LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
Department "Institut für Informatik"
Lehr- und Forschungseinheit Medieninformatik
Prof. Dr. Sven Mayer

Master Thesis

Visualization of Human Behaviour in VR

Michael Hufnagel
m.hufnagel@campus.lmu.de

Bearbeitungszeitraum: 1. 5. 2021 bis 27. 11. 2021
Betreuer: Dr. Sylvia Rothe
Verantw. Hochschullehrer: Prof. Dr. Sven Mayer

Zusammenfassung

Forscher oder Studenten, die sich in der virtuellen Realität (VR) informieren und Studien in diesem Bereich durchführen wollen, müssen sich Gedanken darüber machen, was sie in ihrer Forschung messen wollen. Während es offensichtliche Elemente wie die Verfolgung der Bewegung und Orientierung eines VR-Benutzers gibt, die oft berücksichtigt und analysiert werden müssen, gibt es andere Aspekte, die Forschern helfen könnten, Studienergebnisse sinnvoll zu nutzen und ihre Visualisierung zu unterstützen. Diese Arbeit berichtet über die Implementierung eines VR-Logging- und Analyse-Tools. Aspekte wie Sentiment, Emotion, visueller Fokus und Visualisierung von gesprochenen Inhalten werden in ein Konzept integriert und teilweise umgesetzt. Der umgesetzte Prototyp und die Konzeptideen werden mit einer Expertenbefragung von acht Studenten, die bereits VR-Studien durchgeführt haben, evaluiert. Die daraus resultierenden Anforderungen, Ideen und das allgemeine Feedback werden diskutiert. Dies gibt einen Einblick in die Nützlichkeit einiger der vorgeschlagenen Funktionen für neuere VR-Forscher sowie Ideen für zukünftige Forschung und Verbesserungen oder Erweiterungen für dieses Analyse-Tool.

Abstract

Virtual reality (VR) researchers or students who want to inform themselves and conduct studies in the field have to think about what and how they want to measure in their research. While there are obvious elements such as tracking the motion and orientation of a VR user, which often have to be considered and analysed, there are other aspects that might help researchers make sense of study results and support their visualization. This thesis reports on a VR logging and analysis tool implementation. Aspects such as sentiment, emotion, visual focus and spoken content visualization are integrated into a concept and partially implemented. The implemented prototype and concept ideas are evaluated with an expert interview of eight students who previously conducted VR studies. The resulting requirements, ideas and general feedback are discussed. This gives an insight into the usefulness of some of the proposed features for newer VR researchers as well as ideas for future research and improvements or extensions for the logging and analysis tool of this thesis.

Contents

1 Introduction **1**

2 Related Work **3**
2.1 Analysis Tools in AR and VR research . . . 3
2.2 Analysing Speech . . . 8
2.2.1 Word Occurrences . . . 8
2.2.2 Sentiment Recognition . . . 9
2.3 Differences of Emotion and Sentiment Analysis . . . 10
2.4 Emotion Recognition . . . 10
2.5 Visual Focus . . . 11

3 Ideation and Concept **13**
3.1 Ideation . . . 13
3.2 Concept . . . 13

4 Implementation **17**
4.1 Tool References . . . 17
4.2 Tool Prefab Components . . . 19
4.2.1 Logging Manager . . . 19
4.2.2 Analysis Manager . . . 22
4.2.3 Study Manager . . . 26
4.3 Post-Processing . . . 26
4.3.1 Speech to Text and Spatial Audio Events . . . 26
4.3.2 NLP and Sentiment . . . 27

5 Evaluation **29**
5.1 Interview Design and Questions . . . 29
5.2 Procedure . . . 30
5.3 Expert Interview Participants . . . 31
5.4 Results . . . 32
5.4.1 Analysis Tool Expectations . . . 32
5.4.2 Logging . . . 33
5.4.3 Analysis Tool . . . 35
5.4.4 Sentiment, Emotion and NLP Analysis Questions . . . 37
5.4.5 General Questions . . . 39
5.5 Discussion . . . 41
5.6 Limitation . . . 45

6 Future Work **47**

7 Conclusion **49**

A Expert Interview Pre-questionnaire **57**

B Expert Interview Questions **59**

1 Introduction

Many studies on immersive analysis tools exist that aim to support researchers in their work. They often focus on aspects like gaze, position and interactions in the virtual environment (e.g. [9][34]). There is less focus on audio and analysis of speech or emotion itself which are important parts of human behaviour. Speech can be analysed with Natural Language Processing (NLP) techniques to acquire sentiment or spoken content. Looking at the emotions elicited in an environment could be interesting for researchers since different virtual environments can elicit different emotions in study participants [17]. This might be helpful to e.g. discern if an environment is frustrating the user and therefore impacting the study data.

Dinis et al. [14] studied the design of virtual reality (VR) hospital rooms. They wanted to find out which emotions participants feel when the objects in the virtual environment vary. The study asked questions about the emotional state of participants each time they were shown a different room. They found that more positive elements (e.g. a painting) in a room result in a more positive feedback of participants. The used objects were an artwork, a plant, a landscape and a chair. Specifically, the artwork and landscape resulted in a positive response and the chair did not. Another study by Riva et al. [46] found that designing relaxing and anxiety inducing virtual environments measurably elicits the respectively intended emotion in users. Furthermore, they studied the relationship between emotion and presence. The findings indicated a circular relationship. Reports for presence were higher in emotional environments and the degree of presence influenced the emotional state of study participants. The study measured both presence and emotion with questions while the probands were in the virtual environment wearing a head-mounted display (HMD). A visualization of emotions might also be beneficial for VR researchers because of this relationship. Felnhofer et al. [17] studied five virtual parks which each aim to elicit one of the emotions of sadness, joy, boredom, anxiety and anger. This was accomplished for most of them. Specifically, anxiety and anger environments provided more ambiguous results. In these studies, participants had to fill out questionnaires [17][46][14] or report their emotions inside of the virtual environment [46]. Not all research can use these methods in their study but maybe would still like to know about a study subjects emotional state. This thesis wants to find out what aspects of human behaviour could benefit a researcher in his analysis and explores some of them.

VR researchers should be able to discern the effect of their environment on study subjects as well as analyse aspects of human behaviour such as spoken words, sentiment towards an object, emotion or visual focus. This could support them not only in their studies but also in the design phase. Towards creating a VR logging and analysis tool for VR researchers and finding out which of the proposed aspects are beneficial to visualize this thesis formulates the following research questions:

RQ1. What requirements do researchers have for a VR logging and analysis tool?

RQ2. Which elements of the implemented logging and analysis tool are useful for a VR researcher?

RQ3. What are the benefits if analysis methods for visual focus, emotion recognition, sentiment recognition and speech analysis are added to a VR analysis tool?

This thesis presents a concept and implementation of a VR logging and analysis tool based on the reviewed literature. The prototype and concept ideas were evaluated through an expert interview of eight students with experience in VR research. Through this method information was acquired about their requirements (RQ1), the usefulness (RQ2) and benefits (RQ3) of the proposed additions for measurements of human behaviour.

The following section describes research about previous work about virtual and augmented reality analysis tools. Thereafter, relevant research about the factors word occurrences, sentiment, emotion and visual focus and their visualization is presented (see section 2). The thesis continues

with the concept of the implemented analysis tool and the ideation process behind it (see section 3). Implementation details and references to other software that was used can be found in the following section 4. This thesis then reports the results of the expert interviews and their discussion in section 5. A prospect is given for future interesting works that were found through the aforementioned research and study (see section 6). Lastly, the thesis is concluded with a summary of its results in section 7.

2 Related Work

The following sections describe previous works done on AR and VR analysis tools (see section 2.1). Thereafter, the thesis elaborates on aspects that would be interesting to analyse and visualize such as speech (see section 2.2), emotion (see section 2.4) and visual focus (see section 2.5). There is a focus on word occurrences (see section 2.2.1) and sentiment recognition (see section 2.2.2). Furthermore, a distinction between sentiment and emotion is provided (see section 2.3).

2.1 Analysis Tools in AR and VR research

This section shows current scientific papers about analysis tools for AR and VR environments. The selected research is either a tool that is usable in VR, is an in-situ analysis tool with a wide variety of functionalities or formulates requirements that are relevant for the tool that is implemented. This thesis summarizes the process of informing the design of an analysis tool, the functionalities as well as the evaluation of prototypes of relevant related studies. This information helps to identify key features for the VR logging and analysis tool of this work. Furthermore, it provides insights into the evaluation methods that are currently used in the literature.

Mixed Reality Analytics Toolkit (MRAT) Nebeling et al. [34] developed the Mixed Reality Analytics Toolkit (MRAT). They interviewed eight groups of media, instructional and research teams to find out their requirements for the MRAT. Participants had to select a VR or AR project that the interview should focus on. Nebeling et al. [34] divided their interviews into four parts. They asked about the methods interviewees used for evaluation, the targeted users, the goal of the project and the insights they wanted to gather. The researchers described the following five requirements that were generated from the interviews:

1. “Need for evaluations beyond usability testing.”: Participants were less concerned with being able to evaluate usability aspects such as efficiency, effectiveness and satisfaction. They mostly wanted to answer larger research questions.

2. “Need for quantitative and spatial data collection.”: Interviewees expressed a need for spatial and quantitative data, like information about the users’ paths, their focus and where they clicked.

3. “Need for tracking diverse interactions.”: Another requirement was being able to track a variety of interactions. The interviewees had diverse types of projects and therefore had different needs. They mentioned gaze, the proximity of participants for social interactions, 3D landscape perception for a large screen, AR and VR setup and interaction tracking on different mixed reality (MR) devices. One interviewed team wanted "to use the VR headset as a 6-DOF camera for retaking film scenes" [34].

4. “Need for flexible task definitions.”: Another requirement generated from the interviews was a need for task definitions that are flexible. Participants talked about study designs with specifically designed tasks, timed events and timed tasks. Others did not have a task definition at all, since students were expected to experiment.

5. “Need for key performance indicators (KPI).”: The last requirement was for key performance indicators(KPI) such as the time spent in a location or where and how long someone is looking. Other aspects such as trends in targets or visited locations were mentioned. The interviewers reported: "Teams wished for MRAT to guide them in the development of heuristics and metrics." [34].

MRAT supports "Interaction Tracking, Task Definition, and Session Inspection"[34]. The tracking capabilities capture basic information in an MR study setting like task or virtual objects state, but also information about the participants. This includes position, gaze, gesture and voice command tracking. Gaze is tracked with a gaze cursor in MR applications and not with an eye tracker. Screenshots can be taken when an event happens or at an interval. Performance parameters like memory use and frame rate are logged. The researchers explain that it can be used to estimate the impact of data collection on the performance of the MR application. MRAT writes information about the task state as well as their accuracy, duration and their success which depends on the task definition. Tasks can be of the type completed, skipped or started. Modes are used across tasks for "higher-than-task level" representations of study phases or conditions. Virtual object updates are tracked when they are created, removed, moved or change their orientation. MRAT uses markers to log events associated with a task and to track physical objects in AR. Markers can be physical or virtual. Physical markers are tracked with Vuforia[1]. The assignment of markers is done in unity[2]. The tool also provides custom logging for developers who want to define their own events. MRAT allows the definition of tasks. They can have a type, instructions for a user, target objects or markers for interactions as well as start and end markers. Metrics are used for determining the success of a task and heuristics are implemented for automatically recognizing task completion. The end of a task can also be triggered manually. MRAT lets users inspect sessions with different visualizations. They provide a web interface and an MR app for in-situ exploration. The web interface contains the tracked events on a timeline, global metrics, a 2D floor plan of the used area, event playback, filters and is synchronized with the 3D visualization in the MR application. MRAT can filter the users, modes, events, time and floor plan data. The in-situ analysis shows events at the location they happened. Nebeling et al. [34] use hierarchical clustering to reduce the number of events that need to be rendered. They are represented as a 3D marker that will split again if a user comes close enough. Individual event details can be inspected through interaction in the web interface or in the in-situ visualization. The MR app also allows the selection of groups of events.

MRAT was evaluated in a case study. Two teams of participants had to handle a disaster scenario using a phone, HoloLens[3] and Vuforia[1] markers on t-shirts. They had three tasks and a questionnaire to complete. Instructors for the scenario were tasked to analyse the data with the MRAT web interface, in-situ tool and a HoloLens[3] to make conclusions about the completion of tasks. Subjects of the disaster scenario study provided positive feedback with some usability issues that the MRAT researchers attribute to the design of the t-shirts with the markers. The instructors analysing the data were asked questions about which subjects were tagged as injured and which metrics they would use to compare success between the groups. MRAT was extended with other events that were needed for the study and could not have been known from the interviews informing the original design. The researchers added events that are thrown after an object is focused for a set amount of time as well as custom events necessary for the study. The tools server was extended with the ability to track messages from the study participants phones and the lab assistant who controls the disaster simulation. The goal of the MRAT case study was to evaluate feasibility for such usability scenarios, not usability.

MIRIA: A Mixed Reality Toolkit for the In-Situ Visualization and Analysis Interaction Data
Büschel et al. [9] developed the Mixed Reality Interaction Analysis toolkit (MIRIA). They informed their design through a literature review of previously developed analysis tools for user movement and interaction data. MIRIA is an in-situ tool that focuses on the analysis of AR study data. The literature review that the researchers conducted also contains studies about VR analysis

[1] https://developer.vuforia.com/ (last access 27.11.2021)
[2] https://unity.com/ (last access 27.11.2021)
[3] https://www.microsoft.com/de-de/hololens (last access 27.11.2021)

tools. The application from Büschel et al. [9] can help to inform the design of this thesis tool. Based on their findings they formulated functionalities that could support the process of visual analysis. These requirements are:

1. "Visualization of Position & Movement Data": This includes the visualization of individual trajectories from users or objects, gaze and areas where people linger for an extended amount of time.

2. "Visualization of Event Data": The researchers summarize application events and user interactions under this category.

3. "Visualization of Study Context & Stimuli": This requirement describes the need for any information that helps in understanding the context of a study. This includes study tasks, environment and stimuli. Their literature review showed the use of screenshots or videos to revisit an application state or a study setup.

4. "Filtering": The researchers describe this process as important for the comparison and evaluation of data. They also name the selection of study conditions and sessions as one of the most basic filtering tasks. Other categories they name are time, location, tracked objects or users.

5. "Annotations": Tagging data in an in-situ analysis tool is useful for clustering and filtering such annotated data. Furthermore, the export of it allows researchers to change their analysis tool between desktop and in-situ.

6. "Flexibility and User Preferences": This includes customization of the virtual environments origin for consistent visualization of the recorded data. Furthermore, visualizations should be able to be manually rearranged.

