector pointing from the plane to the camera is calculated. If the dot product is positive, the back is visible to the camera.

Target Indicator Arrow

The target indicator arrow (R7) is implemented according to [25].

First, the image plane's position is transformed from world coordinate space to the screen space of the main camera. A negative z value of the screen coordinates indicates that the image plane is located behind the camera. In this case, the screen coordinates need to be multiplied by minus one, otherwise the position would be inverted. The arrow image is displayed on the two-dimensional screen; thus, the z value does not influence the final presentation, however, it should be set to zero nonetheless. If the z value is too big, the arrow image might be located too far away from the camera frustum, resulting in it being culled for performance and not rendered at all.

Next, the relative x and y distances between the canvas center and the image plane are calculated to determine whether the vector to the image plane intersects first with the y border of the `canvasRect` or the x border. This is required to see which border needs to be set to the maximum value and at which border the arrow image needs to be moved accordingly (up and down if the x value is fixed or left and right if the y value is fixed). The correct position of the nonfixed border can be calculated with the fixed border, the angle to the image plane, and trigonometry (Figure 3.11).

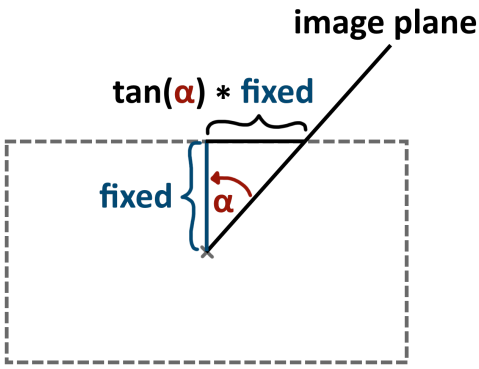


Figure 3.11: Calculate the correct coordinates of the arrow image

4 Pilot Study

4.1 Introduction

During the development of the prototype and the design of the large-scale user study, three iterations of a small-scale pilot study were conducted. The first iteration focused on assessing the functionality of the prototype and getting early feedback, and the second test phase aimed at checking the refinements made to the prototype and finding crucial bugs. The final iteration included the diagrams and respective tasks that would be used in the actual user study. All iterations of the pilot study were conducted with the same two test participants in a consistent location. The two participants were tested separately from each other.

4.2 Methods

The procedure of the first two test iterations was identical. First, the test supervisor presented the two visualization methods and explained his thought process regarding the available features of the prototype. Next, the test participants were handed the prototype to try both methods freely and ask any questions. Instead of a diagram, a geospatial map was used as the enormous size of this image offered a wider range of interaction with the methods and incited the test participants to test the application to fuller extremes. While testing the prototype, the participants were encouraged to immediately mention any strengths or weaknesses to the test supervisor. Afterward, the participants and the supervisor had an informal discussion about how to improve the prototype.

In the third iteration, the geospatial map in the prototype was replaced by the diagrams intended for the large-scale user study. The respective tasks were given to the test participants on a printed sheet. As the participants were already familiar with the functionality of the prototype from the previous two iterations, no more instructions were given by the test supervisor. The test participants were asked to answer the tasks in written form and were not allowed to ask questions while the test supervisor silently observed their behavior. Afterward, the results of the tasks were informally discussed, and the participants explained the thought processes behind their answers. If a false answer had been given, the supervisor and participant tried to retrace and comprehend why the mistake had been made and how the wording could be improved to avoid further misunderstandings in the future.

4.3 Results

In total, the discussions with the test participants yielded the following suggestions for improvement:

Regarding the visualization method in World Space, the test participants quickly discovered the display timeout after a few minutes of not touching the screen in AR mode. Furthermore, the participants expressed they were having difficulties with the white background of the image plane as it was not well distinguishable from a uniform white wall, and the lack of shade and texture on the image plane background made figuring out the correct perspective nonintuitive. Additionally, the first implementation of a 3D arrow pointing in the direction of the image plane if out of sight was met with skepticism from both test participants, who deemed the large arrow resting at the center of the screen as cumbersome and intrusive. One participant noted that the image plane might be too big for a smaller room and suggested adding a manual scale configuration as well as the option to remove a diagram from the 3D scene again. The other participant encountered the issue of moving too close to the image plane so that the camera would go through. Not only did this effectively limit the possible zoom factor to an unsatisfying value, but also would this situation occur relatively suddenly with no clear indication or predictability.

Regarding the Screen Space method, both test participants noted how the current implementation of moving the diagram with a finger swipe felt stiff and abrupt, which helped to improve the underlying mechanics in future iterations. Moreover, the button that resets the diagram's position and scale was initially labeled as the `Reset` button. Both participants hesitated to use this `Reset` button as they were unsure how much of the application would be reset. When they – prompted by the test supervisor – did use the `Reset` button, they stated that it did not feel like an actual reset due to the smooth animation with which the position and scale gradually returned to their default values. One participant also pressed the `Reset` button accidentally while being zoomed in pretty far and expressed frustration because it was not easy to get back to the specific view again. The participant also suggested a minimap, a small summary diagram in a corner to support the user's overall orientation.

4.4 Conclusion

Subsequently, the following changes have been made to the prototype:

- There is no more display timeout.
- The arrow pointing to the image plane was changed from 3D to 2D.
- The white background of the image plane was changed to a uniformly striped pattern.

- Momentum was added to the movement of the diagram via swiping to make it feel more natural.
- The `Reset` button was renamed into the `Center` button to more clearly represent its behavior.
- The recovery functionality to revert to the last position and scale was added to the `Center` button.

The following solutions were derived in discussion with the test participants, but those features were decided not to be relevant for the follow-up user tests. Hence, they were not implemented:

- A long tap on the image plane in World Space opens a pop-up menu to allow the following actions:
  - Scale the image plane.
  - Rotate the image plane on all axes if necessary.
  - Delete the image plane and re-scan the marker (or scan a different marker).
- Getting so close to the image plane in World Space that the camera would go through could be handled with different solutions:
  - Show a text warning if the camera gets too close.
  - A filling bar indicates the distance to the image plane.
  - If the camera gets too close, the image plane is secretly moved backward and scaled up accordingly to give the illusion of being able to get infinitely closer.
  - As the camera gets closer, the field of view is blurred, emulating real eyes or cameras.
- A minimap acts as an overview diagram in the corner of the screen to support the user with overall orientation.

The tasks for the large-scale user study have been reformulated to bring across the intended meaning more precisely.

5 User Study

5.1 Introduction

After three iterations of pilot studies, the prototype and testing procedure were refined enough to start with the large-scale user study. The following four hypotheses are set for the user study:

H1 The World Space method is preferred because it is more intuitive to use and closer to reality.

H2 The Screen Space method is preferred because it is an established and familiar norm.

H3 Members of older generations prefer the World Space method, while members of younger generations prefer the Screen Space method as they grew up with it.

H4 The World Space method exceeds in hedonic qualities (e.g., fun, engagement), while the Screen Space method exceeds in pragmatic quality aspects (e.g., efficiency, precision).

5.2 Methods

5.2.1 Introduction and Consent

The user study is designed as a within-subject study, meaning each participant is exposed to both visualization methods. Due to the broad demographic range of test participants, this design is ideal to exclude any bias that could result from different participants comparing only one method each. However, when exposed to two iterations of the user study, there might be some carry-over effects like learning or fatigue, which can influence the evaluation of the second method. To counterbalance those effects, permutations of the test iterations are required. [26] So, besides the two visualization methods, there are two different diagrams and sets of corresponding tasks. The order in which the methods are evaluated and which diagrams are used is randomized, resulting in four permutations of test iterations.

The user study starts with the test supervisor briefly introducing the scientific investigation, the goals of the study, what is expected of the test participant, and what data will be collected. Once the participant is clear on the procedure, they fill out the form of consent, which is provided by the Department of Augmented Reality for the purpose of user studies. Next, the test supervisor demonstrates the use case of the application by giving an overview of the two visualization methods. The method with which the test participant starts is then explained and shown in more detail, and the participant can try out the method themselves with a sample picture. The test participant can ask questions at any time if anything is unclear. If the test participant feels ready to start with the first task, the test supervisor initiates the test and logging of relevant data.

5.2.2 Diagram and Task

In order for the participants to get working experience with the two visualization methods, tasks were designed for both methods. For each method, there is one diagram and two corresponding tasks the participants need to answer. Two decision tree diagrams were chosen as they do not need specialized knowledge and allow for a broader demographic range of participants. Two online diagrams, DoLaundry[1] and SayHi[2], were used as inspiration to draw the diagrams used for the user study. While the refined DoLaundry (Figure 5.1a) had some decision nodes removed, several nodes were added to SayHi (Figure 5.1b) in order to equalize the two diagrams in terms of complexity. At the same time, possibly sensitive topics were removed from the diagrams to not risk offending any participants. Eventually, both diagrams consist of twelve decision nodes, twenty-four edges, and eleven and ten final states, respectively.

Furthermore, consistent fonts, sizes, and forms were applied to the two diagrams, as well as a uniform color scheme (Figure 5.2), which adheres to the recommended contrast ratio of at least 4.5:1 between text and background color, according to the latest Web Content Accessibility Guidelines (WCAG) 2.1 AA [24]. The color scheme was also ensured to be appropriate for people with any color-blindness disability. By significantly varying the colors in their contrast and brightness values, even a completely monochrome visualization would be distinguishable.

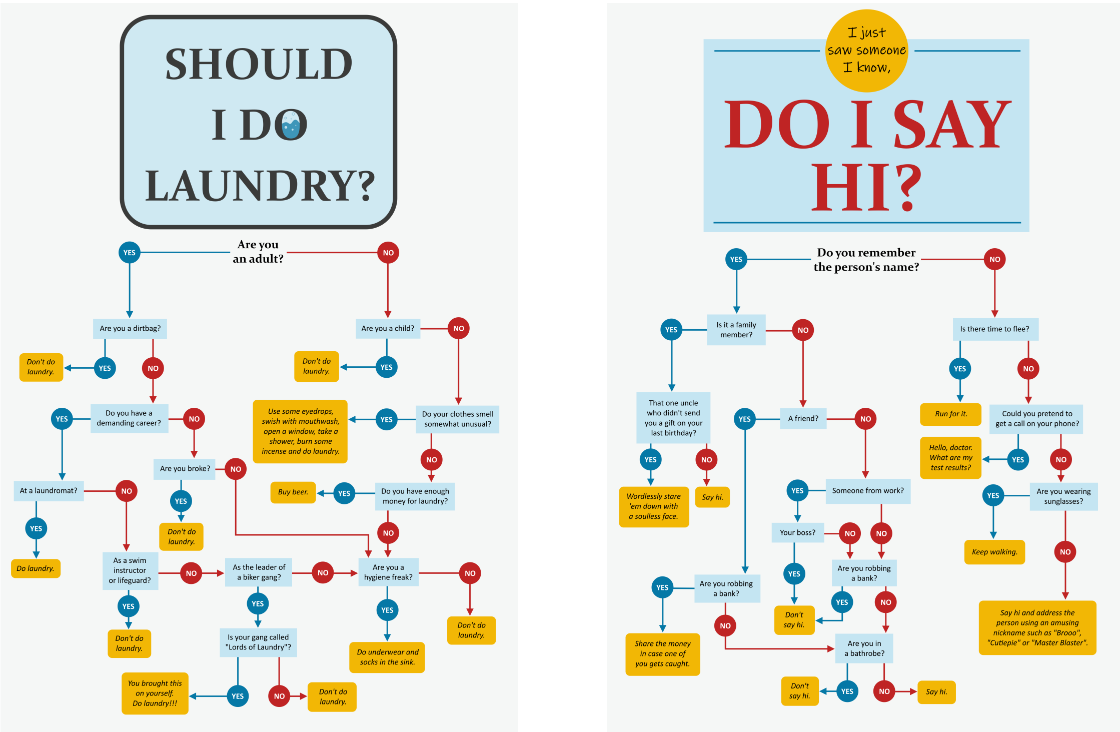
All these accessibility efforts were made to create an experience as uniform as possible for all test participants so that the study results will mostly depend on the two visualization methods without being influenced by too many external factors.

The sets of tasks for both diagrams were designed to be equal in structure (Figure 5.3). The first task asks the participants to count the number of final states that contain a specific phrase. This requires the participants to scan the whole diagram on a macro-level. The second task aims to operate more on a micro-scale of the diagram by asking for a path between two nodes. The previously acquired overview knowledge from the first task might help participants navigate the diagram for the second task more efficiently.