MIRIA saves the recorded study data into ".csv" and metadata associated with the study into ".xml" files. The tool implements various 2D and 3D visualizations. 2D views are shown on a resizable and repositionable rectangle and 3D views are rendered in the environment. User or object movement is visualized with tube segments and called "3D trajectories". Paths are colored in case multiple sessions are visible. An analyst can use color to encode the speed of tracked objects onto trajectories. "3D trails" are used to visualize the movement data a few seconds before a selected point in time. They look like the "3D trajectories". 3D objects like a HoloLens[3] or tablet can be used to mark a data point on trails and trajectories. MIRIA provides other objects to recreate spots in an environment that have changed since the study. Selected positions or interaction events can be visualized as "2D point plots" which are represented as colored circles. "2D scatterplots" visualize everything before a selected point in time. "2D heatmaps" can visualize aggregated event positions. "Media views" visualize images and videos on a 2D plane that can be attached to any moving and tracked object or static object in the scene. The researchers of MIRIA use an "event timeline" to show an events duration and point in time. Furthermore, they provide other menu panels. The "settings panel" is used to configure visualizations. A "session panel" allows an analyst to join an analysis session since MIRIA can be used in collaboration. The "study control panel" loads studies and lets users select study participants, sessions and conditions. In addition, it has the functionality to deactivate or activate 3D visualizations. The "playback panel", which contains the "event timeline", has a progress bar that can be used to select a point in time as well as controls for playback of the study data. MIRIA give multiple possibilities for filtering study data. This includes the selection of time spans, study sessions and conditions as well as being able to disable visibility of virtual objects. The requirement of annotation is addressed with virtual markers that are synchronized between analysts to mark significant spots.

The MIRIA research team decided to use expert "walkthroughs" for two previously published papers to generate feedback and show the technical validity of the tool for visualizing human-computer interaction (HCI) study data. Four participants were each guided by one of the authors. The researchers used a single user and a collaborative study. They performed walkthroughs of two additional studies with two of the authors to show further usage scenarios. Expert feedback for the tool was positive. The playback functionality was the most useful for the experts to analyse study data. Visualizing the study context through e.g. captured videos was also judged as useful for all experts. The engagement and interest for the tool was high for all participants. Some of the critical aspects or suggestions that were mentioned are the limited field of view, fatigue because of the HMD, labeling of trajectories, configuration of color and more support for annotations. The researchers of MIRIA saw through the variety of their studies the importance of spatial user interaction data and its exploration. Furthermore, they argue that the in-situ analysis helps in getting an impression of the movement trails scale and how they relate to physical objects. They mentioned the benefits of the easy deployment of MIRIA, its collaborative aspect and user engagement. This involvement was suggested as an indicator that presenting data in such an in-situ tool can be beneficial.

Immersive analysis of user motion in VR applications Kloiber et al. [26] developed a VR analysis tool to visualize user motion in an immersive virtual environment using the Unreal Engine[4]. The design was informed by a literature review on previous works of 3D trajectory visual analysis and immersive trajectory visualization. The researchers formulated three analysis tasks for VR-based training application data that can be used to understand user behavior. These tasks are:

1. "Analysis of change in behaviour of repeated motion"
2. "Detecting anomalies in task solving"
3. "Finding variations in user motion within single sessions or between multiple session"

Kloiber et al. [26] implemented several visualizations and interactions for their VR analysis tool. Furthermore, they give users the option to observe and navigate through the virtual environment where the data was recorded or get an overview through elevated vantage points that need to be created. It is not specified if they need to be inserted by a user themselves or are already built and only need to be inserted.

The tool uses tubes to visualize trajectories. The researchers try to achieve a motion similar to the original with centripetal Catmull-Rom splines [56][26]. Paths are colored per session and are textured with glyphs that indicate the movement direction. A session contains the head and hands movement data of one tracked user. Movement through teleportation can either be hidden or the resulting tube is reduced in diameter and not affected by spline interpolation anymore. The number of trajectories can be reduced by only showing the average path resulting from the head and hands position data. This average motion helps in recognizing if the movement of one source is very different. Kloiber et al. [26] generate keyframes where they display pose data with 3D objects of hands, an HMD and a triangle that connects all three components. These objects have the same color as the session they belong to. Keyframes are selected based on the magnitude of the spatial activity or through a uniform distribution. Users can choose the number of displayed keyframes. This calculation is done for one session and reused for all others to allow for comparison. The index of a key frame is shown with a number above the HMD object. The environment can be set to a white texture to reduce visual clutter for the mentioned visualizations.

Details of the spatial data can be displayed in multiple ways. Data can be filtered with time ranges. Users can interact with the trajectory to get information about the current time, interpolated

[4]https://www.unrealengine.com/ (last access 27.11.2021)

rotation and position, source of motion and session ID. Tracked user movement can be replayed through the animation of the virtual HMD, hands and triangle which leaves a trail behind. Speed and start time can be configured.

The tool allows for various interactions. Users can use a laser to interact with the trajectory and user interface. Trajectories can be gripped which creates a ball that can be grabbed and moved with the thumbstick or hand motion. A ball can be split by grabbing it with both hands to create two. The position of these objects can be changed and represent a filter that can be used to show the encompassed trajectory. Furthermore, these spheres can be touched to gain either additional information or to create the virtual objects representing a tracked users' pose. Menu interaction can be done with the laser or the index fingers of a user. The menu has options to configure the trajectory, keyframes, global animation and global time filter.

Evaluation of the tool was done with a game as a use case where four wheels had to be mounted onto a race car. The sessions for this scenario were simulated. The analysis of the data was done with five experts. They tried out the task that had to be analysed. The participants were presented with four analysis tasks after a presentation of the tool and a guided introduction of the features. After they completed them they had to fill out a usability questionnaire and talk about their experiences. Experts reported the animation of user motion as their favorite feature. Hiding teleport interactions in the trajectories was also well received. Some experts had problems with understanding the storyboard and the resulting keyframe visualization while others found it useful to compare poses between different sessions. The researchers discussed an improvement through a more understandable clustering algorithm. Participants did not notice a difference when the texture of the virtual environment was simplified in one of the tasks. The researchers theorized that this might be because of the already simple environment used in the study. Experts suggested that visual clutter might be a problem if the number of sessions that are compared increases further.

A Design and Application Space for Visualizing User Sessions of Virtual and Mixed Reality Environments Agarwal et al. [3] provided a design and application space for visualizing sessions in a virtual or mixed reality environment. They illustrated it by describing two example visualization applications. The researchers informed their design space with a literature review of related work about visualizations. Interactions, eye-tracking, physical motion and storylines are categories that were used to order the reviewed papers. The design space consists of the recorded session data, seven visualization categories, visualization approaches and two scenarios. These categories are:

1. "Entity identifiers": Objects or users that are uniquely identified.
2. "Event identifiers": Events that are uniquely identified.
3. "Entity timeline": Changes of entities over time such as interactions, associated events and attributes.
4. "Event timeline": Event occurrences over time.
5. "Event Density Field": Localized event occurrences with information about involved entities and other attributes.
6. "Trajectory View": Movement of entities over time.
7. "Scene View": Scenes that are displayed over time.

The researchers based the application scenarios of their design space on their experience and the literature review. They saw a use for visualizing user sessions for debugging applications and evaluation of user studies in HCI research.

The researchers provided two example analysis tools to demonstrate the use of their design space. One supports the visualization of a debugging scenario while the second is used to exemplify the evaluation of a user session. The prototype for debugging purposes shows a replayable virtual view and a timeline visualization displaying a tracked users hand velocity and an interactable object's velocity. The second prototype visualizes differently colored user and object events on a timeline for multiple sessions with two participants each. The density of events is visualized with a histogram. Waveforms are used to show where participants talked a lot. The prototype contains an overview as well as a playback mechanism of the virtual world. Both prototypes are 2D visualizations.

2.2 Analysing Speech

This thesis focuses on analysing different facets of human behavior in VR. This includes the content of spoken words like word occurrences and sentiment. The following section will inspect relevant related work. While spoken content is not needed for all sentiment analysis techniques it is needed for recognising word occurrences. Yu and Deng [55] talk about automatic speech recognition and the increase of accuracy through deep learning algorithms driven by mobile applications such as short message dictation (SMD), voice search (VS) and virtual speech assistants. They noted:

> "For example, the word accuracy in English or the character accuracy in Chinese in most SMD systems now exceeds 90 % and even 95 % on some systems." [55]

2.2.1 Word Occurrences

The analysis tool concept of this thesis included the visualization of spoken content. There are a number of natural language processing (NLP) tools to analyse text. The focus of this thesis is on the aspect of which words are spoken most often or what study subjects are talking about. For this purpose, the section talks about how to extract this information from spoken content and how it could be visualized.

> "Natural language processing (NLP), is the attempt to extract a fuller meaning representation from free text. This can be put roughly as figuring out who did what to whom, when, where, how and why." [23]

Khurana et al. [25] separate NLP into generating and understanding text. Natural language generation (NLG) produces natural language text while natural language understanding (NLU) or linguistics extracts information. The researchers mention a distinction of syntax, semantics, pragmatics, morphology and phonology for NLU.

This interest of this thesis lies in the visualization of word occurrences for supporting researchers in analysing spoken content. The text separated into words and their occurrences is called Bag-Of-Words (BoW)[23][51][59]. BoW has the disadvantage that semantic meaning is lost [59]. Words like husky and dog are counted separately from each other [59]. This effect can be mitigated for example through techniques that label groups of words like conceptual labeling (CL) [51] or Fuzzy Bag-of-Words (FBowC) using a fuzzy mapping [59].

Visualization of text There are different ways how processed words and text can be visualized. Word clouds mostly display static content for the purpose of summarizing text [20]. Heimerl et al. [20] created a word cloud explorer which can filter and search terms, show additional information about the word cloud, look at words in their original context and highlight words co-occurring in the same sentences. Hu et al. [21] developed SentenTree which visualizes words similar to a word cloud with different sizes depending on their occurrence. These words which are also called

nodes are connected to other co-occurring words. The researchers use this visualization for social media data such as tweets. Collins implemented a radial layout showing document content that can be filtered and searched [12]. The visualization shows occurrences through the level of the nodes transparency.

2.2.2 Sentiment Recognition

This section talks about the different ways sentiment analysis is done in the literature. Sentiment analysis mainly concerns itself with the polarity of opinions and their recognition [50]. Sentiment is comprised of the sentiment holder, an emotional disposition that can be positive or negative and an object that it is formed toward [33][50]. There is research concerning itself with not only the polarity but also the intensity of sentiment [57][50].

Soleymani et al. [50] surveyed different approaches for computational sentiment analysis. They were categorized into approaches that analyse text, visuals and multiple modalities. Sentiment text analysis has a variety of supervised or unsupervised learning approaches. Supervised analysis of sentiment uses annotated data to build predictive models. Text entries are processed into feature vectors that contain information about aspects such as word frequency which are then used to train machine learning algorithms that are validated with other annotated texts. These types of approaches are domain or context-specific. An example is the determination of sentiment polarity in the news or social media such as “Twitter” messages. Deriu et al. [13] trained a convolutional neural network (CNN) on data from “Twitter” [50]. Pang et al. [38] used support vector machines (SVM), maximum entropy classification and Naive Bayes machine learning methods on movie reviews [50]. Supervised approaches need annotated data which makes the process hard to implement across domains since texts can vary in structure [50]. Blitzer et al. [8] concern themselves with this problem of adapting a sentiment classifier to different contexts [50]. Unsupervised approaches use expert knowledge which is often encoded into lexicons. A lexicon contains the sentiment and its associated phrases [50]. A problem for this approach emerges if only the average valence of words in a sentence is considered for the sentiment calculation [50]. Irony, negatives, intensifiers or presuppositions are problems that can be addressed with contextual valence shifters [43][50].

The analysis of speech was described by Soleymani et al. [50] as a young field compared to emotion recognition. Mairesse et al. [31] found that sentiment information is contained in pitch-related features without textual cues [50]. Pereira et al. [39] focused on textual content for sentiment recognition [50]. The researchers used the opinions of queries to retrieve documents [39][50]. Kaushik et al. [24] found that sentiment analysis can be done even with word recognition rates of speech data that are low [50].

Visual sentiment analysis concerns itself with the recognition of sentiment in bodily or facial gestures or sentiment expressed by humans observing visual multimedia [50]. It is argued that visual sentiment analysis alone will not be enough to extract opinions out of multimedia data even though there is an increase of publicly available visual sentiment datasets and high accuracy techniques [50]. Extracting the entity or comprehending context from the visual content might be hard [50].

Sentiment analysis through multiple modalities is described as a good method to capture the sentiment polarity even without text and only audiovisual content [50]. Text analysis is still more effective in subjective recognition and entity aspect extraction [50].

Visualization of sentiment The visualization of sentiment is a well-explored topic. Kucher et al. [27] surveyed 132 techniques and described them with 7 groups and a total of 35 categories. These groups are the data domain of sentiment analysis, data source, data properties, analytic tasks, visualization tasks, visual variable and visual representation. Data domain denotes the area of sentiment analysis such as online social media. The data source describes where the information

that is displayed comes from. Data properties contain the categories geospatial, time series and networks. The analytic tasks group contains the analysis of polarity, emotion and stance as well as opinion mining. Examples of visualization tasks that were found are comparison, overview, monitoring and navigation. Visual variables that were used are color, position, size, shape and texture. Researchers used the visual representations line plot, area, node-link, clouds, maps, text and glyph. Kucher et al. [27] provide a full categorization with the amounts of visualization techniques that fall into each category.

2.3 Differences of Emotion and Sentiment Analysis

This paper briefly talks about the differences of emotion and sentiment described in the literature to avoid confusion between the terms emotion recognition and sentiment analysis.

Sentiment analysis tries to find if the polarity of an opinion is either negative or positive [30][50]. Soleymani et al. [50] surveyed multiple studies on the topic of uni- and multi-modal sentiment analysis. They also tried to define sentiment analysis and its differences from other terms like opinion mining and emotion recognition by looking at definitions in different studies. Studies often use both opinion mining and sentiment analysis interchangeably [50][37]. Soleymani et al. [50] suggest that emotion recognition and sentiment analysis should be separated. The reason for this is, that emotions are short-term and opinions are not [50][33]. It is also mentioned that emotion recognition can still be used to find the polarity of the sentiment [50].

2.4 Emotion Recognition

Emotion could be another interesting factor to visualize and analyse in a VR environment. Reasons for this assumption was the circular connection found between presence and emotion [46] as well as the lack of emotion visualization functionalities in related analysis tools (see section 2.1). The following section talks about models of emotion, the methods that can be applied for emotion recognition and emotion visualization.

Emotion detection (ED) needs emotion models as a foundation since they describe the representation of emotions [1]. It has to be clear which model is used for the ED [1]. Acheampong et al. [1] described discrete (DEMs) and dimensional emotion models (DIEMs) and talked about some examples. DEMs categorize emotion into separate classes. The model of Ekman [16] separated six emotions into the categories fear, surprise, disgust, anger, sadness and happiness. Their synergy can result in emotions such as pride, lust, shame and more. Plutchik [42] described eight emotions that can vary in intensity and produce complex emotions in combination. They are formulated as the opposite pairs surprise vs anticipation, anger vs fear, trust vs disgust and joy vs sadness. Ortony et al. [36] categorized 22 emotions which contain the six emotions of Ekman [16] and dislike, like, gloating, gratification, grief, fears-confirmed, hope, admiration, disappointment, pity, shame, appreciation, self-reproach, reproach, envy, relief. DIEMs assume that emotions are dependent and need a spatial space to represent their relationships [28][1]. The circular 2-dimensional model of Russell [47] differentiates emotions between the arousal domain or deactivation and activation and the valence domain or unpleasantness and pleasantness. It assumes that emotions depend on each other. Acheampong et al. [1] talked about other examples such as the 2-dimensional wheel of emotions by Plutchik [42] using valence and arousal and the 3-dimensional model by Russell and Mehrabian [48] using dominance, arousal and valence/pleasure. DEMs are widely used for the classification of emotion [10][1]. The reason for this is their lower complexity and number of categories compared to DIEMs [1]. “The dimensional models, however, are highly recommended in projects that represent similarities in emotions.” [1]

Some studies work on extracting emotional content directly out of text [4][5][1]. This is a good technique to get the specific emotion a writer of an opinion wanted to convey [5]. Since the implemented tool of this thesis is working in a virtual reality context with a head-mounted display,

with eye-tracking and other sensors there are other approaches to classify specific emotions. Lim et al. [29] surveyed recently studied emotion recognition techniques that are using eye-tracking. Some of the features that can be used are pupil diameter, position and responses as well as fixation duration and motion speed of the eye. Yong et al. [54] omitted eyebrows and eyes from images and were able to accurately classify emotions. Another possibility is emotion recognition based on speech. Harar et al. [19] for example used a german corpus and a Deep Neural Network (DNN) to recognize the emotions angry, neutral and sad.

Visualization of emotion Kucher et al. [27] surveyed sentiment but also emotion visualizations. Emotion categories are often represented with different colors or an interpolation (e.g. [22][58]) depending on the used dimensional model (e.g.[53]) [27]. Another paper reported on building emotional map visualizations [6]. The researchers used image data from social media and facial emotion recognition to determine the emotion of users as well as their geographical location. The recognized emotions are each individually visualized with a heat map covering the cities that were extracted from the geotagged photos.

2.5 Visual Focus

In the following section, this thesis talks about the information that can be extracted from eye-tracking data and briefly about calibration. Some of the possibilities of visualizing eye-tracking data are named.

Gaze points indicate where a person is looking which is recorded by eye-tracking devices [7]. The number of recorded data points can vary depending on the recording device. This data can then be processed further into different data types. Gaze points can be aggregated over a specified time span and area which is called a fixation. Rapid eye movement from one fixation to another is called a saccade. Smooth pursuit is when a participant follows the movement in a dynamic stimulus. A scan path consists of saccades and fixations which reveals a persons search behaviour. An area of interest (AOI) denotes a stimulus part of significance. It could be an object or part of it. Saccadic movement between two AOIs is a transition. An aggregation of fixations over time inside of an AOI is called dwell. These are all types of data from which metrics can be extracted (e.g. the fixation position). Other types of eye movement can be found in the research paper of Singh and Singh [49].

Eye-tracking needs calibration and validation to increase the accuracy of the produced data [11]. VR users perform this task by fixating different points on the HMDs display. The calibration process uses this information to adjust the gaze calculation. The actual coordinates and eye-tracker coordinates are compared to create a quality measure. It is used to validate the calibration. Clay et al. [11] experienced a decrease in eye-tracking accuracy due to e.g. the shift of the HMD. Calibration can be operator-, participant or system controlled [35]. An operator is a person controlling the eye tracker software, the participant is the user and the system is the software itself. The trend goes more towards system control [35] as previously exemplified in the work of Clay et al. [11].

Visualization of visual focus Visualizing eye-tracking data of a user is a well-researched field. Blascheck et al. [7] found 110 papers in their survey with different techniques visualizing eye-tracking data. Visualizations were categorized into AOI-based and point-based analysis. Thanyadit et al. [52] implemented a system for teachers to observe students in VR. They visualized gazes and the students as virtual avatars. Pfeiffer [40] visualized attention in 3D space. The heat map visualization is rendered on top of the targeted object. Pfeiffer and Memili [41] created heat maps that are rendered in real-time on top of moving objects.

3 Ideation and Concept

This section covers the ideation process that led to the idea of implementing a VR logging and analysis tool that visualizes aspects of human behaviour (see section 3.1). Afterwards, the concept that should inform a prototype implementation is presented (see section 3.2).

3.1 Ideation

The initial idea for this master thesis was created through a literature review of interesting VR and AR topics as well as a brainstorming session with two other master students and the supervisor. First ideas were about researching emotion and sentiment in a virtual environment which created the following research questions:

1. Does the absence of questioning about emotions have an impact on presence?
2. Which rooms or parts of the virtual environments are more frequently visited?
3. Which rooms or parts of the virtual environments elicit positive or negative sentiment?
4. Which rooms or parts of the virtual environments elicit which emotion?

In the review process of related studies, it was found that there already are studies that concerned themselves with the design of rooms and their impact on emotion [17][46][14] and the connection to factors such as presence [46]. It was noted though that participants in these types of studies had to fill out questionnaires [17][46][14] or report their emotions inside of the virtual environment [46]. This gave rise to the idea that researchers might profit from the information of their participants' emotional state or their behaviour in general. VR researchers might want to analyse this data to get new insights or make their studies or failures in the process easier to explain through visualizations. A VR analysis tool might be able to support them in this endeavour. This idea led to the following research questions:

1. Can the analysis tool help understand human behaviour in VREs?
2. Which elements of the analysis tool are useful for a researcher?
3. Which elements of the analysis tool are useful for VRE designers?

After researching other analysis tools in VR and AR (see section 2.1) it was considered if the extension of an already existing tool would benefit this and future research. A talk with one of the developers and researchers of MRAT from Nebeling et al. [34] was conducted where the wireframe of the planned analysis tool was presented (see section 3.2). Their tool was developed with VR in mind [34]. Based on the received feedback it was decided to build a separate tool.

3.2 Concept

The concept for this thesis logging and analysis tool was informed by the literature review of previous AR and VR analysis tools (see section 2.1). A wireframe prototype of the planned tool's features was created to visualize the concept and implementation goals. Implementation feedback was provided each week with the thesis supervisor and two master students who were at that time working on their own master thesis in the research field of VR. The wireframe prototype was done with "JUSTINMIND"[5]. The picture of a scene from unity was used as a background for the input elements from "JUSTINMIND" to create an interactable wireframe. In the following the initial

[5] https://www.justinmind.com/ (last access 27.11.2021)

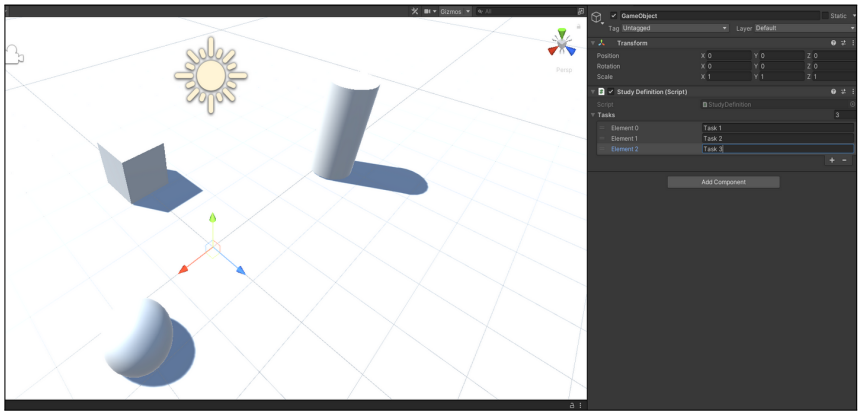


Figure 3.1: Wireframe screen showing a unity environment and a script that is used to define task names.

design ideas are presented. Not all features that are shown here were implemented in the final prototype. The full implementation can be found in section 4.

An important part of the study is the definition of the study itself. Users of the analysis tool should be able to define study parameters inside of unity in a script. One of these are the tasks and their names (see figure 3.1). Researchers should be able to define their studies beforehand to order their study data into their different stages.

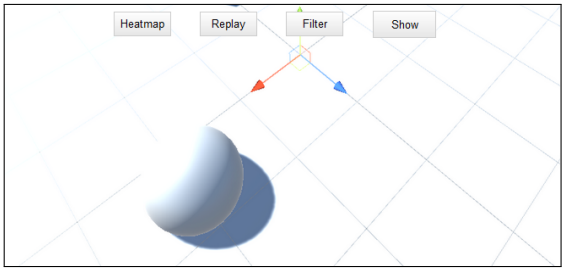


Figure 3.2: Wireframe screen cutout showing the menu items. Buttons are available to control the visualizations visibility, the data that should be displayed through filters and the replay functionality for previous user sessions.

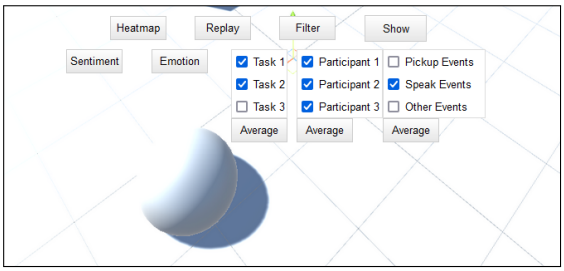


Figure 3.3: Wireframe screen cutout showing the conceptualized filters. Users should be able to filter by task, participant and event.

Figure 3.2 shows the menu of the application. This menu should be visible inside of the application since it should be an in-situ analysis tool. It should be able to easily get out of view so the analyst can inspect and interact with the visualizations. The menu should be able to control all visualizations visibility. Each visualization should be filterable. The sentiment and emotion heat maps should be configurable as well. The replay menu should allow a user to replay previous user sessions and show participant position and focus at a specific point in time.

The filter menu should contain categories for the tasks that were performed, for study participants and for each event that was logged by the application (see figure 3.3). Events that could be implemented are participant speech events and interaction events (e.g. if a user picks up an object). Researchers should be able to create their own events.

Participants movement should be visualized with a colored line (see figure 3.4). Participants color and line should be visualized in a legend. Events should be visible together with the paths participants walked. Multiple visualizations should be possible to use simultaneously. An information window should show on interaction with the events. This should reveal the type of event and all its associated data. The “Speak Event” for example should show the spoken text, the senti-

ment that is expressed and the emotion a participant felt in this moment. Analysts should be able to see what a participant did over a task duration and where.

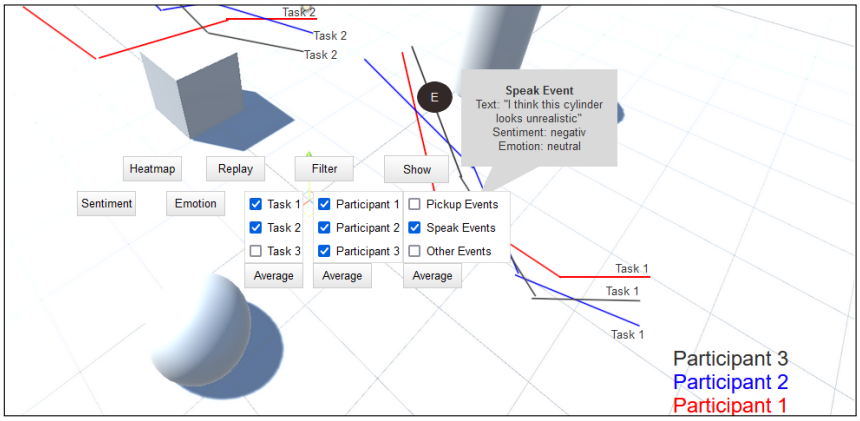


Figure 3.4: Wireframe screen cutout showing the paths and events that are recorded. Participants are distinguished by color. Events and paths show additional information on interaction.

Users should be able to get the average path participants walked (see figure 3.5). The resulting

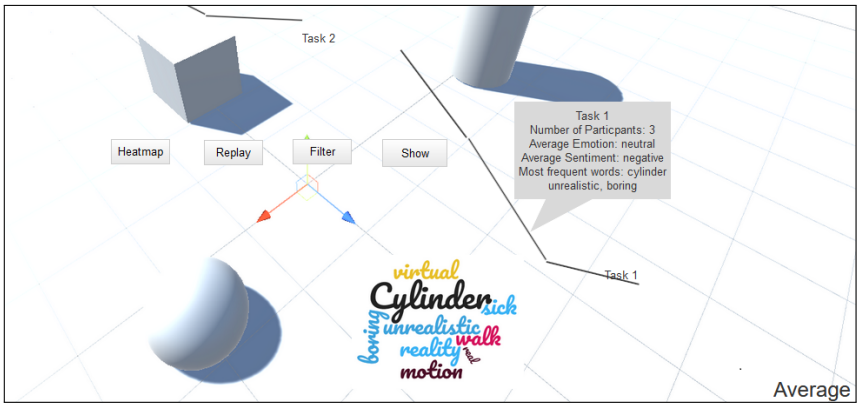


Figure 3.5: Wireframe screen cutout showing the average path that was walked by participants and a word cloud visualizing what was said in a part of the virtual environment. The average path shows additional information on interaction.

line should give information about the average data points over a task. This should include the number of participants, their average emotion and sentiment as well as their most frequent words. A word cloud is used to visualize the most frequent words that were spoken by participants in an area. Analysts should be able to have an aggregated view of what happened in a task, what participants spoke about while the study was executed and where they talked.

Users should be able to show specific points in time of their study to analyse important spots in the environment (see figure 3.6). This includes the events with the participants' position and their current focus. On interaction with the events, general information is displayed. The position events reveal information about the participant, the task that is currently performed and the focused object. Analysts should be able to inspect individual study participants movements and focus to revisit what happened in a task or study.

Emotions should not only be visible individually but in aggregated form (see figure 3.7). The emotion heat map shows categories from a DEM since they are often used for the classification of emotion [10][1]. These categories are shown as a legend inside of the analysis tool as individual colored text. The heat map shows sections in the virtual environment where certain emotions were more frequent. Each emotion category receives an individual color that is displayed on the

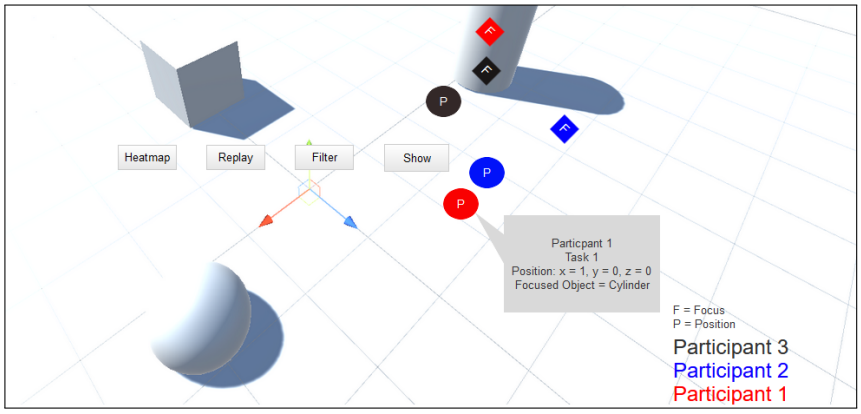


Figure 3.6: Wireframe screen cutout showing the position and focus of recorded users at a specific point in time.