As the participants are not supposed to ask questions during the test, the tasks must be clear and precise. Both tasks provide a specific answer scheme with an example answer. As a result of the pilot study, the answer format for the second task was changed from text input to checkboxes, representing the sequence of decisions as a sequence

[1] https://www.pinterest.de/pin/409475791116140835/

[2] https://www.pinterest.de/pin/71002131601111582/



(a) The refined DoLaundry diagram

(b) The refined SayHi diagram

Figure 5.1: The two diagrams for the user study

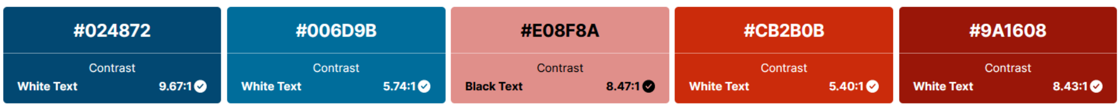


Figure 5.2: An accessible color scheme for the diagrams

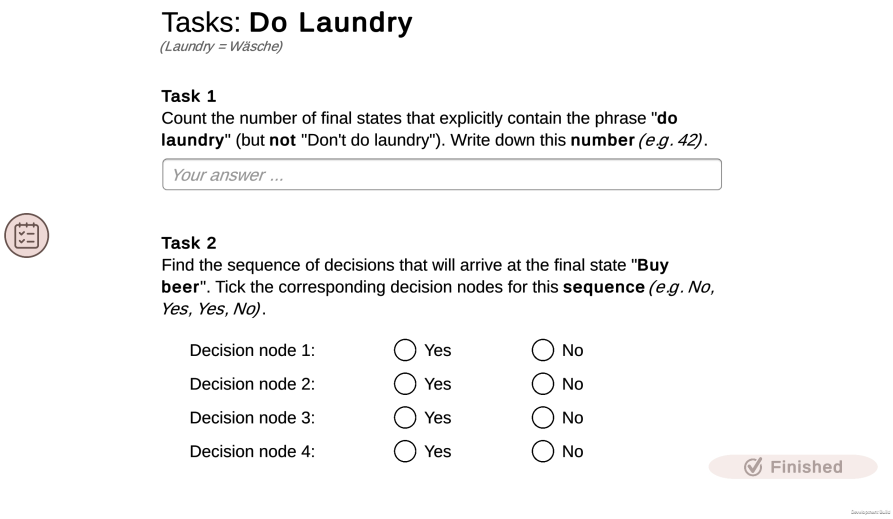


Figure 5.3: Task 1 and 2 for the DoLaundry diagram

of either yes or no (Figure 5.3), as previously – even though an example answer had been provided – pilot test participants were confused in which way to write down their answer.

The tasks can be viewed and answered directly in the application, meaning the test participants can not look at the diagram and the tasks simultaneously.

5.2.3 User Experience Questionnaire

After the test participant has answered the two tasks for the first diagram, they are asked to fill in an evaluation form for the currently used visualization method. The test supervisor emphasizes again that the participants should not evaluate the tasks or diagrams in any way but just the visualization method.

The chosen evaluation form is the User Experience Questionnaire (UEQ)[3] [27]. The UEQ consists of 26 items, each with a pair of expressions with opposite meanings. The UEQ was chosen over other traditional evaluation forms (e.g., SUS, NasaTLX) because it measures a variety of six evaluation scales: attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty. Thus, the UEQ measures not only pragmatic quality aspects (perspicuity, efficiency, dependability) but also overall attractiveness and hedonic quality aspects (stimulation, novelty), which are not explicitly goal-directed but listed as ergonomic requirements of a product in ISO 9241-11 [28]. This norm states that the usability of a product depends on effectiveness and efficiency (the pragmatic qualities) and satisfaction (the hedonic quality). Considering the importance of user engagement in this study's hypotheses, including hedonic quality aspects in the

[3] `https://www.ueq-online.org/`

evaluation form is mandatory.

Each test participant can choose whether to fill in the UEQ in German or English, as pilot test participants reported they did not understand some of the subtleties of the English version. The expression pairs consist of very precise terminology and need to be understood and evaluated spontaneously with little rationale to achieve the clearest results. Due to a broad range of demographics, some test participants have insufficient English skills to adequately evaluate the UEQ in English. Currently, the UEQ is available in 21 languages, and all certified translations were tested to be equivalent and produce comparable results [23].

5.2.4 Second Test Iteration

After the evaluation form has been filled out, this whole process is repeated for the second visualization method. The test supervisor explains the functionality and features of the second method and allows the test participant to test the method with a sample picture and ask questions, if necessary. For the test, the participant receives a new diagram and two tasks, but it is unavoidable that they are now more familiar with the general structure of the tasks.

Finally, the second visualization method is evaluated with another UEQ. It is important not to hand out both UEQs bundled at the end of the study, even though this might make it easier to compare the two methods. However, the UEQ is designed to be answered spontaneously and immediately after the product test to capture the user's subjective perception [27]. Retrospective explicit evaluation, which requires a deeper rational analysis by the user, is not always reliable [29] as the test participants usually are not experts (which is the point of this larger-scale user study) and often do not have the expertise to explicitly evaluate the product in significant detail. This is why the evaluation form should not give the test participants the opportunity to directly compare the two methods. This opinionated comparison is instead inquired about in an informal interview.

5.2.5 Demographic

After both methods have been tested and evaluated, the test participant is asked to fill in a demographic questionnaire. It was decided against acquiring this information at the start of the user study so as not to possibly bore test participants and confront them with too many topic-unrelated questions before the actual study has even begun.

The questionnaire inquires about basic information like age, gender identity, educational background, and English language skills, and also about previous experiences involving applications using AR or traditional screen interaction (e.g., mobile games) to determine if any correlations can be derived. The height of the participant's shoulder above the table and the height of the table above the ground are measured as well.

5.2.6 Interview

At the end of the study an informal oral interview is held with the test participant. The test supervisor asks four open questions, giving the participants the chance to talk more openly about their experience with the application:

1. Which of the two methods do you prefer?
2. What aspects of the two methods did you really like?
3. What aspects of the two methods did you not like or are in need of improvement?
4. Do you have any other feedback or suggestions for improvement?

While signing the form of consent at the beginning of the study, participants could voluntarily agree to the final interview's audio being recorded. This recording would then only be used for a one-time transcript and deleted afterward.

5.3 Results

5.3.1 Overview

A total of 16 participants were tested, four for each within-subject permutation. The device used for the study was the Samsung Galaxy Tab S8 11". Most user tests were conducted at the Department of Augmented Reality in Garching (Figure 5.4). The participants ranged from students to employees, with and without informatics background; thus, it would have been impractical for some participants to come to Garching. They were tested in other locations, either private homes or work environments. Every location featured a table of 70 centimeters in height, enough freedom of movement, and a well-lit environment. The exception is three participants who were tested at a table of 77 centimeters. However, the evaluation of the test results did not indicate any correlations between the results and the table height.

5.3.2 Demographic

The demographic data of the test participants at the time of the user study is as follows:

- The average age of test participants is 34, ranging from 20 to 64 years. 56 % of test participants are between 20 and 23 years old, 13 % between 30 and 31, and 31 % between 53 and 64.
- 25 % of test participants identify most with a female gender identity, 69 % with a male gender identity, and 6 % prefer not to specify.
- The highest degree of education 56 % of test participants have completed is a high school degree or equivalent, 25 % have completed a bachelor's degree, and 19 % a diploma's degree.

Figure 5.4: The main testing location at the Department of Augmented Reality in Garching

- 38 % of test participants rated their English skills as basic or conversational and chose to evaluate the UEQ in German. 62 % of participants rated their English skills as fluent or proficient and chose to evaluate the UEQ in English.
- 37 % of test participants have had no prior experience with AR technology, 44 % reportedly have used AR three times a month or less in the last six months, and 19 % use AR more than three times a week.
- 56 % of test participants do not play games on mobile devices, and 46 % play games on mobile devices at least weekly.
- The average shoulder height above ground is 68.5 centimeters, ranging from 53 to 79 centimeters. 6 % of test participants have a shoulder height between 50 and 59 centimeters, 50 % between 60 and 69 centimeters, and 44 % between 70 and 79 centimeters.
- 56 % of test participants are students, and 44 % are employees.
- 75 % of test participants have a background in informatics.

5.3.3 Observations by the Test Supervisor

During the tests, 38 % of test participants required at least some translation support from the test supervisor regarding the English language. Support was mainly necessary to initially understand the tasks they were given. While navigating the diagrams and performing the tasks, substantially less language support was needed. Those participants all rated their English skills as basic or conversational.

75 % of test participants required at least some assistance from the test supervisor to understand the first task. As this was the overwhelming majority of participants, no clear correlations with answers from the demographic survey could be found. 44 % of test participants required at least some assistance to understand the second task. There is an exact overlap with participants who are employees. As the structure of the two tasks is similar for both diagrams, no participant required additional assistance to understand the tasks in the second test iteration.

5.3.4 Quantifiable Observations – Logged Data

Task Completion

In the first test iteration, 56 % of test participants answered the first task correctly, 25 % made a minor mistake, and 19 % gave completely wrong answers. Of those who answered correctly, 78 % required assistance from the test supervisor to understand the question. Of those who did not answer correctly, 71 % had received assistance from the test supervisor, albeit without a positive impact. No participant required assistance in the second test iteration, but the participants who had made mistakes in the first iteration made them again in the second iteration. No clear correlations with answers from the demographic survey could be found. No participant was aware of their mistakes.

All participants in both test iterations answered the second task correctly.

For each test iteration, the task completion time was measured. There was a separate timer for each of the two tasks, which was further divided into working time (the participant inspected the diagram) and reading time (the participant inspected the tasks). All time measurements were rounded to two decimal places. P-values from the paired t-tests were rounded to four decimal places.

The mean working time for task 1 in seconds is 106.57, ranging from 13.9 to 241.35. The mean working time for task 2 in seconds is 32.06, ranging from 6.7 to 138.65. The working time distribution of both tasks (Figure 5.5) clearly shows that participants completed the second task substantially faster. A paired t-test is performed to check if the values differ significantly. The resulting p-value is 0, which we reject at alpha 0.05, thus providing strong evidence that the first task generally takes longer to solve.

Next, the working times for each visualization method are compared. The mean working time for task 1 (Figure 5.6a) is 98.91 seconds with Screen Space and 114.24 seconds with World Space. The paired t-test provides a p-value of 0.2947; thus, no significant differences can be derived. However, the data can be refined. Most participants

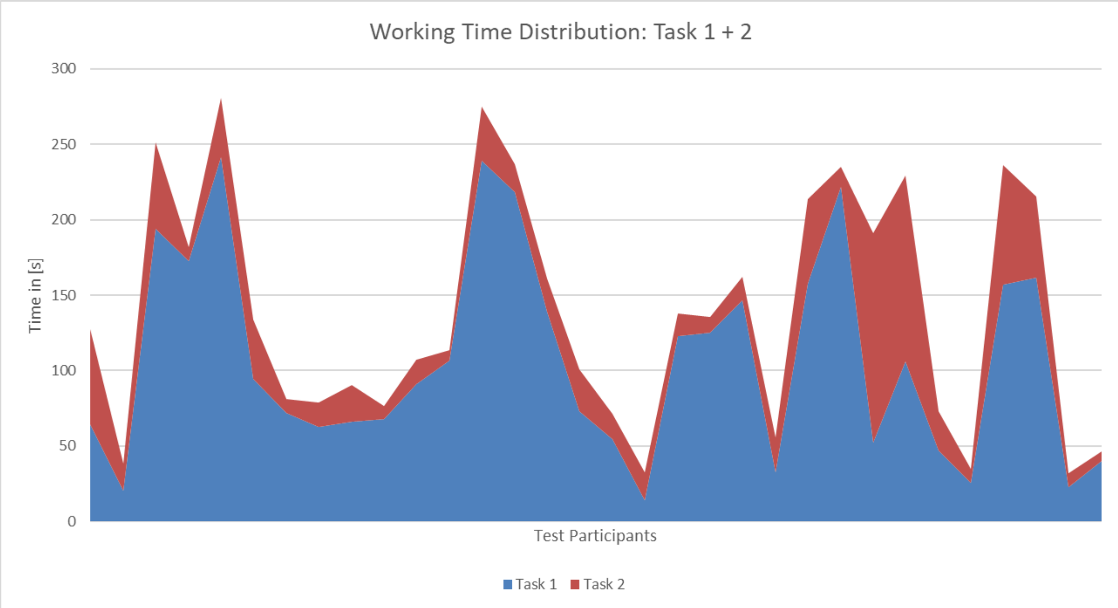
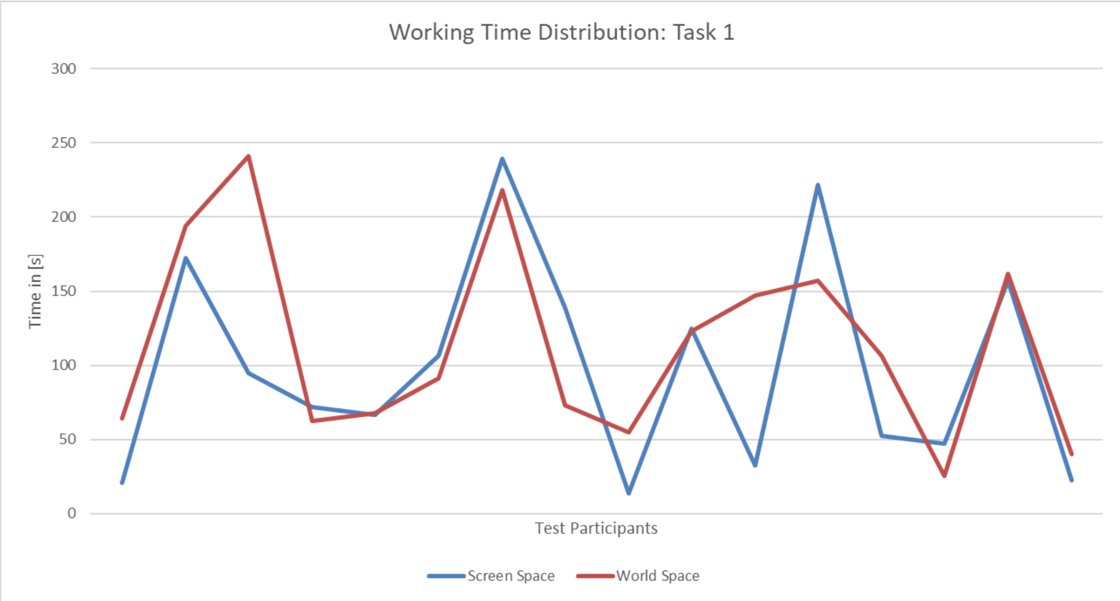


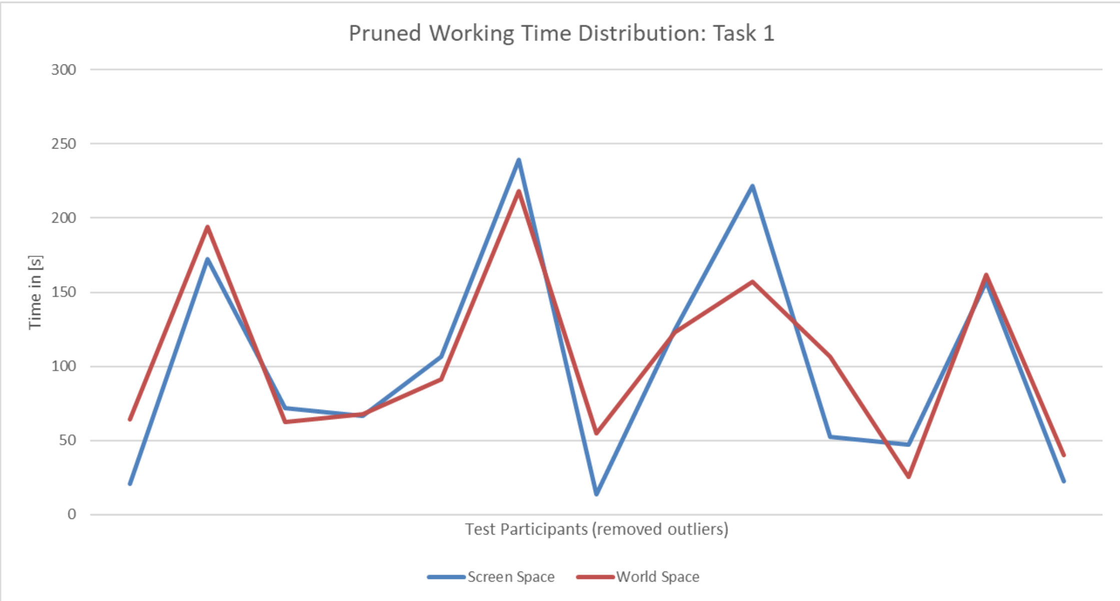
Figure 5.5: Working time distribution

needed relatively the same amount of time with each method, but there are five outliers (basic English skills, no students) with considerable time discrepancies between the two methods. Two of those needed substantially more time with their second method, and the observations by the test supervisor did not reveal any suspicious behavior. The other three participants, however, needed substantially more time with their first method due to requiring a lot of translation support and further explanations from the test supervisor to understand the task. Thus, these large time discrepancies can not be ascribed to the visualization method. If these three participants were pruned from the data set (Figure 5.6b), this would result in a mean working time of 101.24 seconds with Screen Space, and 105.14 seconds with World Space and a p-value of 0.6708. The difference has become even less significant, presenting almost equal working times for both methods.

For task 2, the mean working time (Figure 5.7a) is 33.21 seconds with Screen Space and 30.91 seconds with World Space. The paired t-test provides a p-value of 0.6961; thus, no significant differences can be derived. Again, there are some outliers with more considerable time discrepancies between the two methods. However, there is no overlap with the outliers from task 1. Notably, most of the outliers from task 2 are students with fluent or proficient English skills. Two of them needed substantially more time with their first method and struggled to understand the task for no apparent reason. When asked afterward by the test supervisor, they said they were tired and lacked concentration and that it had nothing to do with the respective visualization method. Observations by the test supervisor revealed that the third outlier double-checked their answers to task 1 several times while working on task 2, thus requiring a lot more time, which was



(a) Normal data set



(b) Pruned data set; removed outliers

Figure 5.6: Working time distribution for task 1 with both visualization methods

eventually not allotted to completing task 2. Thus, these large time discrepancies can not be ascribed to the visualization method. If these three participants were pruned from the data set (Figure 5.7b), this would result in a mean working time of 34.07 seconds with Screen Space, and 28.14 seconds with World Space and a p-value of 0.079. With a rejection threshold of alpha 0.05, this is surprisingly close to a significant difference in favor of the World Space method.

Upon further analysis, no significant correlations between responses in the demographic survey and the task completion times could be found. Tendencies indicate that students, on average, completed the tasks faster, and the 19 % of participants who use AR more than three times a week had the overall smallest task completion times. Most of the participants with the slowest completion times are not students. Notably, the participant with the fastest working time in World Space and the fifth fastest working time in Screen Space is no student and has basic English language skills with no background in informatics or experience with AR technology.

No correlations could be found regarding the participant's age, gender, English language skills, experiences with mobile games, shoulder height above the table, or whether they had a background in informatics. No significant differences between the DoLaundry and SayHi diagrams could be found.

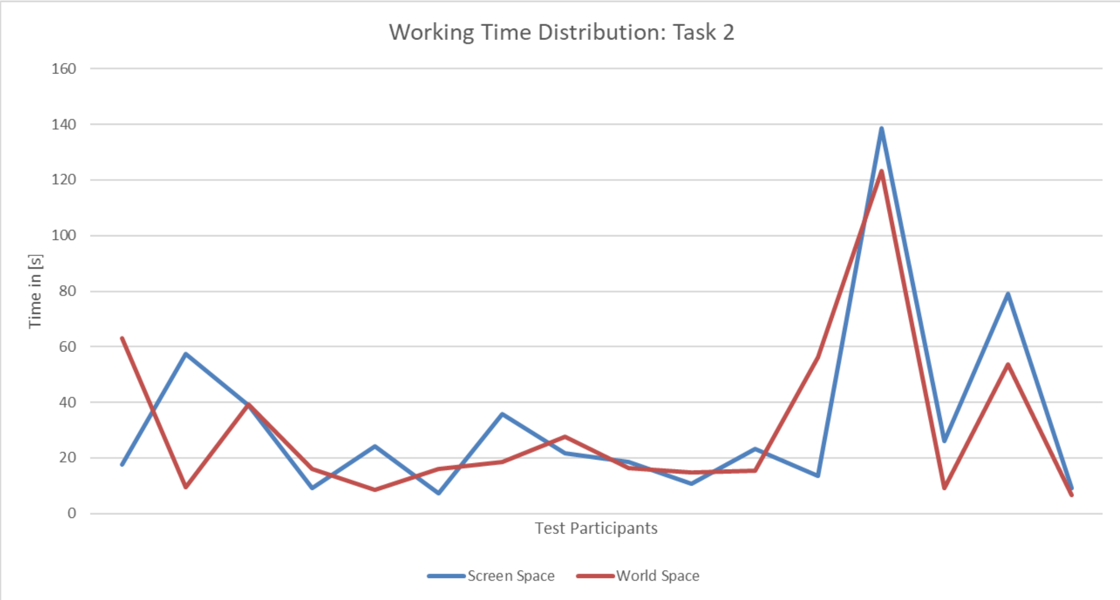
The necessary reading time to understand the tasks significantly decreased in the second test iteration. In the first test iteration, the mean reading time in seconds for task 1 (Figure 5.8a) was 29.78, ranging from 7.9 to 74.76. The mean reading time in seconds for task 1 in the second testing iteration was 13, ranging from 4.33 to 34.3. The paired t-test provides a p-value of 0.

In the first test iteration, the mean reading time in seconds for task 2 (Figure 5.8b) was 34.89, ranging from 2.57 to 66.86. The mean reading time in seconds for task 2 in the second test iteration was 15.77, ranging from 1.05 to 40.37. The paired t-test provides a p-value of 0.0001. Thus, strong evidence is given that the participants needed significantly less time to understand both tasks in the second test iteration.

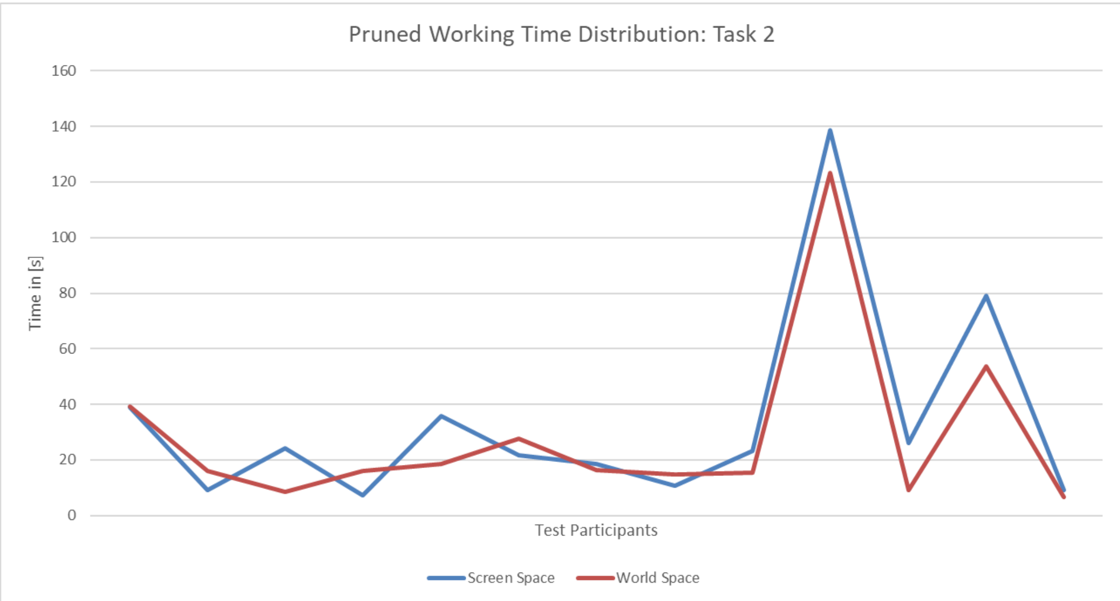
Other Logged Data

For each timestep of one second, four additional values have been logged during the user test. While using the Screen Space method, the scale factor of the diagram was tracked. While using the World Space method, the distance to the diagram, pitch rotation of the diagram, and rotation of the diagram relative to the camera were tracked. The data was normalized for each participant and then accumulated to show the relative distribution among all participants. The presented mean values and standard deviations were rounded to two decimal places.