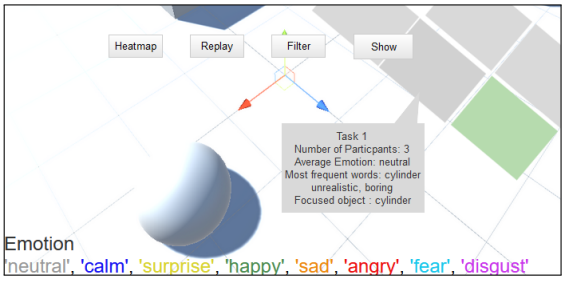


Figure 3.7: Wireframe screen cutout showing a heat map visualizing emotion.

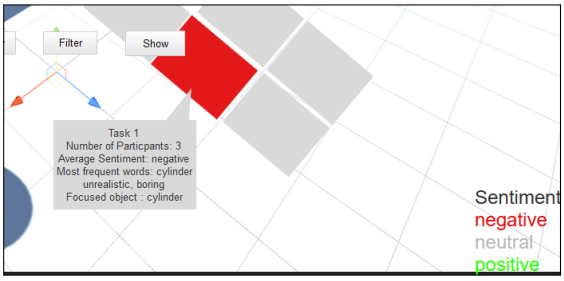


Figure 3.8: Wireframe screen cutout showing a heat map visualizing sentiment.

visualization. The heat map is interactable and shows the number of participants, their average emotion, the task that was performed and the focused objects. Analysts should be able to tell what participants felt in which part of the virtual environment on average.

The sentiment visualization should show similar to the emotion heat map where negative and positive sentiments were expressed (see figure 3.8). The categories negative, neutral and positive are shown as a legend inside of the analysis tool and are colored individually. The sentiment heat map should also be interactable and show the number of participants, their average emotion, the task that was performed and the objects that were focused. Analysts should be able to tell what sentiment participants expressed in which part of the virtual environment on average.

4 Implementation

This implementation of a logging and analysis tool for VR researchers is based on the previously explained concept (see section 3.2) which was informed by the literature review of implemented tools in research (see section 2.1). This thesis presents details about the used tools and why it was decided to use them (see section 4.1) as well as the components of the prototype that were implemented. The sections are divided into the functionalities that are directly programmed in unity (see section 4.2) as well as those that use other languages or tools (see section 4.3). All components that the implemented prototype consists of are illustrated in figure 4.1.

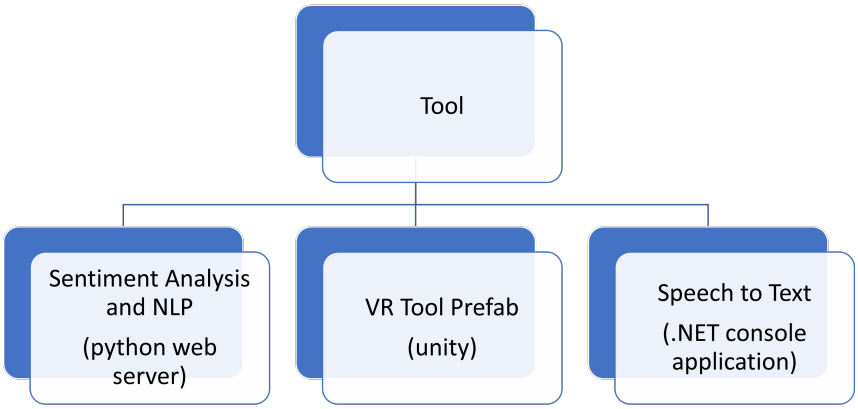


Figure 4.1: All components of the VR logging and analysis tool. It consists of a python[7] web server responsible for sentiment analysis and NLP, a ".NET"[9] console application converting speech in an audio file to text as well as the main application for logging and analysis in unity[8].

4.1 Tool References

The main part of the implemented logging and analysis tool is programmed in unity[8] as a prefab[6]. The sentiment analysis and natural language processing (NLP) was done in python[7]. Logging was done with unity[8] functions or assets. Audio event generation through speech to text was done with ".NET"[9] as a console application.

Logging The tool uses an edited version of "SavWav"[10] to convert the unity audio clip structure into ".wav" audio files. Position data of participants and their controllers depending on the tracking origin space is logged with the "OpenXR" plugin[11] from unity[8]. Mixed Reality Toolkit (MRTK)[12] was considered since Nebeling et al. [34] used it in their work. There was no need for most of the functionality except "OpenXR"[11], which is why the tool of this thesis did not integrate it.

[6] https://docs.unity3d.com/Manual/Prefabs.html (last access 27.11.2021)
[7] https://www.python.org/ (last access 27.11.2021)
[8] https://unity.com/ (last access 27.11.2021)
[9] https://dotnet.microsoft.com/ (last access 27.11.2021)
[10] https://gist.github.com/darktable/2317063#file-savwav-cs (last access 27.11.2021)
[11] https://docs.unity3d.com/Packages/com.unity.xr.openxr@1.2/manual/index.html (last access 27.11.2021)
[12] https://docs.microsoft.com/de-de/windows/mixed-reality/mrtk-unity/?view=mrtkunity-2021-05 (last access 27.11.2021)

Post-processing The python tool uses an edited version of a python server[13] implementation to respond with sentiment and NLP data.

The **NLP** was done with spacy[14]. The tool uses this library for sentence extraction and counting word occurrences.

Sentiment analysis was done with the classification model from Guhr et al. [18]. "Spacy-sentiws"[15] was considered which uses the German-language resource from Remus et al. [45]. It did not provide sentiment scores for a whole sentence but for the individual words and was therefore not used.

The ".NET"[9] console application uses windows desktop speech technology[16] for generating audio events through **speech to text**. Unity provides a "DictationRecognizer"[17] which was considered and tested. During the implementation process, some significant drawbacks were discovered that discouraged from using the functionality for this thesis. Most notably, users have to permit dictation in their speech privacy settings[17]. Researchers would have to ensure that the setting is activated on a study participants PC since the tool has to be executed while the study is running. In addition, the "DictationRecognizer" provides no indication when exactly the user spoke, which means the start of speech has to be guessed through a hard-coded offset. Lastly, since this functionality only works on windows it would make the virtual reality logging tool unusable for other HMDs. An alternative speech to text tool is the Windows desktop speech technology[16]. The contained API is used after the study by processing the recorded audio. An advantage of this approach is that the speech recognition works offline. This is important since this tool is processing audio data from studies which should not be passed on to third parties without the participants' consent. Another advantage is that the point in time when a speech event took place can be estimated accurately. The API has the disadvantage, that the location of the speaker is not known and has to be estimated based on the recorded spatial data. Other alternatives might be the speech SDK from Microsoft's cognitive services[18]. One of the main concerns for this tools speech recognition is the data security for study participants. For this reason, it was decided to use the windows speech technology, even though it is inferior in terms of accuracy. It has to be mentioned that the speech SDK[18] might still be a valid option if it can be configured so that there is no data trace as Microsoft claims[19].

Visualizations This tool used and edited an already implemented heat map[20]. One reason was that the grid mesh implementation provides an easy way to potentially add interactability for the grid tiles. An alternative would have been a shader-based heat map[21]. It requires calculating the heat for each pixel on a surface. While this approach is useful for small areas it would be too time-intensive for large virtual reality environments. The grid used for this study in turn can be adjusted. The granularity can be scaled down for bigger virtual worlds in case the grid mesh gets too large. For smaller areas, the size of the grid could be scaled down reducing the required vertices.

The hands from the unity "Oculus Integration"[22] were used to visualize movement for the

[13] https://gist.github.com/nitaku/10d0662536f37a087e1b (last access 27.11.2021)

[14] https://spacy.io/ (last access 27.11.2021)

[15] https://spacy.io/universe/project/spacy-sentiws (last access 27.11.2021)

[16] https://docs.microsoft.com/de-de/dotnet/api/system.speech.recognition?view=netframework-4.8 (last access 27.11.2021)

[17] https://docs.unity3d.com/2021.2/Documentation/ScriptReference/Windows.Speech.DictationRecognizer.html (last access 27.11.2021)

[18] https://docs.microsoft.com/en-us/azure/cognitive-services/speech-service/speech-sdk?tabs=windows%2Cubuntu%2Cios-xcode%2Cmac-xcode%2Candroid-studio (last access 27.11.2021)

[19] https://docs.microsoft.com/en-us/legal/cognitive-services/speech-service/speech-to-text/data-privacy-security?context=/azure/cognitive-services/speech-service/context/context (last access 27.11.2021)

[20] https://unitycodemonkey.com/ (last access 27.11.2021)

[21] https://github.com/ericalbers/UnityHeatmapShader (last access 27.11.2021)

[22] https://assetstore.unity.com/packages/tools/integration/oculus-integration-82022 (last access 27.11.2021)

tools replay functionality (see section 4.2.2).

In-situ Interactions and controls for the in-situ analysis tool were implemented with the unity "OpenXR" plugin[11] and the "OpenXR Toolkit"[23]. This logging and analysis tool should support as many VR HMDs as possible. It should not only be used for experiments inside of the lab but also outside with remote participants and different HMDs.

Test virtual environment The thesis used an edited version of the virtual environment designed by Reinhardt [44] to illustrate the functionalities of the logging and analysis tool to experts in the field of VR research (see section 5).

4.2 Tool Prefab Components

The prefab programmed in unity[8] is divided into three major parts "Study Manager", "Logging Manager" and "Analysis Manager" (see figure 4.2). The following section will explain what they contain and how they work. Changing between logging and analysis can be done through a script attached to the root node of the prefab.

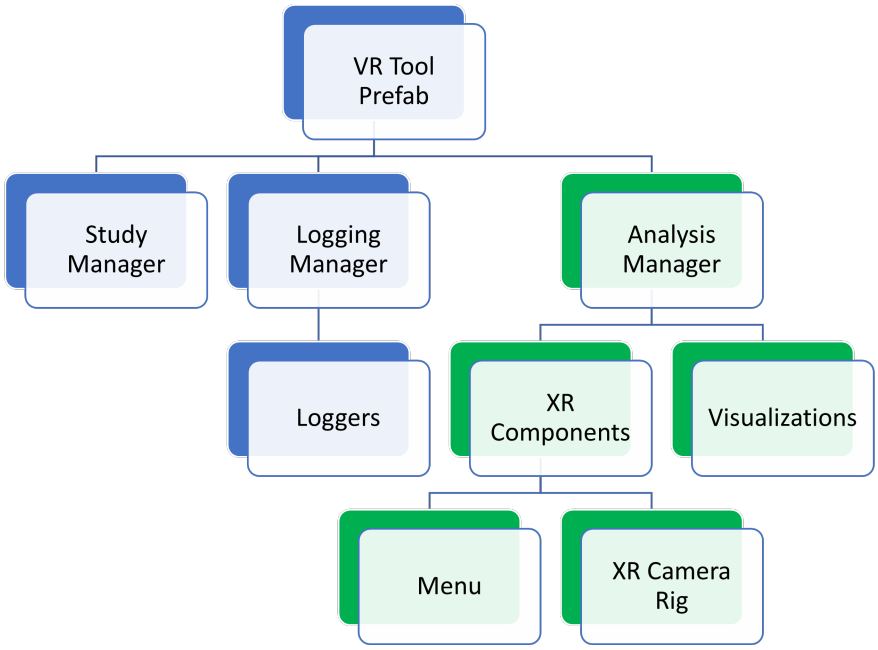


Figure 4.2: VR logging and analysis tool prefab components. They consist of the "Study Manager" containing study information administration and definition, the "Logging Manager" containing logger administration and configuration and the "Analysis Manager" containing the visualizations and components to interact and move inside of the virtual reality environment.

4.2.1 Logging Manager

The "Logging Manager" has a script that contains references to all loggers. Any loggers that are not needed for the study of a user can be removed or added here. Additionally, the number of logs that should be generated per second can be defined here. This is only relevant for the "ParticipantLoggers" and "TransformLogger" since they are logging data continuously. Users can

[23] https://docs.unity3d.com/Packages/com.unity.xr.interaction.toolkit@1.0/manual/index.html (last access 27.11.2021)

check if the data they want is properly logged here with a debug mode where they can define a logging length and the required parameters for starting the logging (see section 4.2.1).

Loggers and Events All loggers are "MonoBehaviour"[24] scripts. Some of them have options that can be configured in the unity editor itself. The loggers that were implemented are:

1. "AudioRecorder": Is responsible for recording audio from the HMDs and saving them as ".wav" to disk. The tool records multiple audio clips since recording audio in unity requires a fixed length [25] and because participants might have multiple tasks which should be recorded (see figure 4.4). The "AudioRecorder" scripts generated clips length can be configured in the unity editor from one second up to an hour. This only denotes the maximal length of one audio file. Audio is recorded until the logging is stopped (see section 4.2.1). Every time the recording is stopped the "RecordingInfo.json" file is created or updated with the start and stop of a recording session and the associated audio clips. The process of saving audio can impact the performance of the study program it is running on. This tool uses unity "Coroutines"[26] to convert the unity"AudioClip"[27] asynchronously into a byte array and threading from ".NET Standard 2.0"[28] to save it to disk in parallel, which reduces the effect. There might be studies that record tasks successively. Stutters of the program between these study sections might impact the results.

2. "ControllerLogger": This is an abstract class. Currently, two classes are inheriting from it, logging the right and left controller position through "OpenXR" as well as the controller transforms position in the virtual world if it is attached. The transform can be added through the editor.

3. "GazeLogger": This was intended to provide eye-tracking data, which is currently not possible through "OpenXR"[29].

4. "HMDLogger": Is logging the head position through "OpenXR" as well as the head transforms position in the virtual world if it is attached. The transform can be added through the editor. The script also logs the focused objects name, since this is done without eye-tracking but with a ray cast using the forward vector of the camera, which gives a rough idea where the participant is looking.

5. "InputActionEventLogger": This is logging the actions a study participant performs. Users can reference new input actions that should be tracked through the script. They can select the type of input action and add the transform that is executing it to log the position.

6. "CustomSpatialEventLogger": This is logging the custom events that the researcher or developer can define in his study program (see section 4.2.1).

7. "TransformLogger": This is logging the position of any transform with a "TrackableTransform" script attached.