- The mean value of the scale factor is 2.55 and the standard deviation 0.86 (Figure 5.9a).
- The mean value of the distance in centimeters is 60.54 and the standard deviation 18.44 (Figure 5.9b).

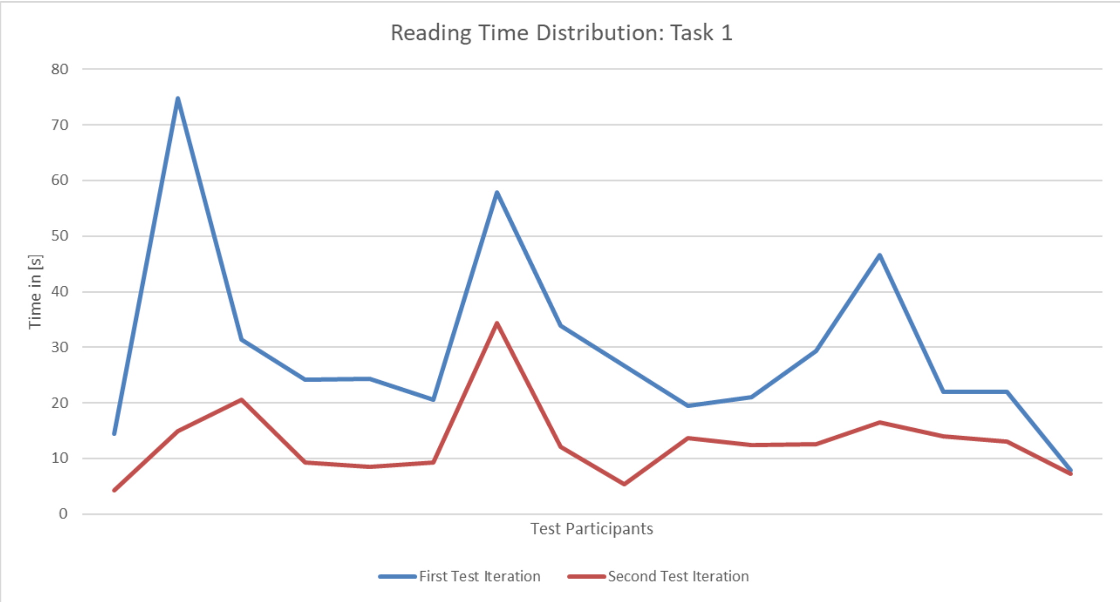


(a) Normal data set

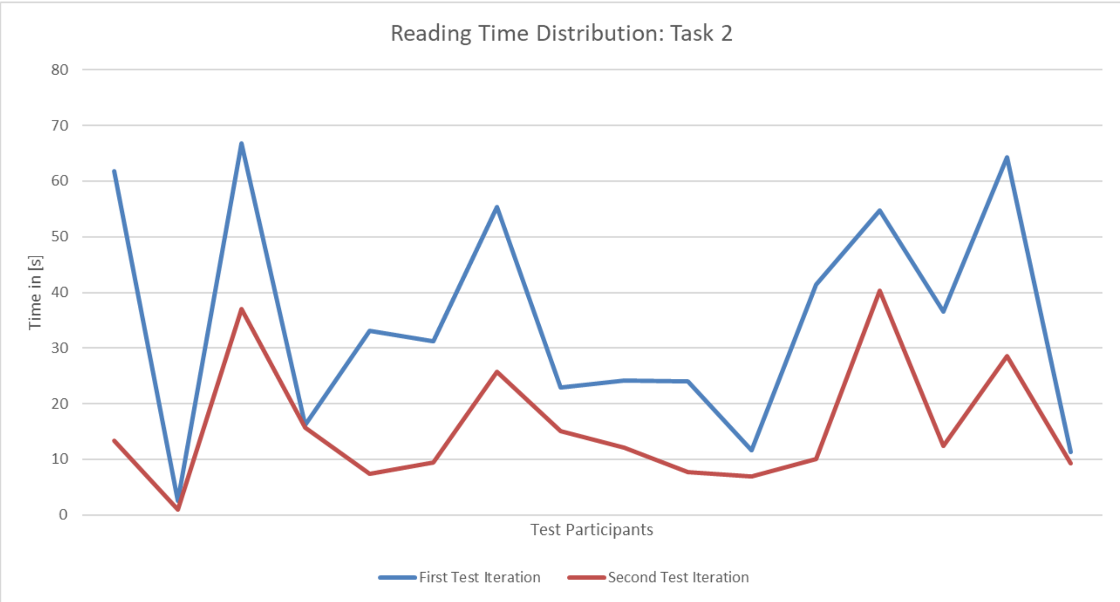


(b) Pruned data set; removed outliers

Figure 5.7: Working time distribution for task 2 with both visualization methods

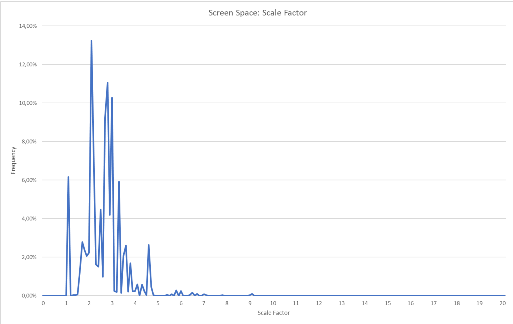


(a) Task 1

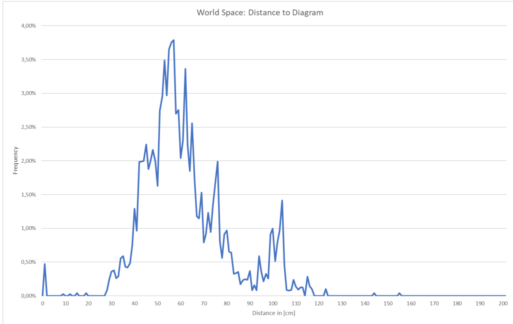


(b) Task 2

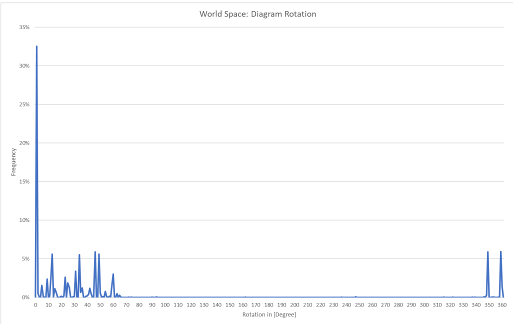
Figure 5.8: Reading time distribution for first and second testing round



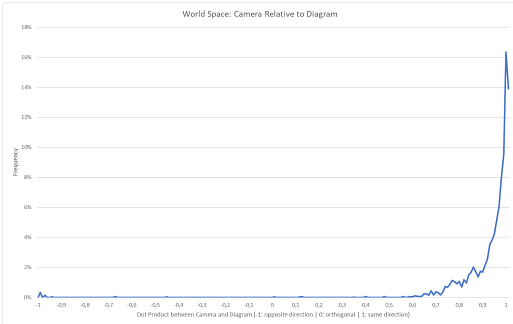
(a) Scale factor in Screen Space



(b) Distance in World Space



(c) Pitch rotation in World Space



(d) Relative rotation in World Space

Figure 5.9: Frequency distribution of logged data

- The mean value of the pitch rotation in degrees is 20 and the standard deviation 19.26. There is a substantial peak at 0 degree rotation with a frequency of 32.53 %, indicating many participants have not rotated the diagram at all (Figure 5.9c).

- The rotation of the diagram relative to the camera was measured as the dot product between the two respective forward vectors[4]. The mean value of the dot product is 0.92 and the standard deviation 0.17. There is a substantial peak at 1 with a frequency of 16.35 %, indicating many participants directly looked at the diagram (Figure 5.9d).

No significant differences between the DoLaundry and SayHi diagrams could be found.

[4]A value of 1 indicates the two vectors point in the same direction, a value of 0 indicates they are orthogonal, and a value of -1 indicates they point in opposite directions

5.3.5 User Experience Questionnaire

General Results

The UEQ consists of 26 items, each with a pair of expressions with opposite meanings. The resulting values for each item range from negative three (corresponding to the term with a negative connotation) to positive three (corresponding to the term with a positive connotation). The mean values of each of the six evaluation scales are used to compare the results of Screen Space and World Space (Figure 5.10). All mean values were rounded to two decimal places. P-values from the paired t-tests were rounded to four decimal places.

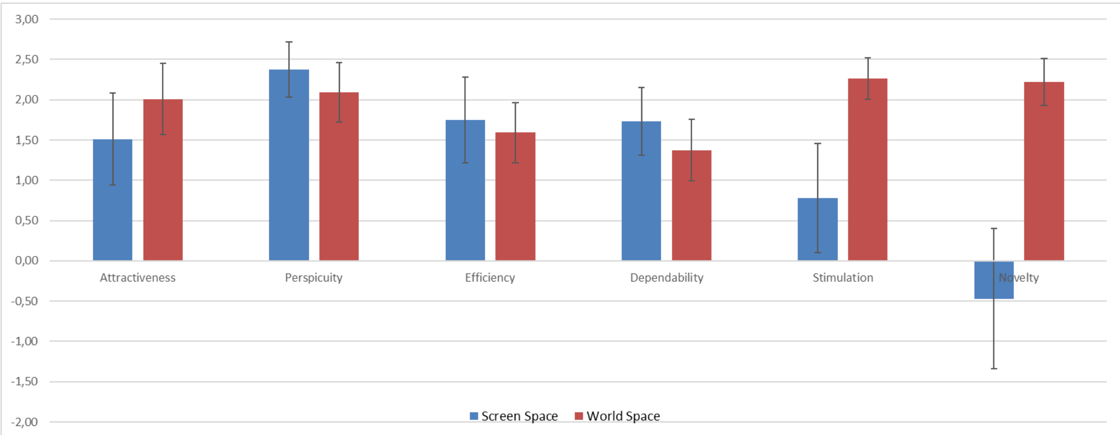


Figure 5.10: Comparison of the UEQ results

The comparison indicates that users overall liked the World Space method a bit more (mean value attractiveness Screen Space: 1.51, World Space: 2.01). The difference in the pragmatic quality aspects is not very substantial, but Screen Space takes the lead in all three categories (mean value perspicuity Screen Space: 2.38, World Space: 2.09, mean value efficiency Screen Space: 1.75, World Space: 1.59, mean value dependability Screen Space: 1.73, World Space: 1.38). The comparison of the hedonic quality aspects is much clearer, showing a considerable predominance of the World Space method (mean value stimulation Screen Space: 0.78, World Space: 2.27, mean value novelty Screen Space: -0.47, World Space: 2.22).

A paired t-test is performed to check if the six scale means of the two methods differ significantly. The resulting p-values for the scales of attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty are 0.1835, 0.2827, 0.6419, 0.2236, 0.0007, and 0, respectively. The p-value is rejected at alpha 0.05, providing strong evidence that only the hedonic quality scales of stimulation and novelty differ significantly. The other scale comparisons can only be viewed as tendencies.

Correlations with Demographic Survey

Upon further analysis, no significant correlations between responses in the demographic survey and the results of the UEQ could be found. Tendencies indicate that members of older generations generally rated the hedonic quality aspects of both visualization methods higher, while members of younger generations rated the pragmatic quality aspects higher. The same applies to English language skills; participants who reported they were rather less confident in their English skills and needed some translation help from the test supervisor tended to rate the hedonic quality aspects higher and the pragmatic quality aspects lower. There was a substantial overlap between those two demographic groups.

Participants with a background in informatics also rated the hedonic quality aspects lower. For the perspicuity scale, their mean value was higher, and their standard deviation was smaller. The same applies to students compared to employees (all students study an informatics-related course). Preexisting experience with AR technology did not make any difference in the mean values. However, participants who regularly used AR technology had a smaller standard deviation in the pragmatic quality scales.

No differences could be found regarding the participant's gender, experiences with mobile games, or shoulder height above the table.

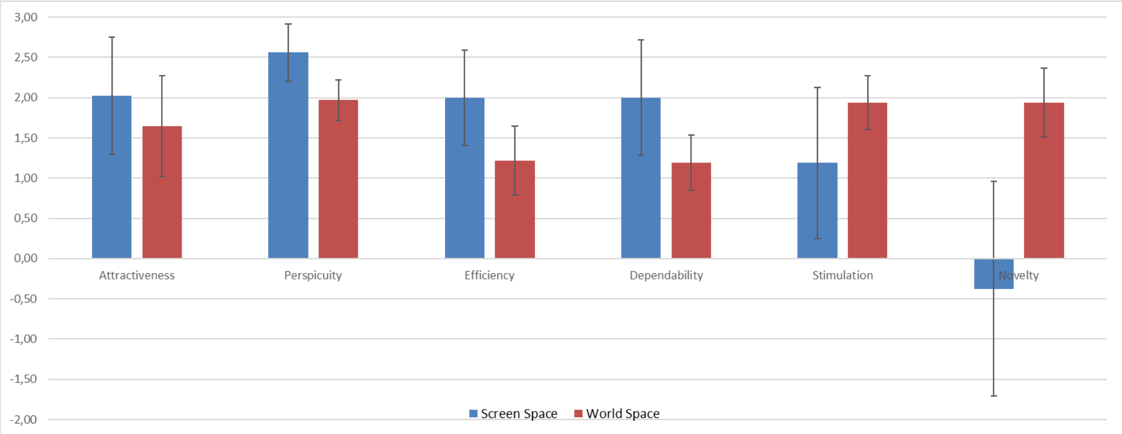
All of those tendencies are not significant; thus, no tables or specific values are given as they would not have any reliable meaningfulness.

Correlations with the Study's Execution

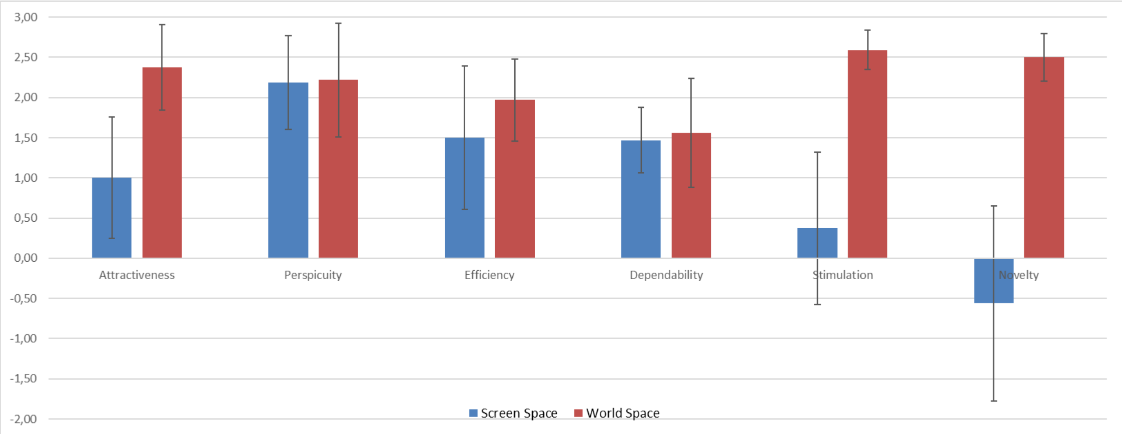
Whether the DoLaundry or the SayHi diagram was used for either visualization method did not reflect significantly in the results of the UEQ. Participants who were exposed to the World Space method first tended to rate the hedonic quality aspects of Screen Space a bit less benevolently.

A notable difference, however, occurred regarding the two sequential test iterations. Participants in their first test iteration (Figure 5.11a) tended to overall like the Screen Space method a bit more (mean value attractiveness Screen Space: 2.02, World Space: 1.65). The comparison of the pragmatic quality aspects is a lot more favored to the Screen Space method (mean value perspicuity Screen Space: 2.56, World Space: 1.97, mean value efficiency Screen Space: 2, World Space: 1.22, mean value dependability Screen Space: 2, World Space: 1.19). The difference in the hedonic quality aspects is less substantial, but World Space still takes the lead, albeit less clearly on the stimulation scale (mean value stimulation Screen Space: 1.19, World Space: 1.94, mean value novelty Screen Space: -0.38, World Space: 1.94). Those indications only show clear tendencies for the perspicuity and novelty scales.

In contrast, after having experienced both methods and being more familiar with the task format, participants in their second test iteration (Figure 5.11b) tended to clearly like the World Space method more (mean value attractiveness Screen Space: 1, World Space: 2.38). The values of the pragmatic quality aspects have nearly equalized, with World



(a) Results of the first test iteration



(b) Results of the second test iteration

Figure 5.11: UEQ sequential test iterations

Space actually taking a very close lead in all three categories (mean value perspicuity Screen Space: 2.19, World Space: 2.22, mean value efficiency Screen Space: 1.5, World Space: 1.97, mean value dependability Screen Space: 1.47, World Space: 1.56). The difference in the hedonic quality aspects has now clearly shifted to favor the World Space method (mean value stimulation Screen Space: 0.38, World Space: 2.59, mean value novelty Screen Space: -0.56, World Space: 2.5). Those indications only show clear tendencies for the attractiveness, stimulation, and novelty scales.

5.3.6 Interview

Even though the informal interview was structured with four questions, many participants answered them in such a way that the strengths of one method were listed again as weaknesses of the other method (rightfully so). This, however, would make a presentation of the results, organized by questions, unnecessarily bloated. Thus, the results are presented in a clustering, which somewhat resembles the six evaluation scales of the UEQ. Where meaningful, the explicit evaluations of the participants are justified (or refuted) with the implicit results of the UEQ and other logged data.

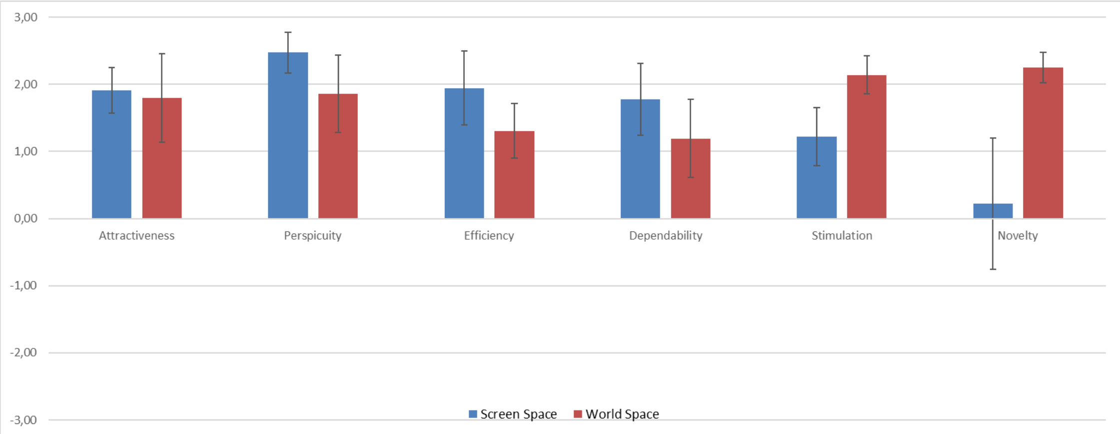
Attractiveness – Preferred Method

When asked explicitly which visualization method they prefer, 44 % of test participants preferred World Space, 50 % preferred Screen Space, and 6 % found both methods to have equally preferable qualities. The method to which each participant was exposed first does not show any influences in the explicit evaluation.

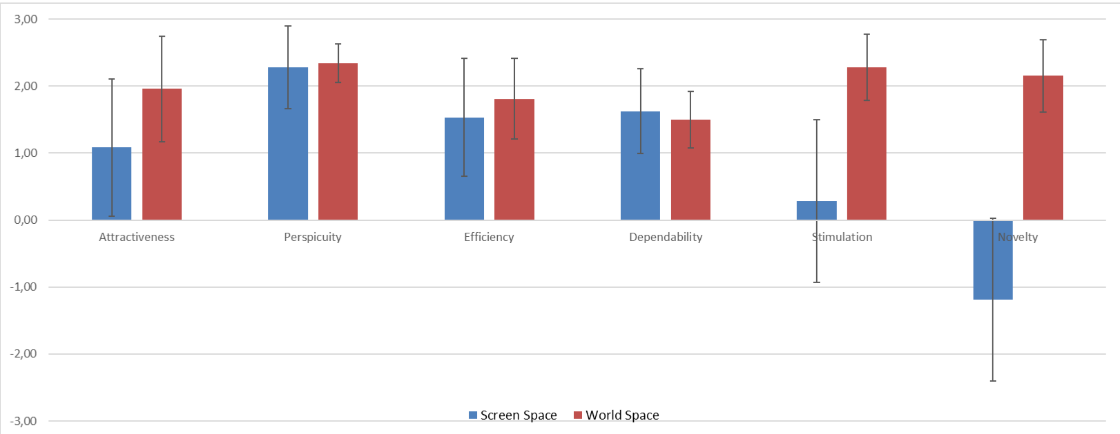
The results of the UEQ show tendencies that participants who explicitly preferred Screen Space also implicitly preferred Screen Space in four of the six UEQ scales: attractiveness, perspicuity, efficiency, and dependability (Figure 5.12a). People who explicitly preferred World Space seemed to – though not as clearly – implicitly prefer World Space in three scales: attractiveness, perspicuity, and efficiency. Their UEQ values for the dependability scale were almost identical (Figure 5.12b). All test participants unanimously preferred World Space over Screen Space in regard to the two hedonic quality scales: stimulation and novelty. However, the difference is significantly greater among participants who explicitly preferred World Space.

This match in implicit and explicit evaluation is a pleasant indicator that the evaluation of the whole user study is built on a plausible foundation.

Furthermore, there seems to be a correlation between the explicit preference and the task completion time. The working time distribution of task 1 (Figure 5.13a) clearly shows that participants needed less time while working with the visualization method they later explicitly preferred (mean time with preferred method: 88.07 seconds, mean time with non-preferred method: 125.07 seconds, p-value of paired t-test: 0.0046). However, no such clear indications can be derived from the working time distribution of task 2 (Figure 5.13b), which vaguely suggests that participants actually needed more time while working with the visualization method they later explicitly preferred (mean