All events that are created by the loggers are saved as data points into ".csv" files which makes them easy to analyse even outside of the analysis tool (see figure 4.3). A ".json" file records information about the used logging rate and recorded tasks (see figure 4.3). The "ControllerLogger"

[24] https://docs.unity3d.com/ScriptReference/MonoBehaviour.html (last access 27.11.2021)
[25] https://docs.unity3d.com/ScriptReference/Microphone.Start.html (last access 27.11.2021)
[26] https://docs.unity3d.com/Manual/Coroutines.html (last access 27.11.2021)
[27] https://docs.unity3d.com/ScriptReference/AudioClip.html (last access 27.11.2021)
[28] https://docs.unity3d.com/2019.1/Documentation/Manual/dotnetProfileSupport.html (last access 27.11.2021)
[29] https://forum.unity.com/threads/accessing-eyetracking-device-on-vive-pro-eye.1119559/#post-7208284 (last access 27.11.2021)

scripts, “HMDLogger” and “GazeLogger” are “ParticipantLoggers”, which write information to one “.csv” file. All data is saved per participant per session locally on the disk and can be found in a persistent path[30] on the HMD or PC running a study application with the logging tool. Study data is written immediately persistent to avoid loss of data in case the study is stopped because of errors or other unforeseen circumstances. This is currently not the case for all audio data. There might be data loss, in case the study is interrupted while a research subject is performing a task. This can be mitigated by reducing the clip length in the “AudioRecorder”.

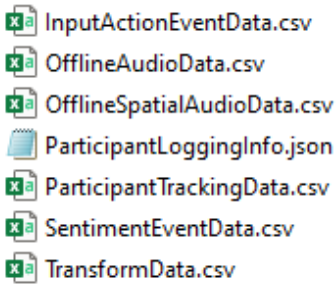


Figure 4.3: Logging “.csv” files that are saved per user and a “.json” file containing information about the logging session.

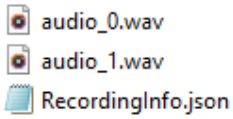


Figure 4.4: Example for recorded audio clips as well as a “.json” file containing information about the recording session.

Extension The tool of this thesis currently only creates events that have a 3-dimensional vector as position, except for audio events that are generated from post-processing (see section 4.3.1). It can be extended by creating an event class and a logger. Events can either inherit from “LoggingDataPoint”, “SpatialLoggingDataPoint” or any already existing classes that inherit from them. The visualizations only work with “SpatialLoggingDataPoints” since they need a position. Researchers might require events that only depend on time, which is why this tool makes the distinction between spatio-temporal and spatial data. Users can also add fields to the “ParticipantTrackingDataPoint” and “TransformTrackingDataPoint” if they want to track additional data continuously for a study participant or object in the virtual world. An example would be the addition of eye-tracking data from “OpenXR” if it becomes usable to the “ParticipantTrackingDataPoint”. Loggers for events should inherit from the “EventLogger” class and can then be added to the “Logging Manager”. Researchers can also use the custom spatial events directly in their programs code if they do not need a specific structure of their data.

Usage Starting the logging is simple. Users can call the functions of the static “Logger” class and provide it with information to start logging. This includes the session, participant and task ID. Identifying the session is necessary for collaborative studies since in that case, the ID is the same.

```
using UnityEngine;

public class DemoLogging : MonoBehaviour
{
    // Start is called before the first frame update
    void Start()
    {
        var sessionId = "1";
        var participantId = "1";
        var taskId = "Running␣Task␣1";
        var startInformation =
```

[30] https://docs.unity3d.com/ScriptReference/Application-persistentDataPath.html (last access 27.11.2021)

```
            new LoggingStartInformation(sessionId,
                                        participantId,
                                        taskId);
        Logger.StartLogging(startInformation);
    }
}
```

Custom events can be defined by providing the events name, message, value, value type and position and are written to disk when they are invoked.

```
using UnityEngine;

public class DemoLogging : MonoBehaviour
{
    // Start is called before the first frame update
    void Start()
    {
        var eventName = "EventName";
        var message = "Message";
        var value = 1f;
        var valueType = value.GetType();
        var position = transform.position;
        var customEventArgs =
            new CustomSpatialEventArgs(eventName,
                                       message,
                                       value.ToString(),
                                       valueType,
                                       position);
        Logger.InvokeCustomSpatialEvent(customEventArgs);
        Logger.StopLogging();
    }
}
```

After logging is concluded researchers can take the data and copy it to the persistent directory of the device they are using to analyse the data. The folder of a participant can be moved into the folder with the same session ID in case it was a collaborative study. If it was a non-collaborative study the session folder containing the participant can be transferred.

4.2.2 Analysis Manager

The “AnalysisManager” script starts the initial load of the logging data. It is attached to a prefab that contains the visualizations that will be rendered as well as the “OpenXR” camera rig and menu that are used to navigate through the virtual world and interact with the analysis tool.

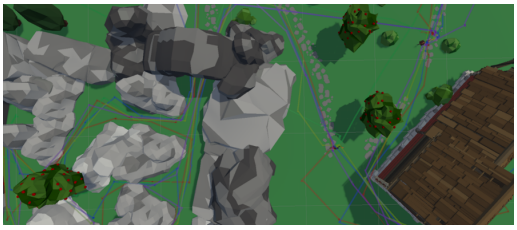
Visualizations The Visualizations (see figure 4.5) that were implemented in this tool are:

1. “Point Cloud”: Spatial logging data points are shown as labeled and colored cubes (see figure 4.5a). The cube’s size can be scaled and their transparency can be customized through the unity editor, but not in the virtual world. This tool uses a cube that can be exchanged for any other labeled geometry by exchanging the prefab used to render the visualization.

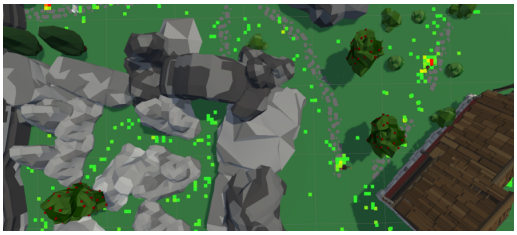
2. “Line Graph”: Spatial logging data points are connected through a colored line (see figure 4.5b). The lines width can be scaled and their transparency can be customized through the unity editor, but not in the virtual world.

3. “Heat Map”: Occurrences of spatial logging data points are shown on a 2-dimensional plane through a color gradient (see figure 4.5c). The default behavior shows spots in the virtual

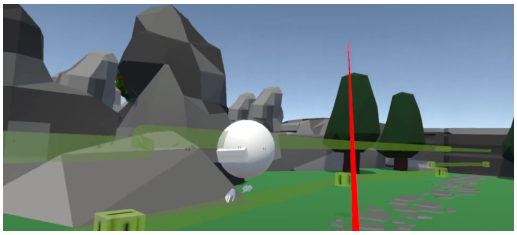
(a) The "Point Cloud" shows data points as a labeled and colored cube.



(b) The "Line Graph" shows data points, which are ordered through the time stamps, as a colored line.



(c) The "Heat Map" shows data point occurrences through a color gradient.



(d) The "Replay" functionality visualizes the movement of a single participant or transform.

Figure 4.5: Visualizations implemented for the in-situ analysis tool.

reality environment with no events as transparent. Occurrences are colored with a green to red gradient. The heat map visual can be customized with a blue gradient. Users can define their own heat map coloring by creating a material with a one-pixel height ".png" and adding it to the "HeatMap" prefab.

4. "Replay": Movement of a participants head and controller, as well as transform movement, can be replayed with this functionality (see figure 4.5d). "ParticipantTrackingDataPoints" and "TransformTrackingDataPoints" data is used for this (see section 4.2.1). Transform position, rotation and scale changes can only be visualized for non-collaborative studies. The transform information is currently tracked per participant. There is no distinction between which object moved for which participant. Participants head and controller position and rotation data is visualized with a sphere and hands (see section 4.1). The models can be customized in the unity editor through the "ReplayManager" script. This tool creates "AnimationClips"[31] from the logged events to achieve the "Replay" functionality.

Visualizations are created per event. This allows for combining different ones. A researcher can for example create "Point Clouds" for audio and input action events, while simultaneously showing the path participants took through a "Line Graph". This tool distinguished the "Point Cloud" and "Line Graph" as grouped visualizations and the "Heat Map" as an ungrouped visualization. The reason for this is that the former ones are visualizing data that is grouped for example by a study participants id. The "Heat Map" shows the occurrences of events which does not need to be grouped.

The program can be extended with more visualizations, by inheriting from the abstract classes enforcing this distinction.

XR Components The components used for the in-situ exploration of this tools visualizations are the XR camera rig from the unity "OpenXR" integration[11] and the menu. There can only be one camera in the virtual reality environment which is why all cameras will be disabled if the analysis

[31] https://docs.unity3d.com/ScriptReference/AnimationClip.html (last access 27.11.2021)

Figure 4.6: Interaction with an audio event visualized as a point cloud cube. Users can point and click on it with a ray cast. This will show the menu with information about the event. In the case of audio events, it will also show a play button to instantly replay the spoken content of a study participant.

mode is selected. This tools camera rig will then be set to active. Consequently, moving and interacting with the analysis tool in the virtual world is done with different controls from the study it will be used for. This program uses teleportation with a curved ray cast indication to move quickly through the virtual reality environment. Interaction is also done through ray casts both for the menu and the analysis visualizations (see figure 4.6). This tool implemented interactions with the point cloud cubes which shows information about an event in the menu. Audio events specifically provide the possibility of replaying spoken text from study participants. This functionality uses the audio events information that was generated from the speech to text application (see section 4.3.1). Such additional content can be customized through the "SelectionPanel" class for other events as well. Interactions with other visualizations like the "Heat Map" and "Line Graph" can be done with the "HapticInteractionManager".

The tools menu is divided into different tabs. Participants can use the study manager to select the study they want to analyse by setting it to active (see section 4.2.3). As an alternative, they can choose the study they want to analyse in-situ through the provided menu (see figure 4.7). Study information that was defined will be shown when the research name is selected and activated. Researchers have the "Line Graph", "Point Cloud" and "Heat Map" as visualization options. Every

Figure 4.7: User interface for selecting and choosing the study that an analyst wants to inspect.

selection will be shown in the filter and grouping tab for further definition of what data a user wants to inspect. Visualizations can be marked with a checkmark in the menu (see figure 4.8). A legend describing the color coding will be shown when a visualization is selected and the "Show" button is pressed. In figure 4.9 the visualizations are grouped by participant id, which is why the colors are shown with the associated study subject. "Heat Maps" show the occurrences of events. The legend that is shown for this type of visualization has a numeric value or an interval that is associated with a color (see figure 4.10). The maximum value shown depends on the spot in the

Figure 4.8: The “Show” tab is used to visualize the selected items with the selected filtering and grouping.

“Heat Map” with the highest number of data points. Researchers can filter their data for each

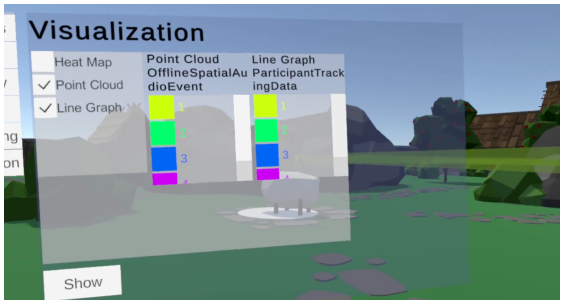


Figure 4.9: Legends that are shown when “Point Cloud” of audio events and a “Line Graph” of participant positions are selected.

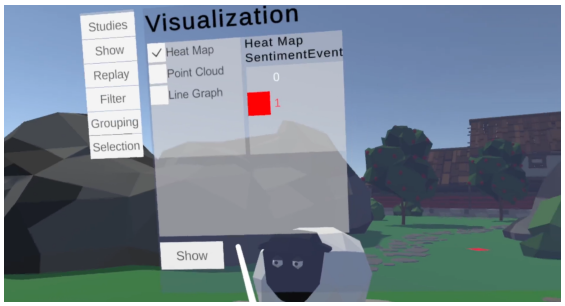


Figure 4.10: Legend that is shown when a “Heat Map” for positive sentiments is selected.

selected visualization (see figure 4.11). They can choose different participants or tasks they want to analyse. Visualizations are rendered for each selected event, thus allowing for multiple ones of the same type. Another category will appear for the “Heat Map” if there are sentiment events for the study. Researchers can choose to visualize where negative, neutral or positive sentiments were expressed. The “Line Graph” and “Point Cloud” need data that is grouped. Currently, researchers

Figure 4.11: Users can filter the data points they want to visualize for each selected visualization.

can either group by their study participants id or by the task id (see figure 4.12). The replay tab gives controls for the recorded participant and transform data (see figure 4.13). Users can select an entity by name that they want to rewatch. They can then use the play button on the left to start and the pause button on the right to stop the recording. The timeline allows researchers to jump further ahead in time. There is currently no way to select and replay multiple participants or transforms as mentioned in the limitations of this thesis (see section 5.6).

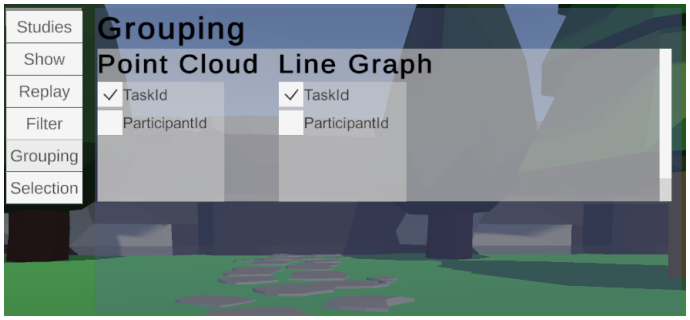


Figure 4.12: Users can group the data points they want to visualize for each selected visualization that needs it.

Figure 4.13: Users can replay participant and transform movement with this menu.

4.2.3 Study Manager

The "Study Manager" is used for managing and defining study information through the "StudyManager" and "StudyDefiner" scripts. Currently, studies have a name, an ID and boolean values denoting collaboration and activeness. This can be extended with additional fields in the "StudyDefinition" class. Studies can be created through the "StudyDefiner". Researchers can load and select already defined ones or edit them. The "StudyDefinition" is saved as a ".json" file at a persistent path[30] in a directory with the name of the study.

4.3 Post-Processing

Not all data needs to be immediately generated while the study is running. This section of the thesis describes how the speech to text, NLP and sentiment data is generated and how it can be used. These processes mostly use tools outside of unity (see section 4.1)

4.3.1 Speech to Text and Spatial Audio Events

The tool uses windows desktop speech technology[16] in a ".NET"[9] console application (see section 4.1). Users can start the speech to text conversion by providing the directory path of the study. The generated "OfflineAudioEventDataPoints" contain information about when and about what a study participant talked. Furthermore, it provides a confidence level indicating the accuracy of the speech to text conversion. The acquired data points are saved in a ".csv" file per session per participant (see figure 4.3).