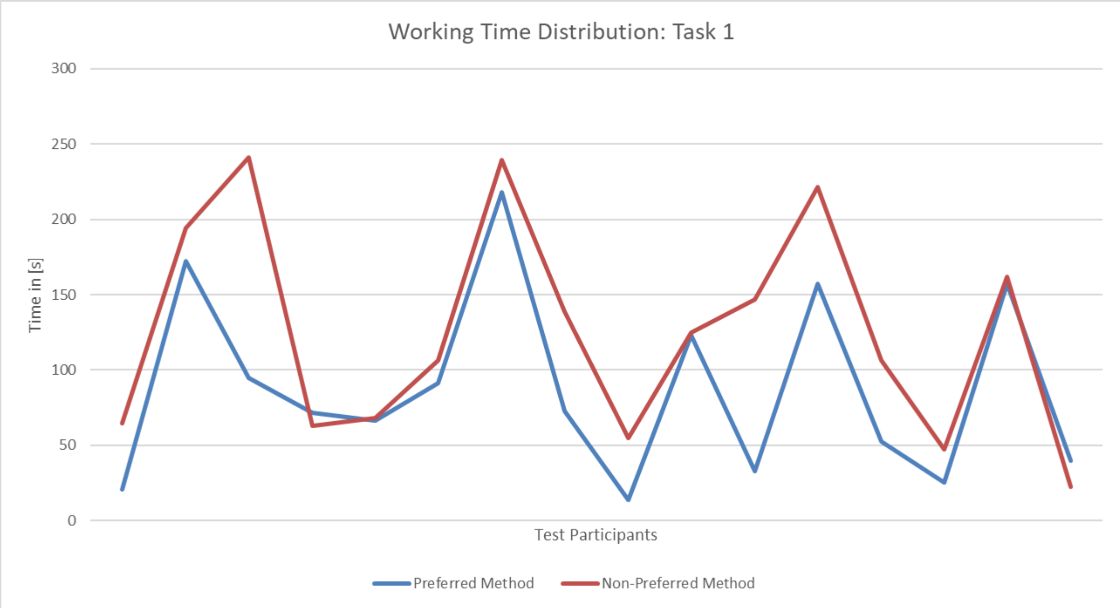


(a) Participants who explicitly preferred Screen Space

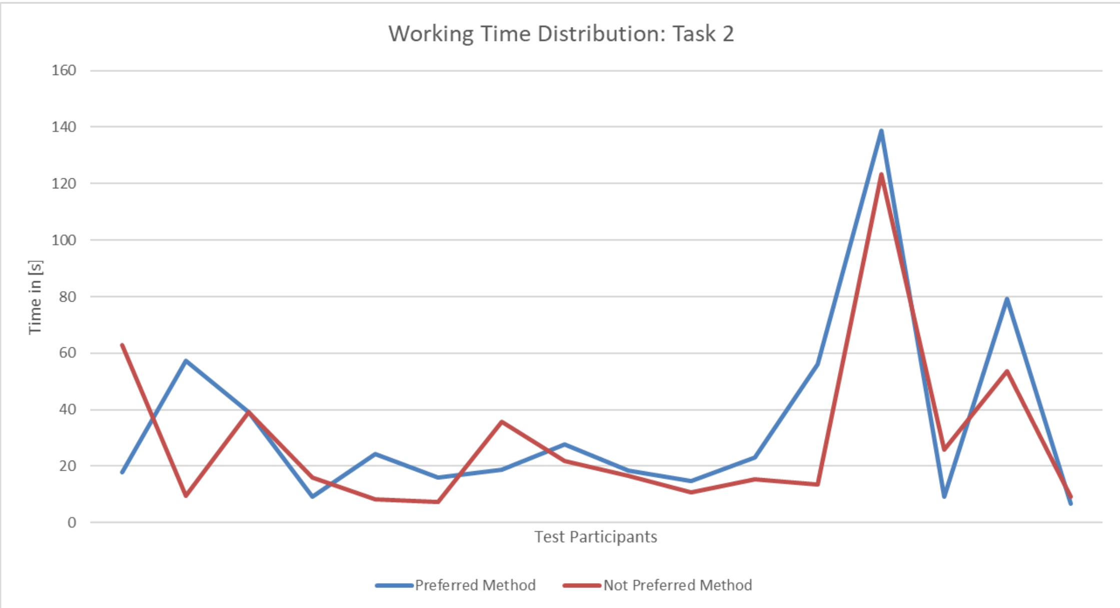


(b) Participants who explicitly preferred World Space

Figure 5.12: UEQ preferred method



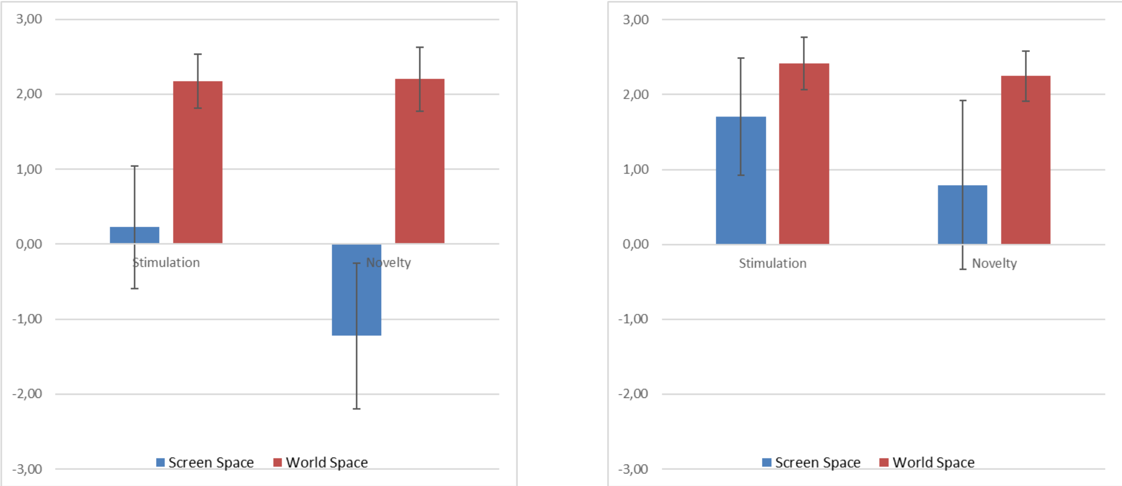
(a) Task 1



(b) Task 2

Figure 5.13: Working time distribution with explicitly preferred method

time with preferred method: 34.78 seconds, mean time with non-preferred method: 29.34 seconds, p-value of paired t-test: 0.3498).



(a) Participants who found World Space motivating

(b) Participants who did not find World Space motivating

Figure 5.14: UEQ hedonic qualities

Hedonic Qualities – Stimulation, Novelty

More than half of the test participants stated that the World Space method was more motivating, fun, and activating, ensuring excitement and attention through varied interaction. Even though some of those participants explicitly preferred Screen Space, they were open to the advantages of World Space. The participants who did not mention a motivational aspect of the World Space method exclusively preferred Screen Space and voiced the harshest rejection against World Space. When asked as to why by the test supervisor, they replied they either would not want to put in the effort to learn a new method, or they felt so highly proficient with the Screen Space method that they feared a new method would only throw them back in terms of efficiency. This sentiment is also reflected in the results of the UEQ, where the participants who were not motivated by the World Space method evaluated the Screen Space method much more benevolently and almost equally in stimulation (Figure 5.14).

Due to the motivational aspect, one participant suggested the World Space method might be helpful in school education as children would perhaps learn better if the learning method evoked more emotions in them. Many participants also mentioned how they found the World Space method more immersive; they felt there was a greater sense of feedback that they were actually doing and accomplishing something. Some assumed this sensation was caused by more informed and increased engagement due to the involvement of the vestibular body system providing additional information and navigation cues.

In contrast, many test participants reported low engagement with the Screen Space method – interestingly, this sentiment was voiced predominantly by participants who explicitly preferred Screen Space. They said the method was boring, put their brain into "sleep mode", and they felt they were working less efficiently because the method was too familiar and banal. However, this self-proclaimed reduced efficiency was not coherently reflected in the task completion times.

Perspicuity

Roughly half of the test participants judged the World Space method more intuitive because it is more natural and easy to learn (particularly for spatial representations, e.g., paths). This advantage, however, is not reflected in the results of the UEQ nor the task completion times. Two of those participants additionally deemed the Screen Space method too artificial, unnatural, and non-intuitive. They assumed the method is only popular due to its ubiquity and that it would be more challenging to learn for people who have no experience with it (e.g., people of very young or old age). Both those test participants are students with informatics background, young, and have reportedly had little experience with AR technology.

Many other test participants presumed the brain would have to multitask less with the World Space method if the corresponding task involved moving and interacting with the environment due to a similar situational context. Some of those participants additionally presumed the Screen Space method would be more challenging to use if the corresponding task required moving due to a mismatching situational context.

Three-quarters of the test participants – among them, everyone who explicitly preferred Screen Space – stated that the biggest strength of the Screen Space method is its ubiquity. The method is firmly established in everyday life and well-known to a majority of people. Due to this high degree of familiarity, the method can easily be used efficiently without requiring any particular practice or skill set.

Half of those participants additionally stated that a significant weakness of World Space – and AR in general – is its unfamiliarity. The technology is still relatively new and not widely used, and many people lack the necessary practice to work efficiently with this method. Some participants would have wished for more training with the World Space method (beyond the scope of the user study) and assumed they might evaluate the two methods differently if they had been equally proficient with both of them.

Most participants – in equal parts preferring either World or Screen Space – believed the World Space method to be clearly superior if people had more practice with it, grew up with using this method, or did not have any prior knowledge about any of the two methods.

Pragmatic Workflow – Efficiency, Dependability

One test participant said that the World Space method offers more freedom to view the diagram without restrictions and from any angle or size. Two other participants who explicitly preferred World Space wished for even more degrees of freedom regarding the rotation via swiping. Additionally to the existing pitch rotation, they also wanted yaw rotation to be possible. However, they proposed that only one degree of freedom should be available at a time, and the diagram should be rotated exclusively either yaw or pitch, depending on which dimension of the swiping input vector was greater. This approach should ensure freedom without making it chaotic. This additional freedom, or more the lack thereof, seemed to be a key issue, as many test participants stated location dependency to be a significant weakness of the World Space method. The method requires freedom of movement and the user's willingness to move, but at the same time, the necessity of a marker limits the freedom of movement and location.

Interestingly, two other participants who preferred World Space and reportedly use AR technology weekly assessed the existing pitch rotation as redundant with no use for mere 2D image planes in an immersive 3D scene. They feared such a feature might rather confuse users by giving them unnecessary degrees of freedom. The only applicable use case would be to counteract suboptimal marker placements, which, however, should also be avoidable with enough precaution. They concluded that the swiping interaction would be more appropriate for visualizing actual three-dimensional objects.

Another unwanted interaction feature was the `LookAtMe` button. Two participants directly said it was an unnecessary feature, but, in essence, no participant used the `LookAtMe` button during the user test. Instead, some participants suggested a button to reposition the image plane directly in front of the camera, wherever it is currently facing. As this feature would not only change the rotation of the image plane but also its position, this might solve the location-dependency problem.

However, such a feature could only be realized by relinquishing the marker, which is used to stabilize the tracking. Removing it would only increase another weakness of the World Space method that many test participants – all of them preferred Screen Space – perceived even with the marker: unsteady tracking. A number of factors, for example, poor lighting conditions, an environment that is too uniform, a weak camera, or health problems (such as shaky hands), can lead to an unsatisfying tracking experience. Two of those participants wished for some kind of interpolation to make the AR tracking smoother. Even then, one of those participants noted that the Screen Space method would be more precise.

Roughly half of the test participants remarked on the spatial flexibility of the Screen Space method. The user is not tied to sufficient space or lighting conditions and can work while resting and sitting. The device can also be put down to pause work and later resume it without losing the visual state. Regarding the loss of the visual state, one of those participants stated a hurdle for collaboration among multiple people on a single device with the World Space method; two people have different lines of sight on the same display. When one user hands the device to another person, the area that is visible on the screen changes.

To accommodate for this, several features were suggested for the World Space method. A few participants thought World Space would offer a good collaboration opportunity by having an entire team view the same 3D immersive scene from multiple devices. Two participants had a similar idea and suggested a feature to draw notes and annotations on the image plane, enabling synchronous and asynchronous interaction with other people. There should be the possibility to upvote annotations and make virtual comments ("sticky notes") to collectively work in an immersive environment that furthers the users' engagement and motivation.

A different approach to counteract the location-dependency problem and loss of visual state was a screenshot suggestion by two participants. This screenshot preserves the current visual state and could be interacted with by moving and scaling, essentially acting as a combination of both methods. Another participant proposed a feature to save several viewports and later retrieve them again, restoring the visual state and offering a quick alternation between different working views. All those suggestions would have to rely on markerless tracking, returning to previously mentioned problems.

Some test participants also mentioned that the World Space method would work best with HMDs, freeing the user's hand. HMDs would be greatly suited for even more immersive collaboration in the same scene. However, other suggested features, such as the screenshot functionality, would not be feasible in this manner. Additionally, HMDs are not nearly as widespread and common as cellphones or tablets – even though this might change in the future, and AR technology might someday be ubiquitous in any everyday glasses. Generally, those participants agreed that the World Space method would not be suited for devices with smaller screens. By being able to more easily and permanently adjust the size of diagrams in Screen Space, this method is more applicable to a broader range of devices. However, three participants stated that – especially on smaller screens – the finger takes up a part of the field of vision while operating on the diagram, thus obscuring some information – a frustrating disadvantage that the World Space method does not struggle with.

Furthermore, two participants positively remarked on the mechanics for error robustness: the arrow in World Space and the `Center` button with undo functionality in Screen Space.

Navigation

Four test participants stated the biggest strength of the World Space method to be navigation on a micro-scale. Just small, subtle movements of the camera would suffice to quickly and effortlessly view proximity areas of the diagram and get a good low-level overview. However, roughly half the participants (among them the just mentioned along with some more who preferred the Screen Space method) stated the most frustrating weakness of the World Space method to be navigation on a macro-scale. Strong zooming in and out of the diagram requires moving the entire device back and forth, possibly taking several steps, which is substantially more effort than using the pinching gesture to quickly zoom in Screen Space. In return, the Screen Space method struggled to deliver satisfying micro-scale navigation, as pointed out by some of those participants, primarily ones who preferred World Space. They found swiping just for minimal position changes to be cumbersome and annoying, especially when needing to go back and forth several times.