Processing the audio after the study occurred has the disadvantage that there is no indication where participants spoke in the virtual reality environment. This tool uses the position information of the "ParticipantTrackingDataPoints" to estimate the location that an audio event took place. The program takes the "ParticipantTrackingDataPoint" with the closest timestamp to an "OfflineAudioEventDataPoint". This process creates "OfflineSpatialAudioEventDataPoints". They contain

the start and end positions that a study participant had while talking. The time differences between "ParticipantTrackingDataPoints" and a "OfflineAudioEventDataPoints" start and stop timestamp are recorded as well since it is an estimate. These offsets should be small as long as the logging rate is high, which can be regulated through the "LoggingManager" script (see section 4.2.1). The spatial audio data points are saved in a ".csv" file per session per participant separate from the audio data points generated from speech to text (see figure 4.3). This tools visualizations can then be used to look at the data. A user can look at the individual events and replay the audio if a "Point Cloud" is used (see section 4.2.2).

4.3.2 NLP and Sentiment

The created spatial audio events can be processed further. This tool uses a python server that receives "http" requests from the unity tool. Spatial audio events are sent per session per participant in "JSON" format. The python application then calculates the sentiment per sentence and the uttered words and their occurrences. This information is received by the unity tool. The sentiment information is written to a ".csv" file. Currently, there is no functionality to save the word occurrences. The generated "SentimentEventDataPoints" can be analysed with the implemented visualizations. This tools menu allows generating a heat map for the negative, neutral or positive sentiments (see section 4.2.2).

5 Evaluation

This thesis contains an evaluation of eight expert interviews conducted with people who have previously conducted a VR study. The following section shows the interview design and talks about the questions that were asked (see section 5.1). Then the interview procedure is explained (see section 5.2). The thesis presents the experts that were recruited for the study (see section 5.3) and the categorized results of the interviews (see section 5.4). The gained interview insights are discussed in relation to three research questions (see section 5.5). In addition, the limitations of the interview and the presented prototype are explained (see section 5.6).

5.1 Interview Design and Questions

A semi-structured expert interview was conducted for several reasons. First of all, experts in the field of VR research were needed to evaluate the logging and analysis tool, since they are the main users in the end and understand what might be necessary and useful to analyse. Second of all, the research questions of this thesis do not aim to answer the question of the usability of the implemented program. While an experiment where users test the tool might give other insights, this thesis is more concerned about the functionality aspects that might be most interesting or missing from the existing analysis tool. This and the ongoing COVID-19 pandemic are the reasons that an interview approach was chosen. Practically testing the tool would have still been a good approach since scientific papers like Büschel et al. [9] for example used "expert walkthroughs" to generate feedback for their tool MIRIA. Semi-structured interviews require mostly open-ended questions, which have the advantage of getting an individuals independent thoughts [2]. This is useful to answer RQ2 and RQ3. A focus group approach was also considered since it aims to discuss as many facets of a topic as possible [32]. Even though they are good to analyse diverse opinions, they are not recommended if expertise from the participants is necessary [32], which is why it was not used.

The study used semi-structured interviews and presentations between the individual question sets which explained the implemented tools functions and ideas. The questions are divided into the initial question, questions about logging and questions about analysis (see appendix B). The first one is about the general functional expectations of a VR analysis tool. The goal was to find out what people would expect, without showing what the program can do. The following questions were separated in logging and analysis since they are two different use cases with different requirements. While the main goal is to find opinions about the benefits of things like emotion recognition, sentiment recognition and speech analysis (see RQ3), it was also planned to generate feedback on the analysis and logging tool as a whole.

Therefore, the first questions were about what and how interviewees wanted their data to be saved. The question about what data the participants needed was split into two parts since we wanted to make the questions more open and give the participants more room to explain their answers in detail. The first part was about specific events they wanted to change, add or remove, which were previously shown in the presentation. In the second part, respondents were able to name any additional data they need which was not specifically event-related. Questions about the logging locations were also divided. We wanted to give participants room again to write and elaborate on their reasoning for their preference. Audio and the generated speech to text data is a specific part of logging since it is used for many of the analysis functions. This includes for example the sentiment analysis. That is why we wanted to ask if the recorded audio and generated text would benefit the interviewees' research. We conclude the logging questions by asking for any other remarks, ideas or requirements for logging or the logging tool, in case we missed any specific aspect that participants wanted to mention.

The analysis questions were divided into three parts. The analysis tool questions were used to gather mainly feedback and enhancement propositions for the program that was presented.

Aspects like. Sentiment, emotion, and natural language processing (NLP) questions were mainly used to gather insights about the potential benefits of the proposed analysis functions. Questions were formulated for all three aspects about the usefulness of a visualization in a collaborative and non-collaborative context. Specifically for emotion and sentiment visualizations, we asked for a comparison, since they may provide similar insights. The last section contained general questions. Since the implemented tool is mainly an in-situ tool we wanted to know the general opinions about such an approach and for what interviewees would use it. We also asked about the general logging and analysis tools they are normally using. This could have also been done in the pre-questionnaire (see appendix A) but we decided to incorporate it into the interview. The reason for this was that we might be able to gather additional insights if they had already used a similar in-situ tool or other things outside of for example Excel or SPSS. All other questions were used to gather any positive or negative feedback that participants wanted to express about the analysis or logging tool and to express any remarks that they felt were not addressed by the previous questions.

5.2 Procedure

The interviews were conducted over zoom[32] because of the ongoing pandemic. There was no need for an experimental setup since the prototype was presented with slides and a video. The interview started with introductions and a question about permission to record the interview. The first part was a brief explanation of what the goal of the study is, which is gathering feedback about the tool and opinions about other features like sentiment analysis. Then participants were invited to a miro board[33] (see figure 5.1) which contained all questions of the interview (see appendix B). They were asked the first question about the general expectations of an analysis tool. Interviewees were able to verbally answer the questions and in written form with post-it notes. Questions were formulated in English but participants were able to answer both in German and in English, depending on which they preferred and how they could better express their opinions. Explanations were given if necessary since English language capabilities were not requirements in the recruiting process. After the first question (see appendix B) the presentation about logging started. Google slides[34]

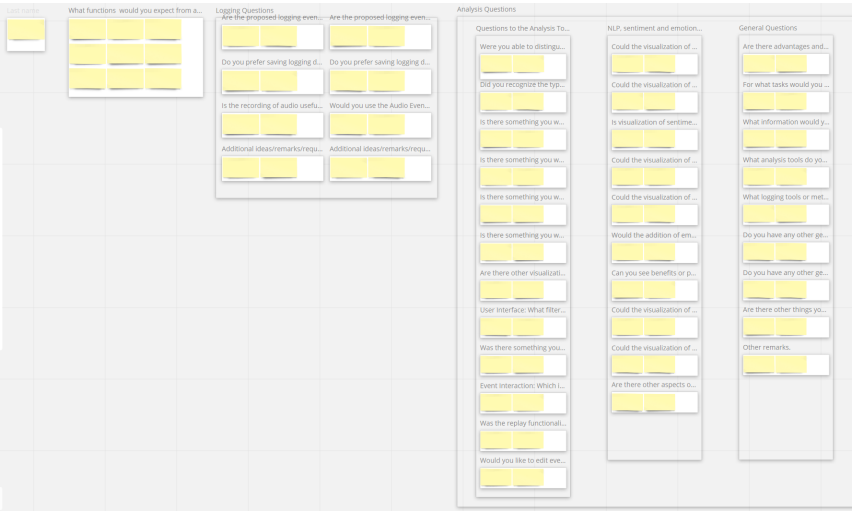


Figure 5.1: Empty miro board used for the expert interview. Participants could either verbally communicate their answers or write them on a yellow post-it themselves.

were used to present screenshots of unity showing the integrated tool, implementation examples

[32] https://zoom.us/ (last access 27.11.2021)
[33] https://miro.com/ (last access 27.11.2021)
[34] https://www.google.de/intl/de/slides/about/ (last access 27.11.2021)

and information about what and where data is saved. The presentation began with explanations on how to select the mode for analysing or logging data and showed what the tool is logging (e.g. controller position, input actions, audio). Next, interviewees were shown and explained how users can define studies they want to log data for. Furthermore, the code that is necessary to start logging and to define custom events was presented. Lastly, it was shown where data is saved, in which files and in which format. The presentation part was concluded by asking if there are any additional questions. Then the interviewees returned to the miro board.

After participants answered the questions about logging, the study continued with an explanation of the analysis portion of the tool. Screenshots of the in-situ user interfaces tabs were presented with a short explanation of what each of them does. Next, the three different visualizations were shown in bird view. Mock data for five participants was recorded and used. Positional data was displayed with the line graph and the heat map and audio data with the point cloud. Furthermore, an explanation about the visualizations color meaning was provided. For line graph and point cloud, it was to distinguish between participants and for the heat map to distinguish between high and low amounts of position events. Since the explanations about the usage of the in-situ tool were brief the study continued with a demo video showing with mock data parts of the functionality. The example data was recorded in an edited version of the virtual environment created by Reinhardt [44]. The video included a demo of the point cloud for audio and input data, the line graph for position data and the heat map for sentiment data. It was shown how to interact with point cloud events and which data is provided (e.g. spoken text in audio data event). The feature for audio events to replay what was spoken was presented. The replay functionality for positional data of the participants and tracked transforms was also shown with explanations about the limitations (see section 5.6). The user interface operation was also explained again. The explanation was concluded by showing ideas that were not implemented in the prototype but conceptualized. This includes the emotion heat map and the word cloud for spoken text. Again participants were asked if there are any questions about the explained details. The interview continued with the questions about analysis (see appendix B). For a detailed explanation of the questions see section 5.1. After these questions, the interview was concluded by thanking the interviewees and asking about their desired reward.

5.3 Expert Interview Participants

Eight participants were recruited for the interview through e-mail. All of them were currently in the process of a VR study or have already performed one, which was the only requirement. The invitation e-mail contained a link to the google forms[35] pre-questionnaire (see appendix A) and to Doodle[36] for registration purposes. The e-mail also contained a letter of acceptance template concerning the study participation, which the interviewees signed and sent before the study. Participants received an invitation link for zoom after registration in Doodle. The pre-questionnaire was used to find out their fields of expertise and education level. Of these interviewees five were currently studying HCI for their Master's Degree, two had a bachelors degree in media informatics and one a Bachelor in informatics (see table 5.1). Table 5.1 shows the anonymized participants' education level and fields of expertise. Each interview was recorded with the permission of the subjects. The length per interview was estimated to be one hour, but in the end lasted between 74 min and 100 min (M = 89.31 min, SD = 9.28 min). Each participant was compensated with a 10 euro amazon gift card or one MMI point.

[35] https://www.google.de/intl/de/forms/about/ (last access 27.11.2021)
[36] https://doodle.com/ (last access 27.11.2021)

Id	level of education	expertise in VR research	most recent VR studies
P1	B.Sc. Informatics	VR Interactions	Interactions with a museums object
P2	B.Sc. Media Informatics	Narrative in VR, VR attention/gaze guiding	VR attention guiding using diegetic cues in an interactive environment with collaborative storytelling
P3	studying Master HCI	Collaborative VR	Collaborative Learning Environment in VR
P4	B.Sc. Media Informatics	Subtitles in VR, usability in general, guidance for VR users	usability
P5	studying Master HCI	effects of color schemes on emotion amd cognitive performance in immersive VR	effects of color schemes on emotion amd cognitive performance in immersive VR
P6	studying Master HCI	VR locomotion	VR locomotion
P7	studying Master HCI	collaboration, interaction and learning	Social Interaction in Virtual Reality
P8	studying Master HCI	subtitles in Cinematic VR	master thesis optimized sound in Social VR

Table 5.1: Participants (anonymized) and their pre-questionnaire answers.

5.4 Results

The results from the semi-structured interview of eight participants are presented. Each interview was recorded but not fully transcribed since the participant data was collected with the miro board. When the answers of a participant were not understandable or abbreviated, recordings were used to get the initial context. These parts were then transcribed and commented with the participants' ID for further analysis. MAXQDA[37] was used for the analysis of the interview data. The analysis was oriented on the coding process described by Dresing and Pehl [15]. Codes or in other words categories were generated for each question by importing the "CSV" data from the miro board where participants answered the questions. Main- and sub-codes were created to categorize the answers [15].

5.4.1 Analysis Tool Expectations

It was asked which functions interviewees would expect from a VR analysis tool since this thesis wants to learn about their requirements (see RQ1). It was found that many participants had a need for **interaction tracking** (P1-P4, P7). Specifically **how often** does an interaction happen (P7), **when** or in which time frame (P2, P3, P7), **what** is the target and who is the interactor(P4) and what is the interaction itself (P2). P1 also mentioned an interest in **when an interaction becomes less random**. It was described as useful to know when a study participants' e.g. random button presses or looking around stops. The interviewee said that people often have difficulties following instructions in VR and it would be nice to know when their random actions stop and become more intended. Interview participants expressed an interest in **position** (P1-P4, P7), **focus** (P1-P5, P7) and **audio** (P5-P8) tracking.

[37] https://www.maxqda.de/ (last access 27.11.2021)

They were interested in how long **participants are in one spot** (P4, P7), the **path** they took (P1, P2), the **position** (P2) and **all avatar values in general** (P3). P7 mentioned an interest in **what triggers people to move**. P7 and P3 specifically talked about **controller** movement tracking. P7 mentioned controller movement through the air and how it could be interpreted. It was mentioned that it might be possible to guess the context from the movement or the frequency of interaction, which touches on tracking the controller position in the tracking space. P8 mentioned the interest in **head movement** tracking and visualization. P7 talked about an **example where tracking the movement might be less relevant**. In the study of P7 participants sat on a chair because of the small study room size.

P1-P5 and P7 were interested in **tracking focus**. P3 specifically mentioned only the tracking of **eye gaze**, P4 and P5 **eye-tracking**. P4 mentioned where and how long people are looking as interesting and suggested a **heat map** as visualization. **Other interests** of the participants concerning focus were, where the user is looking (P1), when an object enters the field of vision (P2) and how often a subject changes their focus (P7).

There were different interests in extracting **information from audio** (P5-P8). P7 and P8 mentioned the **length of speech**. P7 talked about **how often participants talk** and **information about the voice**. The interviewee wanted to know how interested, bored or distracted a person is. P8 mentioned the need for recording study audio and the visualization of not only speech length but also the point in time and speaker. P6 named **audio transcription** as a feature. P5 talked about an interest in **speech recognition** or in other words what is said by participants in a VR study.

Some interviewees mentioned **physiological measures** (P3, P4, P7). Things like heart frequency (P3, P4) for presence (P4), or like previously mentioned in the category **information from audio** measurements to find out if participants are interested, bored or distracted (P7), or brain wave and stress measurements (P3).