Four of those participants made some suggestions for a combination of the two methods to take advantage of their respective strengths and minimize their weaknesses. The basis is the World Space method. Additionally, the user can quickly scale the image plane up and down with, for example, the pinching gesture, a slider, or buttons. The scale then gradually returns to its starting value, either automatically, with a reset button, or by using the pinching gesture again and having the scale magnetically snap to its starting value (and possibly to additional discrete values). One participant gave an example use case: a building plan as a 1:1 AR overlay over the real world to guide the user, with the ability to zoom the whole plan out by either of the previously mentioned operations and get an overview of the whole building complex. Afterward, the size of the plan returns to its original 1:1 scale.

Roughly half of all participants – in equal parts preferring either World or Screen Space – mentioned that due to its anchoring in the physical environment, the World Space method promotes easier spatial orientation and the classification of diagram parts in the overall context. Especially with complex diagrams, a stronger memory effect is created, which makes mental navigation easier. Observations by the test supervisor showed that several test participants in World Space – after attentively scanning the diagram for the first task – were able to navigate to the required nodes for the second task very purposefully and unerringly.

Fewer participants – only those who preferred World Space – felt that while using the Screen Space method, constantly swiping back and forth and zooming in and out without spatial context hindered the development of a supportive spatial understanding. One participant described it as knowing the approximate direction of where to go but not the distance, resulting in constantly needing to check if the destination node has been reached. While using the World Space method, in contrast, they had an exact feeling of where the destination node was located in space and could quickly and precisely maneuver to their target.

As mentioned in a previous section, two test participants felt the World Space method should not offer the interaction functionality of pitch rotating the image plane by swiping. They believe that if the user dynamically changes the diagram's position, rotation, or scaling, this could hinder the development of a supportive spatial understanding. However, a different test participant voiced a contrasting opinion that transforming the diagram in any way would not negatively affect the spatial awareness as the 2D image plane is likely to always be viewed from the front, providing a similar enough spatial context. For three-dimensional objects with high complexity, however, the participant felt that transformations could indeed negatively affect the spatial orientation and possibly even severely confuse the user. Those three test participants all preferred World Space and reportedly used AR technology weekly.

In a similar manner, some test participants felt like the World Space method did not make use of its full potential by just providing a 2D image plane. They especially thought that large diagrams could be better organized and grouped in a three-dimensional arrangement, reducing complexity and making even more use of the spatial awareness.

Three test participants who preferred World Space stated they would find the Screen Space method useful for small diagrams that require no moving around since they can benefit from the omnipresence of the visual state in Screen Space, and the lack of complexity would just make the navigational advantages of World Space redundant. They said, however, that Screen Space would not be useful for small areas of the same size, which are part of a larger diagram, due to the impeded spatial understanding of where the current area is located in the context of the whole.

The feedback in this section comes primarily from participants who regularly use AR technology (monthly to weekly) and have a background in informatics. They are members of younger and older generations in similar numbers.

Accessibility

Roughly half the test participants praised the more pleasant ergonomics of the World Space method. Using the method does not require the fingers, making it more accessible to people with limited fine motor skills and other disabilities. As the tablet can be held comfortably with both hands, the weight of the device is more evenly distributed, and less strain is put on the wrists. The method encourages working while standing or in motion, which provides healthy variety for the spine and neck. Some of those participants additionally mentioned how working with the Screen Space method to perform a task that requires standing or motion puts unnecessary strain on the user. The tablet can only be held with one hand as the other is needed to operate on the screen. Especially over a more extended period of time, this is too strenuous for the arms and wrists, particularly with heavy devices. Furthermore, the static way of looking down on the screen could more easily lead to neck issues. All of those participants have worked with both visualization methods while standing and regularly use AR technology (monthly to weekly). No predominance in age or the preferred method was discernible.

A contrasting opinion was voiced by another – a bit smaller – group of test participants. Notably, they all had no prior experience with AR technology and sat down while working with the Screen Space method. Most of them were female and explicitly preferred the Screen Space method. No predominance in age was discernible. They all negatively remarked on the constant movement required for the World Space method and the inability to work while sitting.

Even though those two groups of participants strongly supported these contrasting opinions, no such division is reflected in the results of the UEQ nor the task completion times.

Four test participants with glasses stated they liked that the Screen Space method offered users with glasses or impaired eyesight the opportunity to more easily and permanently adjust the zoom strength to their individual needs. Despite this, those participants either still preferred World Space or at least considered the method favorably.

One test participant remarked on the better hygienic aspects of the World Space method as its use does not require the user to touch the screen.

5.4 Discussion

Since neither visualization method is clearly preferred over the other, neither H1 nor H2 could be verified. However, the underlying causalities are reflected in the explicit interview opinions of many test participants. Even though no substantial influences on task completion time or implicit evaluation could be found, most participants viewed World Space as more intuitive and natural, and nearly all participants noted the firm establishment and ubiquity of Screen Space. This strong sense of familiarity the participants felt with the Screen Space method also indicates that the implementation of the Screen Space method as described in chapter 3 The Application is a faithful adaptation of the visualization method.

H3 proved to be a false assumption. There were participants of all ages who either thrived excitedly in the new World Space method or had greater difficulties accepting and adapting to new ways of working. Tendencies in the results of the UEQ indicate that generally, members of older generations rated the hedonic quality aspects of both visualization methods higher, while members of younger generations rated the pragmatic quality aspects higher. In many cases, the young age of participants coincides with fluent English skills and a background in informatics. Thus, it can be assumed that this group of participants, on average, is more universally proficient with many technologies and has experience with comparable products, explaining the slightly higher and more precise ratings for pragmatic quality aspects. The opposite group likely is, on average, less exposed to such a multitude and variety of novel technologies, resulting in higher ratings for the hedonic quality aspects.

H4 could be partially verified. The results of the UEQ show that World Space significantly dominates in the hedonic quality aspects. In the explicit interview, most test participants praised the method as more motivating, exciting, and immersive, and no negative aspects regarding novelty and stimulation were given. In contrast, no participant voiced any positives in those categories for the Screen Space method, with the only feedback being the boredom and banality of the method. Since the Screen Space method is ubiquitous in a multitude of everyday activities and World Space is a relatively novel and uncommon visualization method, these results are not necessarily surprising, but it is encouraging to have the method be received with acceptance, confirming the initial assumption with strong evidence.

The second assumption of H4 can not easily be verified nor refuted. Even though the results of the UEQ tend to lean more toward the Screen Space method in regard to pragmatic quality aspects, the difference is not significant. Certain refinements of the data set, in particular looking only at the evaluation of the second test iteration, even resulted in World Space slightly surpassing Screen Space in all three pragmatic quality aspects. In the second UEQ evaluation, the participants have experienced both methods, giving them the opportunity to compare the two methods. However, it is unclear whether test participants filled out their second evaluation form with deliberate comparison in mind, which would shift the evaluation to be less implicit than the first evaluation. As Nielsen said, an explicit evaluation by the user is not always reliable [29].

Though not significantly, the task completion times indicate that the first task was completed faster with the Screen Space method, while the World Space method resulted in faster working times for the second task. Those tendencies resemble the reported strengths and weaknesses of the World Space method. The first task requires constant zooming in and out to scan the whole diagram for specific nodes. To achieve such strong zooms with the World Space method, it is necessary to move the entire device back and forth, possibly taking several steps, which is substantially more effort than using the pinching gesture to quickly zoom in Screen Space. This inefficient navigation on a macro-level is a key issue in need of improvement if work on this visualization method should continue.

However, the working time of the second task shows promising results, especially since pruning out three outlying participants almost shifts the lead of the World Space method to become significant. The second task requires following a path between nodes, thus navigation on a micro-scale, which many participants reported to be more natural with the World Space method. The explicit evaluation and the working time for the second task support the assumption that this natural navigation by sensorimotor contingencies and the anchoring of the data in the physical environment enhance the creation of a "mental model", which – according to information visualization researcher Robert Spence [30] – is essential for efficiently navigating the visualization of complex data. A clear and reliable mental model provides vital cues for perceptual integration. Such perceptual cues facilitate wayfinding and navigation around complex diagrams by offering clear orientation, route monitoring, and destination recognition [9]. While this was positively remarked on for the World Space method, some test participants felt that constant zooming and moving without clear perceptual cues in the Screen Space method hindered the development of such supportive mental maps. It is unclear whether manual movement of the image plane in World Space (e.g., pitch/yaw rotation, `LookAtMe` button, option to reposition the image plane in front of the camera) hinders the development of a mental map.

Regarding the explicit evaluation of the pragmatic quality aspects, many test participants stated that the World Space method was very intuitive and natural to learn. In contrast, the Screen Space method offered more spatial flexibility and was well established and familiar, thus easy to use without practice by most people. Besides the spatial flexibility, the ubiquity of the Screen Space method was, for most participants, the only advantage they stated. There were a few participants who preferred Screen Space because they were substantially more proficient and efficient with it; however, most participants who preferred Screen Space did so because they were more used to this method. In a similar manner, some participants would have wished for more training with the World Space method and assumed they might evaluate the two methods differently if they had been equally proficient with both of them. This inevitable discrepancy in proficiency makes a clear comparison in terms of pragmatic quality aspects difficult. This user study aimed to acquire a broad range of feedback and impressions from a broad range of demographic groups. To draw more meaningful conclusions about H4 and which visualization method is more pragmatically efficient, a more narrow user study focusing on participants with high (or equally low) proficiency in both methods is recommended.

5.5 Limitations

In addition to the inevitable discrepancy in proficiency, there were some other limitations to the user study, as well.

First, the tasks invited the participants to work stationary. Thus, it is no surprise that no one used the `LookAtMe` button. In the given use case, corrective maintenance in chemical industries, the associated chemical plants could potentially extend over several floors and even buildings, requiring the user to move away from the image plane in World Space. In general, many common use cases might prompt movement, which was not tested in this user study. In such cases, the `LookAtMe` button might be seen as more useful, or perhaps the new functionality proposed by some participants to reposition the image plane in front of the camera becomes necessary.

Second, while using any form of Situated Analytics, special care needs to be taken to preserve the physical safety of users and bystanders. For example, a large image plane in World Space must not be placed in a way that obscures critical information, like a sudden staircase behind the image plane. No such considerations were taken into account for the user study, as all testing locations were small and sheltered and did not provide any risks to the users' physical safety.

Third, the testing locations were sufficiently lit. The AR marker was clearly visible to the camera at all times, and no participant lost tracking during the test. More common use case environments might induce unstable tracking or make using the World Space method entirely impossible. Alternative ways for visualization should be provided in such cases.

Fourth, while the size of the diagram in Screen Space can freely and permanently be adjusted to accommodate vastly differently sized diagrams, the size of the image plane in World Space is fixed. If the image plane is not adequately sized proportionate to the user and environment, this can substantially impact the visualization method's usability. For example, if the image plane is sized too large, there might not even be enough space in the physical environment for the user to take as many steps backward as necessary to be able to view the whole diagram. For the user study, the size of the image plane has been hardcoded to fit the testing locations. As a pilot study participant suggested, the user should be able to adjust the size of the image plane if needed. Alternatively, great care must be taken to use the visualization method only with image sizes that are compatible with the current environment.

Fifth, 62 % of test participants performed the tasks with both methods while standing, while 38 % sat down while working with the Screen Space method. This resulted in conflicting opinions about the two methods' ergonomic qualities and physical exhaustion, providing no clear conclusions. The given use case of corrective maintenance in chemical industries offers little opportunity to sit down, and also, for a meaningful comparison of the two methods, it would have been required to enforce using both visualization methods while standing.

Sixth, even though the working time was measured, the test participants were told to take the time they needed to complete the tasks and not to hasten. However, this resulted in many participants allotting time to non-task-related activities such as further exploring the visualization method (even though the opportunity to explore freely had been given beforehand) or careful reading of the complete diagram, which – while perhaps beneficial for some – was not necessary to solve the tasks correctly. The participant with the fastest working time in World Space had only basic English language skills and did not understand a majority of the diagram content. This resulted in them focusing solely on the tasks which they completed quickly and flawlessly. If a refinement of this user study was to be made, the diagrams should rather consist of abstract data and forms which do not invite a closer inspection. Alternatively, geospatial data such as maps could be used. Those would also offer advantages concerning the spatial orientation and development of a mental map, as the location of elements in geospatial maps has meaning, as opposed to, for example, P&IDs or decision tree diagrams.

Finally, 88 % of test participants either needed assistance from the test supervisor to correctly understand the first task in their first test iteration or did not answer the task correctly. After being familiar with the question format, in their second test iteration, no participant needed assistance from the supervisor, and the measured task reading time was substantially shorter. On average, participants who are students needed slightly less assistance, as well as reading and working time, probably because they are more familiar with such question schemes. Nonetheless, the overall performance of all test participants indicates that task 1 was unnecessarily difficult.

The first task for the DoLaundry diagram asks to 'count the number of final states that explicitly contain the phrase "do laundry" (but not "Don't do laundry").' Despite this, most test participants asked the supervisor whether they should count "Don't do laundry" as it contained "do laundry". Even then, some did count it. One final state consisted of five lines with "do laundry" in the last line. Most participants did not find this one. If a refinement of this user study was to be made, the positive and negative affirmations for "Do laundry" and "Don't do laundry" should be reformulated to not contain each other, such as "Do laundry" and "Leave laundry". Furthermore, each final state should consist of at most two lines as otherwise many participants might skip over it again. Any final state that contains "Do laundry" or "Leave laundry" should not contain any additional words. The SayHi diagram should be adjusted analogously.

It is unclear whether the difficulties with the task had any influence on the implicit and explicit evaluations. The results of the UEQ show quite substantial differences between the first and the second test iterations. However, this could also be a consequence of knowing both methods and being able to compare them against each other, which – if done with explicit intent – could very well have distorted subjective opinions. As the evaluation of both UEQs was deliberately not put sequentially at the end of the study, this direct comparison was not intended. Thus, a more meaningful refinement of the user study would be to let the test participants perform the tasks first on paper to get familiar with the question format. Afterward, the participant is exposed to only one of the two visualization methods. This between-subject design, however, would require more test participants to obtain significant results.

6 Conclusion

6.1 Outlook

The results of this user study are not clear enough to draw meaningful conclusions about which visualization method is better. The study acts more as a starting point to the problem statement, discovering which aspects work and which do not through several early pilot studies and a formative evaluation, disclosing the opinions of a broad demographic group. Moreover, while the results of the UEQ mostly allow no precise comparisons – partially due to the discussed limitations, partially due to the early stage in production – both visualization methods have received overall high scores of acceptance. The question of which method is better is a complex topic with many facets. Currently, both methods prevail in certain areas, and the usefulness of each method is highly dependent on the given use case and environment. Comparing more specific aspects of the two methods requires an improvement of the World Space method according to the feedback from this study and further focused user studies with an appropriate range and selection of demographic groups.

With immersive display technologies constantly improving, some identified problems of the World Space method, such as unsteady tracking, might be solved automatically in a few years. According to Mark Weiser's [20] vision of ubiquitous computing, combined with the possibilities of Situated Analytics and the rise of mixed reality technologies, it is highly likely that AR will be used for the presentation of data, even if it should be revealed that this is not the most suitable medium for this purpose. Therefore, continued research on these kinds of visualization techniques is essential to develop evidence-based guidelines on how to best integrate these technologies into the visualization process.

As a result of this user study, a linked 2D and 3D approach is recommended. World Space provides great overall orientation and navigation, while Screen Space offers more precise manipulation and spatial flexibility. A possible way to realize this combination could be to offer a way (e.g., a button) to switch dynamically between the two visualization methods. Switching from World Space to Screen Space projects the image plane from world coordinate space onto screen coordinate space and smoothly resets the rotation. The user can use all interaction methods of Screen Space until switching back to World Space, which seamlessly projects the image into the world scene. An alternative to a `SwitchMethod` button could be to have World Space be the default method and only temporarily change to Screen Space while the user performs the swiping or pinching gestures.

6.2 Further Research

The formative evaluation suggests the following areas of further research:

- According to the research of Makowski et al., there is an indication that presence and emotional engagement can positively influence recall and memory effect [31]. Research needs to be conducted on whether the increased engagement and immersion participants felt with the World Space method enhances the learning experience.
- Many test participants presumed the brain would have to multitask less with the World Space method if the corresponding task involved moving and interacting with the environment due to a similar situational context. Some of those participants additionally presumed the Screen Space method would be more challenging to use if the corresponding task required moving due to a mismatching situational context. However, those are just assumptions, as the use case of the user study was limited to performing stationary tasks. The effect of task-related movement on the usability of both visualization methods needs to be further investigated. In addition, it is crucial to find effective ways to counteract the location-dependency problem of World Space.
- To draw more meaningful conclusions about which visualization method is more pragmatically efficient, a more narrow user study focusing on participants with high (or equally low) proficiency in both methods is recommended. Perhaps the best time for such a comparison is in a few decades with a new generation of users for whom mouse and keyboard are no longer the primary "language" of computer interaction.
- A considerable perceived strength of World Space was the facilitated development of a mental map through perceptual cues. It needs to be explored whether the manual movement of the image plane in World Space (e.g., pitch/yaw rotation, `LookAtMe` button, the option to reposition the image plane in front of the camera) hinders the development of a mental map.
- Some test participants felt like the World Space method did not make use of its full potential by just providing a 2D image plane. They especially thought that large diagrams could be better organized and grouped in a three-dimensional arrangement, reducing complexity and making even more use of spatial awareness. Past attempts at such multi-dimensional data structures like the cone tree were met with skepticism in subsequent user studies [32], however the cone tree did not make use of 3D immersive AR technology. Research needs to be conducted on whether and how three-dimensional information visualization can be usefully integrated into the World Space method.
- As some participants performed the tasks while standing and others while sitting, no clear conclusions could be drawn from the participants' conflicting feedback.

Thus, the ergonomic qualities and physical exhaustion of the two methods need to be further investigated.

List of Figures

Bibliography

[1] R. Spence and A. J. Drew. “Graphical exploration in electrical circuit and model design”. In: *Proceedings, 2nd Man-Computer Comunication Seminar*. 1971, p. 61.

[2] R. Spence and M. Apperley. “On the use of interactive graphics in circuit design”. In: *Conference IEEE of circuits and systems*. 1974, pp. 558–563.

[3] I. E. Sutherland. “A head-mounted three dimensional display”. In: *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*. 1968, pp. 757–764.

[4] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. “Augmented reality: A class of displays on the reality-virtuality continuum”. In: *Telemanipulator and telepresence technologies*. Vol. 2351. Spie. 1995, pp. 282–292.

[5] T. Munzner. *Visualization analysis and design*. CRC press, 2014.

[6] A. Cockburn and B. McKenzie. “Evaluating spatial memory in two and three dimensions”. In: *International Journal of Human-Computer Studies* 61.3 (2004), pp. 359–373.

[7] R. Palmarini, J. A. Erkoyuncu, R. Roy, and H. Torabmostaedi. “A systematic review of augmented reality applications in maintenance”. In: *Robotics and Computer-Integrated Manufacturing* 49 (2018), pp. 215–228.

[8] A. Büchner, G. Micheli, J. Gottwald, L. Rudolph, D. Pantförder, G. Klinker, and B. Vogel-Heuser. “Human-centered Augmented Reality Guidance for Industrial Maintenance with Digital Twins: A Use-Case Driven Pilot Study”. In: *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 2022, pp. 74–76. DOI: 10.1109/ISMAR-Adjunct57072.2022.00024.

[9] D. Moody. “The “Physics” of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering”. In: *IEEE Transactions on Software Engineering* 35.6 (2009), pp. 756–779. DOI: 10.1109/TSE.2009.67.

[10] T. Brüggemann. *ba Ws2324 Tim Brueggemann Comparing Diagrams*. https://gitlab.lrz.de/IN-FAR/Thesis-Projects/ba-ws2324-tim-brueggemann-comparing-diagrams. Accessed: 01/06/2024.

[11] K. Marriott, J. Chen, M. Hlawatsch, T. Itoh, M. A. Nacenta, G. Reina, and W. Stuerzlinger. “Immersive analytics: Time to reconsider the value of 3d for information visualisation”. In: *Immersive analytics* (2018), pp. 25–55.

[12] T. Dwyer, K. Marriott, T. Isenberg, K. Klein, N. Riche, F. Schreiber, W. Stuerzlinger, and B. H. Thomas. “Immersive analytics: An introduction”. In: *Immersive analytics* (2018), pp. 1–23.

[13] N. ElSayed, B. Thomas, K. Marriott, J. Piantadosi, and R. Smith. "Situated analytics". In: *2015 Big Data Visual Analytics (BDVA)*. IEEE. 2015, pp. 1–8.

[14] B. H. Thomas, G. F. Welch, P. Dragicevic, N. Elmqvist, P. Irani, Y. Jansen, D. Schmalstieg, A. Tabard, N. A. ElSayed, R. T. Smith, et al. "Situated Analytics." In: *Immersive analytics* 11190 (2018), pp. 185–220.

[15] Y. Jansen and P. Dragicevic. "An interaction model for visualizations beyond the desktop". In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (2013), pp. 2396–2405.

[16] W. Willett, Y. Jansen, and P. Dragicevic. "Embedded data representations". In: *IEEE transactions on visualization and computer graphics* 23.1 (2016), pp. 461–470.

[17] M. Slater. "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1535 (2009), pp. 3549–3557.

[18] J. K. O'regan and A. Noë. "A sensorimotor account of vision and visual consciousness". In: *Behavioral and brain sciences* 24.5 (2001), pp. 939–973.

[19] B. H. Thomas. "Have we achieved the ultimate wearable computer?" In: *2012 16th International Symposium on Wearable Computers*. IEEE. 2012, pp. 104–107.

[20] M. Weiser. "Some computer science issues in ubiquitous computing". In: *Communications of the ACM* 36.7 (1993), pp. 75–84.

[21] M. Appel and N. Navab. "Registration of technical drawings and calibrated images for industrial augmented reality". In: *Machine Vision and Applications* 13 (2002), pp. 111–118.

[22] I. O. F. Standardization and I. O. f. Standardization. *DIN EN ISO 9241-110, Ergonomie der Mensch-System-Interaktion. Teil 110, Interaktionsprinzipien (ISO 9241-110:2020) - Ergonomics of human-system interaction. Part 110, Interaction principles (ISO 9241-110:2020)*. Frankfurt am Main: Beuth Verlag GmbH.

[23] M. Rauschenberger, M. Schrepp, M. Pérez Cota, S. Olschner, and J. Thomaschewski. "Efficient measurement of the user experience of interactive products. How to use the user experience questionnaire (UEQ). Example: Spanish language version". In: (2013).

[24] W3C. *Web Content Accessibility Guidelines (WCAG) 2.1 - Success Criterion 1.4.3 Contrast (Minimum)*. `https://www.w3.org/TR/2023/REC-WCAG21-20230921/#contrast-minimum`. Accessed: 12/22/2023.

[25] Omti9. *Unity Tutorial - Off-Screen Target-Indicators*. `https://www.youtube.com/watch?v=Ffi8kz6AheA`. Accessed: 12/21/2023.

[26] G. Charness, U. Gneezy, and M. A. Kuhn. "Experimental methods: Between-subject and within-subject design". In: *Journal of economic behavior & organization* 81.1 (2012), pp. 1–8.

[27] B. Laugwitz, T. Held, and M. Schrepp. "Construction and evaluation of a user experience questionnaire". In: *HCI and Usability for Education and Work: 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2008, Graz, Austria, November 20-21, 2008. Proceedings 4*. Springer. 2008, pp. 63–76.

[28] I. O. F. Standardization and I. O. f. Standardization. *ISO 9241-11 - Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs): Part 11: Guidance on Usability*. Berlin: ISO, 1998.

[29] J. Nielsen. "First rule of usability? Don't listen to users". In: *Retrieved August* 5 (2001), p. 2001.

[30] L. Ehrenstrasser. "Robert Spence. Information Visualization Design for Interaction". In: *Information Design Journal* 17.3 (2009), pp. 279–280.

[31] D. Makowski, M. Sperduti, S. Nicolas, and P. Piolino. ""Being there" and remembering it: Presence improves memory encoding". In: *Consciousness and cognition* 53 (2017), pp. 194–202.

[32] A. Cockburn and B. McKenzie. "An evaluation of cone trees". In: *People and Computers XIV—Usability or Else! Proceedings of HCI 2000*. Springer. 2000, pp. 425–436.