P1 and P2 spoke about **study information**, specifically how long the study tasks were. P4 talked about the **visualization with a heat map** in the context of gaze. P5 named it as a general need. P8 elaborated that the representation of the data hast to make sense and **depends on the use case**. P6 mentioned that **recording and replaying** the study to look at specific points in time and throwing **custom events** would be useful. Other expectations were that the tool should be **easy to use** (P6), data collection should be precise (P8), it should be able to exactly analyse what the researcher wants (P8) and it should allow drawing conclusions from the collected data (P8). P7 wanted **quantifiable data** in addition to things like Likert scales and questionnaires. It was mentioned that it is hard to support a studies hypothesis without something tangible and that tracking subconscious processes may help since the self-assessment of participants may be different from what really happened. There may be additional knowledge in the difference between self-assessment and actual behaviour. P6 mentioned the expectation of graphical UI analysis tools.

5.4.2 Logging

Logged Data: The first two questions both were about what logged data the interviewees need additionally to the proposed tools events. Since participants answers were similar for both, they were categorized together.

P6 and P8 found the data provided by the tool sufficient and had no suggestions for data they need or events they would like to change, add or remove. P4, P5 and P7 suggested that **eye-tracking** data would be beneficial. P7 elaborated that the gaze logger might be too inaccurate since study participants do not only focus on the middle of their field of view but mentioned that an HMD that supports eye-tracking is then needed. P3 wanted a way to **track network data**. The given example was that if there was a need to assign colors to study participants it would be helpful to know if they were transmitted and assigned correctly. It was also of importance for the interviewee that this is possible for **debugging** purposes and after the study to ensure that there were no errors. P3 talked about **tracking custom variables**. An example was the need for tracking

user interface elements like sliders and their changes. P1 wanted the **interaction target**, not just the input actions and when they took place. Even though it was out of the scope of the question the participant named the requirement for an **option to disable different loggers**. P2 mentioned data to analyse the **users pose** and body language.

Logging Location Preferences: Both questions about the preferred saving location of data were categorized together since they both contain the reasons and the preference itself.

Most users preferred to have **both** (P2, P3, P6-P8). Participants preferred **remote** (P1, P4, P5) more than **local** (P7, P8).

Interviewees that wanted both ways either wanted that **both work** (P7, P8) or to have a mixture for **redundancy** reasons (P2, P3, P6). P2 thought that the local saving could be used as a fallback. P6 mentioned that connectivity issues in the study might be a problem.

The advantages that were named for local saving were that it is **easier to implement** (P4), **less data loss risk** (P6, P8) and less problems with **data security** (P2, P7). P6 reasoned a lesser risk of data loss because there would be no dependency on the network.

Participants criticized that saving locally makes it **hard to access data** (P1, P4, P5, P6). It is harder to get the logged data if the **study takes place remotely** (P1, P5, P6) or if the **hardware** it is saved on is difficult to access (P4). P4 assumed **data loss risk** when just saving locally since the PC used for the study could break. The interviewee P4 also touched on how **concurrent access** for multiple analysts is hard if the data is not accessible from outside. P6 mentioned that not only accessing the data on a remote PC where you have no control over the setup takes more time but also the **sorting** of logging files since they are not in one place. P5 and P6 reported on **different logging paths** depending on the platform. P5 exemplified this with the process of accessing data on an oculus[38] HMD. In that case, a connection to the pc and an additional program was necessary to first of all access the data. Then participants had to find the logging information with the program which is hard if they are not familiar with it.

Audio: Here the questions were about the usefulness of audio recording for the interviewees and their reasoning. Participants who answered **no** (P1, P6) either **did not need it** (P6) or **did not have audio implemented** (P1). Of the interviewees who answered **yes** (P2-P5, P7, P8) three said it is because their research was **collaborative** (P3, P7, P8) and three because it could be generally useful (P2, P4, P5). P7 mentioned that it is hard to interpret movement alone if the virtual avatar has a basic representation. P2 said that it **depends on the use case**, but may help in general e.g. when there are **questions during the study**. P4 elaborated that audio is useful when **context additionally to tracking data** is needed. The interviewee described that the reason for an abnormality is lost if a researcher only has the tracking information. P2 and P5 reported scenarios from their studies where **participants talk** even though there is no collaborative partner in the virtual environment. P2 described how participants tried to talk to a character seated in the virtual environment. Subjects in the study from P5 talked out loud if they had problems, did not know how something works or when they wanted to give feedback.

Speech to Text: Six participants **would use** the generated audio events from speech to text (P2-P7). Some talked about the factors it would **depend** on (P1, P7, P8).

The benefits interviewees who would use them saw was a **reduced workload** (P2, P4, P5, P6), **seeing user communication** (P3) and **recognizing audio commands** (P2). P2 mentioned that some corrections might be necessary. P4 elaborated that having the transcription may make it more **probable** that the data is **used even if it is not the main focus** of the study. It also makes it **easier to search** the data compared to navigating through the audio (P4). P1 and P7 mentioned that the use of such events would depend on the **speech to text accuracy**. P1 would not use it for

[38] https://www.oculus.com/ (last access 27.11.2021)

a **low number of participants**. P7 theorized that it may not be enough as the **only tool** for audio analysis but would be useful for getting an **overview** of points of interest where participants talked a lot. P7 mentioned it **depends on the study** and may be helpful as support. P8 would use it for the **speaker and time of speech** and not for the content. If it is not necessarily needed P8 would generally not use it because of **privacy** reasons. P7 talked about the possibility of **other study designs** where a subject could talk instead of interacting e.g. with a controller if the audio events would be marked.

Logging Requirements: Participants answered any requirements, ideas or remarks they had for logging in general or for the tool that was presented. P1 was interested in **tracking the real-world movement of a user** and not just the virtual. The interviewee suggested that the **option to disable different logging capabilities** could be more user friendly and that the logging tools **csv files could be divided by task**. P7 mentioned that **synchronized data** is important so that it can be compared especially for multiple participants and when the study setup is offline. P3 was interested in a **reminder for users that logging is activated**. P1 talked about the use of **performance logging** e.g. for researching **VR sickness** by looking at the study participants devices performance differences. P6 wanted the ability to **measure time spans** since they would have to be calculated with the events after the study. Good **documentation** (P5), **intuitiveness for people who have less knowledge in VR** (P5), **fail-safe** execution (P8), **network data** like connection stability and latency (P7) and **structured logging data** that is not just a wall of text (P4) are other requirements participants had.

5.4.3 Analysis Tool

The results in this section were used to gather feedback about the presented analysis tool.

Line Graph Paths: Most participants were **able to distinguish** the line graph paths (P1, P3, P4, P6-P8) through the **different colors** (P1, P6, P8). Some expressed **problems** in case there are **too many lines** (P5, P8) or if the **color matches the background** (P2). One suggestion was to create dotted lines in case there are too many participants (P5). P8 explained that many superimposed lines may benefit from shadows, patterns instead of monochrome colors or gradients.

Event Recognizability: Participants were mostly **able to recognize** events (P1-P4, P6, P8). P7 mentioned the **problem** of **multiple events overlapping** which may get very confusing. P1 had no problems recognizing point cloud and line graph events but with **distinguishing heat map events**. The interviewee elaborated that multiple selected events may overlap. This makes it harder to know which heat map they belong to. Other suggested improvements were a **different labeling** (P4, P6-P8) and **geometric forms** for the events (P5, P7). P4 thought about marking input events not with only an "I" but with the first letter of the actual action. An example was using "T" for teleport or a combination like "I(T)". The interviewee also mentioned that this might result in too much writing and therefore confusion. P8 suggested that custom events could be improved since they are all marked with the same letter which makes them hard to distinguish. P6 proposed icons and P7 symbols instead of letters. P5 elaborated that different geometric forms may help events stand out more and that letters could be used additionally.

Visualization Changes: These are the results of the three questions asking what participants would change about the three visualizations.

Interviewees would like to change the **line parameters** of the line graph (P2-P4, P7, P8). They mentioned **transparency** (P3, P4), **size** (P7, P8), **customizable colors** (P3, P7), **height** (P3, P4) and **emission** (P2). P2 elaborated that emission could be used to make lines stand out from the

background. P4 wanted to transpose lines downward because they are in the field of vision. Other suggestions were a **color blind mode** (P3), **bird view** (P8) and **direction arrows** for the line (P7). P3 suggested the **color blind mode** for all visualizations. P8 mentioned that **handling the overlap** of lines in a smaller room e.g. by adding zoom might be useful.

Participants wanted to customize the **point parameters** as well. They named **transparency** (P3, P7), **size** (P3, P4, P7), **height** (P3) and **customizable colors** (P3, P7). Participants mentioned again that **different geometric forms** (P7) or **labeling** (P6, P7) of the points would be beneficial. Suggestions were to use icons (P6) or symbols (P7). P3 elaborated that **customizing the label** could help for example if a color blind person wants to use the letters to differentiate people instead of with colors. Participants also suggested a **bird view** (P8), **clustering points** that are in one spot (P5) and **marking the position of points** (P1). P1 explained that showing the position of the events e.g. through a light cube in the air that is marking the spot is necessary. Researchers could only find events by walking through the whole environment if there is no marking or line graph.

The **heat map parameters** participants wanted to customize (P3, P4, P7) were **transparency** (P3, P7), **form** (P4), **size of squares** (P3, P7, P8), **height** (P3), **color** (P3). P4 suggested circles instead of squares to make the heat map more vivid. Other ideas were a **color blind mode** (P3), a **bird view** (P5, P8) and a way to **find heat map spots** (P1-P3). P1 suggested locating the heat map events by **marking spots** where they occur through a light cube similar to the point cloud events. P2 proposed either a visualization on **eye level** or in 3D. P3 talked about using a **pointer to find outliers**. P5 suggested a **mini-map** to **find heat map spots** and to get a **bird view** of the whole virtual environment.

Replay Changes: Most participants wanted to **look at multiple animations** at the same time (P1, P3, P4, P7). This includes playback of **participants and trackable transforms** (P4, P7) as well as of **multiple participants** (P1, P3) at the same time to find groups building (P1). P1 expected to **get the participants point of view**. P2 wanted to see **when input events were executed** and **controls** to e.g. jump forward or configure playback speed. P3 said it should be **clear which animation is running**. Participants also wanted to be able to **select a session or task to replay** (P6), see **timeline details** like seconds (P7) and **comprehend the current replay state** (P8).

Other Visualizations: Most interviewees proposed a **gaze visualization** (P3-P5) as a **heat map** (P3-P5). P8 talked about **screenshots** for research papers. P7 described a **3D bar chart** that could be visible from the side and use color encoding for something else than occurrences of events like the heat map. P2 proposed a visualization of the **length of stay with the line graph**. P1 wanted to see **how a participant interacted with which object**.

Event Interaction: P8 wanted to know the **point in time when pointing at a part of the line graph**. P6 would have added **information about the trackable transforms** if a user interacts with them. P1 wanted the **amount of events when the heat map is clicked**. P1 proposed an interaction with the **line graph to get the length of the participants stay** and through this their speed. P7 had **no suggestion** but noted that looking at **details and the audio was most important**.

Audio Replay: All participants thought the replay functionality of voice recordings **was useful** (P1-P8) but P3 noted that it **depends on the research question**. P4 mentioned that a researcher can **get the emotion, thoughts and context** at the same time with the function. Participants said it helped to be **more involved with the data** (P2, P4) and to **check if the NLP functions worked** (P1, P5-P8).

Editing Events: Most Interviewees wanted to edit events **both** inside and outside of VR (P2, P4, P7, P8). P8 mentioned that editing outside is helpful if the VR glasses are no longer available. P7 proposed to edit in VR for small changes since it might be exhausting for larger edits.

P4 and P5 would prefer editing **inside** since it would be hard to get outside for **simple edits** (P5) and there is **less of a chance to forget** the necessary change (P4). P1, P3 and P6 expressed a preference for changing events **outside** since it is **easier to write** (P3, P6) or **hard in VR** (P1).

Filter Categories: Participants proposed filter categories for **object interaction** (P1), **gaze** (P3), **session** (P6) and **time** (P8). P2, P4, P5 and P7 did not have any suggestions.

User Interface Problems It was unclear for P5 in which **order a user has to go through the menu** (P5). Some participants suggested that the **show button should be in each user interface tab** (P2, P3). P7 and P8 suggested using **less scroll bars** either through a **bigger menu** (P7) or a second **hierarchy for the menu buttons** that allows selecting of the filter for a visualization for example (P8).

5.4.4 Sentiment, Emotion and NLP Analysis Questions

Sentiment Visualization in Collaborative Studies: All participants thought that visualizing sentiment for collaborative studies is **useful** (P1-P8). P2 said it can be used as a **reminder** of how a participant felt in a moment in a study. P6 mentioned that it reduces **workload** and gives an estimate of how users felt. P8 elaborated that it is useful since **participants might form an opinion** when they talk with each other. P4 added that **subjects talk more** in collaborative studies which helps to track the sentiments. The interviewee also mentioned that the analysis helps to get a **quick overview** over were participants were negative or positive and which sentiment category is the majority. P1 said visualizing sentiment is useful **if it is needed**. P3 and P4 mentioned that it is **useful in combination** with events (P3) or qualitative data and other things that describe what happened exactly (P4). Other reasons were that participants **mainly use speech** to communicate (P5) and that it is often relevant to know **how they interact** (P7).

Sentiment Visualization in Non-collaborative Studies: All participants thought that visualizing sentiment for non-collaborative studies is **useful** (P1-P8). P4 said there might be a **difference of emotion expression compared to collaborative studies**. The interviewee theorized that participants maybe show more diverse emotions when they are alone. P1 again said visualizing sentiment is useful **if it is needed**. P2 argued that it is important where in the VR environment which emotion is expressed which can be important for measurements like **presence and immersion**. P3 and P4 mentioned that it is **useful in combination** with events (P3) or qualitative data and other things that describe what happened exactly (P4). P7 said that participants have to be encouraged to **talk aloud** about their emotion and perception which has to be integrated into the study design. P8 mentioned that it can help to find out **user problems** that happened in the study. P6 said it may be helpful for **categorizing utterances quickly**.

Sentiment Visualization in Research in General: Participants thought that the visualization of sentiment is **useful** for VR research in general (P1-P8). but some also expressed that they were **unsure about the accuracy of the sentiment analysis** (P4, P7, P8). P4 mentioned the problem of **irony**. P7 noted that **manually rating sentiment is also subjective**. Participants thought it is useful to get an **overview** (P4, P6) of the parts in the virtual environment that are more positive or negative (P4). P1 mentioned that is useful **if it is needed** in the research field. P6 theorized that researchers will probably use their own sentiment analysis tools and check everything manually. Because of this P6 thinks that visualization of sentiment is **useful in studies that are not focused**

on sentiment analysis. P4 mentioned that it is **useful in combination** with qualitative data and other things that describe what happened exactly (P4).

Emotion Visualization in Collaborative Studies: All participants thought that visualizing emotion for collaborative studies is **useful** (P1-P8). P8 was **unsure about the accuracy of the emotion analysis**. P6 said it is useful to **quickly estimate** the users' feelings during an interaction. P5 thought it might help to find out if the **user is comfortable in VR** and get the **feelings that a participant does not want to verbally communicate**. P4 said there might be a **difference of emotion expression compared to collaborative studies**. Participants might express more positive emotions in a collaborative study. P3 mentioned that it might be **useful in combination** with events. P2 elaborated that emotion and sentiment analysis are similar and that emotion visualization would be useful as a **reminder** of how a subject felt at one point in time in the study. P1 mentioned that is useful **if it is needed** in the research field. P7 said it is often relevant to know **how participants interact**.

Emotion Visualization in Non-collaborative Studies: All participants thought that visualizing emotion for non-collaborative studies is **useful** (P1-P8).P1 mentioned that is useful **if it is needed** in the research field. P8 was **unsure about the accuracy of the emotion analysis**. P5 thought it might help to find out if the **user is comfortable in VR** and get the **feelings that a participant does not want to verbally communicate**. P4 said there might be a **difference of emotion expression compared to collaborative studies**. Participants might express more diverse emotions in a non-collaborative study. P2 elaborated that emotion visualization would be useful as a **reminder** of how a subject felt at one point in time in the study. P3 mentioned that it might be **useful in combination** with events. P5 and P6 elaborated that participants can **express their emotions without talking** even though they are alone (P6). P7 said that participants have to be encouraged to **talk aloud** about their emotion and perception which has to be integrated into the study design.

Emotion and Human Behaviour: Five interviewees thought that emotion visualization could help in understanding human behaviour (P1-P4, P6) P5 said **no** and P7 that it **depends** on the **accuracy** of the emotion analysis. P7 elaborated that emotion might be too complex to get precise insights about human behaviour through an external tool. P5 said that emotion analysis enough is not enough to understand human behaviour in VR since there are **more factors** involved. P2 mentioned that it would help to visualize what people felt. The participants' example was the **fear** of heights when conducting a study in high places. P1, P3 and P4 thought it would help to find the **reasons for the actions of participants**. P4 mentioned **uncertainty** as an example. P6 elaborated that it could reveal **patterns**. An example was how many participants are scared of the dark corner in a house.

Sentiment and Emotion Visualization Comparison: P3 and P4 saw an **accuracy benefit for emotion analysis** since **irony** is hard to detect with sentiment analysis. P5 and P7 thought there might be an **accuracy problem for emotion analysis**. P5 gave the example that even a very **temporary emotion** might be included in the visualization which could result in the wrong impression. P7 thought it might be **harder to categorise emotion** compared to sentiment. P4 and P7 said that **emotion is more fine-grained**. P2 and P6 saw a **benefit in combining both**. P4 mentioned that **sentiment gives a faster overview** of where negative and positive things happened. P1 explained that it **depends on the specific case**. The choice between sentiment and emotion may depend on the researchers' question or on the situation that is analysed. P1 mentioned that sentiment is more useful if a **simpler categorization** is necessary. P7 said emotion might allow **more conclusions on a participants behaviour**.

Spoken Words Visualization in Collaborative Studies: Most participants thought that the visualization of spoken words in collaborative studies is **useful** (P2-P4, P6, P8). P2 saw it as a **reminder** of what was said in the study. The participant considered the spoken words as necessary for a **complete data set** where nothing is lost. P3 mentioned that the spoken words could be used as a basis **for voice commands**. P4 elaborated that it is more interesting for collaborative studies since the **participants speak more**. P6 said it provides a **quick overview**. P3 and P8 mentioned that it is helpful to find out if participants used **the same words**. P3 thought it might be interesting which words or phrases subjects used. P8 said visualizing the words that were frequently said would make it easier to find **trends**. P7 mentioned it **might be useful** but would **need to test it**. The participant also suggested that it might be more useful to visualize collaborative studies spoken content since there might be more insights into the **quality of collaboration**. P1 and P5 expressed **problems** about visualizing spoken words in collaborative studies. They elaborated that it might be **too much information**. P1 said it might create **screen clutter**. P5 mentioned that most participants will **talk about topics that are not relevant to the study**.

Spoken Words Visualization in Non-collaborative Studies: Most participants thought that the visualization of spoken words in non-collaborative studies is **useful** (P2-P6, P8). P2 saw it as a **reminder** of what was said in the study. The participant considered the spoken words as necessary for a **complete data set** where nothing is lost. P2 also mentioned an example where participants started talking with a virtual character and how the spoken content might be an indication of how **involved a subject feels in the virtual world**. P4 and P5 said that people might **vocalize their problems in the study**. P5 explained that visualizing spoken content could be **more useful in non-collaborative studies** since participants say things that are often **relevant to the study**. P3 and P8 mentioned that it is helpful to find out if participants used **the same words**. P3 thought it might be interesting which words or phrases subjects used. P8 said visualizing the words that were frequently said would make it easier to find **trends**. P6 said it provides a **quick overview**. P3 mentioned that the spoken words could be used as a basis **for voice commands**. P7 mentioned it **might be useful** but would **need to test it**. The interviewee said that **participants need to be encouraged to talk aloud** for it to be helpful. P4 expressed **problems** about visualizing spoken words in non-collaborative studies. Participants **could talk less than in collaborative studies**, **only talk for special occasions** or **some people might not talk at all**. P1 saw it as **not useful** and would rather use a recording of a person walking through the world or read the transcript.

Other NLP Visualization Ideas: P2 suggested visualizing **changes between second and first language**. Other suggestions were **speech length** (P7), **speech speed** (P2, P7) and **voice quality** (P4, P7). P4 mentioned that **tone height** might reveal if a **sentence is a question**. P7 named **the excitement of a person** as a factor that could be visualized. P8 proposed a **line graph that visualizes the sentiment** with color coding.

5.4.5 General Questions

In-Situ Analysis Advantages and Disadvantages: P5 mentioned that an in-situ analysis tool can be helpful as a **reminder** for the data that still needs to be analysed. P5 said it provides **faster access to the data**. P6 mentioned that you have a **similar perspective as the participant** and can **view places that are hard to see on the desktop**. P2 and P7 explained that the researcher is more in the **context** of the study. P8 mentioned that it is **easier to imagine the data** compared to only looking at the numeric results. P3 said it is **more pleasant to follow a participants path** in VR.

Participants said it is **harder to read** text in VR (P2, P3) and more elaborate to **change between desktop and VR** if the researcher has to look something up (P2). Multiple people mentioned that it is **exhausting** to be in VR for a long time. P4 said that **shortcuts are harder to implement** and that **editing is harder**. P5 mentioned that there could be **less options** than on

desktop and that it is **more difficult to use**. Other disadvantages are that an **HMD is needed** (P6, P8), **pictures are hard to get** inside of VR (P7), it is less useful for **looking at specific isolated points** and it is **no desktop replacement**. There is also **less overview** (P6, P7) since the **view is restricted** (P6) and the in-situ tool **cannot replace free fly through the world** (P7).

P4 noted that a **bird view in VR could also be useful**. P3 mentioned that it **depends on the task** if the in-situ or desktop analysis should be used. P1 mentioned a **preference for only using desktop**.

In-Situ Analysis Tool Tasks: Participants would use the analysis tool to get an **overview** (P7, P8), to recognize the **context** of a study (P7) or to find **patterns** (P6-P8). P6 wanted to view and compare user **paths**. P7 would use the tool for revealing **action patterns** and P8 for getting **similarities in participant behaviour**. Participants wanted **data from audio** (P2, P5, P6) which is **speech to text** (P2), **audio events** (P2, P5) and **utterances** (P6). Interviewees would use the tool for the **visualizations** (P4, P5), **object movement** (P4) and **object interaction** (P2), **participant movement** (P2, P3) and **participant gaze** (P2, P3) and **focus group testing** (P2). P2 explained that the tool could not only be used for studies but for focus testing of storytelling in VR to see where people are looking or going or for **focus group testing** in general. P4 elaborated that the researcher or a colleague that does not know the study could use the tool as a **reminder**. P1 would **rather look at a recording** showing what the user saw but would use the tool **without a VR headset** in bird view.

Exported Information: Many interviewees wanted to be able to **export everything** from the in-situ analysis tool (P1, P2, P7) because they want to analyse outside of VR (P1) or with a free fly camera (P7). Participants expressed a need for a bird view of the visualizations (P2, P4, P5, P8) and screenshots (P2, P4). P4 wanted to export the **diagrams with the applied filters**. P2 said it would be useful to mark and **export data subsets**. Other data participants wanted to export was **gaze** (P5), **position** (P5), **speech to text data** (P3-P5), **audio** (P3, P4) and **audio events** (P5).

Participants Analysis Tools: Participants named the analysis tools **Excel** (P1, P2, P5-P8), **R** (P7), **SPSS** (P2, P5, P6, P8) and **python** (P6). P3 mentioned **none**. P8 used a 2D **heat map tool** for an eye-tracking study. P4 talked about tracking user mouse clicks with **unity analytics** in a game but did not use it to record and analyse data in VR.

Participant Logging Tools: Participants mentioned that they use **zoom recordings** (P1), **qualitative data** from interviews (P4), **surveys** (P5) and **self-made logging tools** (P2, P4, P6-P8). P4 and P8 used an **already implemented eye-tracking tool** but P8 had to **stop using it** since the calibration process was hard to explain to participants which were deaf.

General Logging Tool Feedback: P2 mentioned that the tool **should be easily portable** to unreal engine and that it is **well structured**. P4 and P5 said it is **easy to understand**. P6 mentioned it **covers what is personally needed**. P7 said the provided **data is accessible**, **covers many situations** and is **useful**.

Participants suggested a **confirm logging prompt** that informs them that they are being recorded (P3), a **need for better user experience** (P1), **documentation** (P4) and **customizability** for special cases of a study (P8).

General Analysis Tool Feedback: P2 said the tool is **comprehensive** since the information that is needed is provided. P4 found the **filters useful**. Participants mentioned that it is **easy to understand** (P2, P4, P5) and has **no unnecessary functions** (P4, P5). P7 said the tool is **good for context analysis**.

P3 mentioned the need for a **color blind mode**. P1 said that the **heat map data points are hard to find**. Participants wanted an **option for analysing outside of VR** (P1, P6, P7) because there might **not be a VR headset available** (P1, P6) or in case they want to analyse with a **free-flying camera** (P7). P4 and P7 suggested **improving the UI** and P8 expressed a **need for customizability**.

General Tool Feedback: Interviewees had two more questions to give feedback on the tool which they could not express yet in the previous questions. P6 proposed adding the **functionality to measure and display time spans**. P8 mentioned that **significance tests** in the analysis tool would be interesting. P5 remarked that good **documentation is important**.

5.5 Discussion

What requirements do researchers have for a VR logging and analysis tool? The results from the initial question asking about the participants' general expectations for a VR analysis tool show that the most of interviewees talked about analysing positional data, interactions and focus. This makes sense since VR studies often want to let people go through a virtual environment and interact with or look at elements inside of it. Some studies are more static like in the example that P7 provided of participants sitting in a chair where position may be less important but tracking participant focus is not. Even though participants talked less about getting data from audio, they provided insights that might be interesting to track or visualize. Speech content, length and point in time are things that are easy to track. Extracting information like boredom from the voice of study subjects is harder but may be beneficial for researchers to know in general. Three participants talked about physiological measures like measuring heart frequency or brain waves. This may be beneficial for researchers in general since heart frequency for example may reveal how nervous a person is but needs additional measurement tools for VR headsets. Some of this data may also be interesting but hard to interpret by researchers who are not experts in the field.

The tool seems to have fulfilled most of the basic needs of the researchers like position tracking of participants, interaction tracking and focus in the form of gaze tracking. The first two questions about which data participants are missing reveal that eye-tracking may be a desirable addition to the implemented tool and a potential requirement for VR researchers. From the general questions about what logging tools participants used it can be seen that there are still obstacles like calibration that impact the practicality of eye-tracking in research. Another reason that eye-tracking may be currently hard to consistently utilise is if the studies are done with users in their homes. Many HMDs currently have no eye-tracking capabilities which is no problem for studies in the lab but is for remote. Other data that is missing from the tool and which might be relevant is network data. Even though participants did not talk about this topic much it might help to identify problems before and after a study as mentioned by one of the interviewees. Interaction tracking is done through input actions in the unity analysis tool. One problem of this approach is that it is hard to get the interaction target only based on the logged data and without using the analysis tool. Researchers should be able to analyse this association of interaction and VR object even outside of the virtual environment.

Many interviewees wanted to have a choice where their data is saved or a mixture of both proposed methods for redundancy reasons. One of the main goals of an analysis tool should be that the logged data is usable which can be ensured through combining both remote and local saving. Interview participants elaborated on different problems and advantages of both approaches. Saving data remotely can be a problem if there are connection issues and data is lost. This is of course only a problem if the conducted study is collaborative or if the used HMD does not require a PC. Saving data locally can be problematic for standalone VR headsets since remote participants of a VR study might not have the technical knowledge to access the generated logging information from a user study. Additionally, researchers themselves have to invest more time to access and

transfer this data. While combining both approaches for redundancy seems to be the best approach there might be problems in case the used HMD cannot handle writing to their internal storage and sending data packages over the network at the same time without impacting performance.

Audio logging was desirable for most of the interviewees. It seems to be beneficial if study participants vocalize their problems or questions. Researchers can listen to the audio after the study and better understand what may have triggered the failure of a specific study task. Some interviewees talked about studies where subjects did not have a collaborative partner but still talked or gave opinions. This may be an indicator that recording audio is helpful for VR studies in general.

Most interviewees would use speech to text audio events. Automatically transcribing audio reduces workload if the researcher needs to get information about what people said. One of the experts suggested that even if speech is not the main research data it may be helpful. Researchers might be more inclined to use the spoken content of subjects if it is already transcribed and easy to search through. Going through audio is probably more time-intensive than searching for keywords in a text. One expert mentioned that only the time of speech and the speaker is needed for their purposes and because of privacy concerns. Researchers have to ask for consent and present which data they will track for their experiment. It might happen that a participant is not willing to provide his recorded voice. The researcher should be able to easily adapt which data can be recorded in the case that the audio data is not of high importance.

Participants general feedback for the presented tool and logging data in VR revealed some other topics that may be interesting even though there were not many overlapping suggestions. Tracking the users real-world movement is currently done by the tool through the "OpenXR" plugin[39]. It may be interesting for researchers to not only visualize the movement of the participants' character in the virtual environment but also their actual movement in the real world. This could reveal for example if participants often left their tracking space or bumped into real-world obstacles which could impact their performance or feeling of presence. Specifically for collaborative research, it is important that the recorded data is synchronized so it can be compared or replayed. An analysis tool should handle this automatically. The implemented tool uses Coordinated Universal Time (UTC) timestamps which avoids the problem of study participants who are in different regions. Depending on how an analysis tool generates the timestamps, inaccuracies can arise. For studies where high precision for the data tracking of multiple participants is necessary this might pose another problem. Another useful addition to this or other analysis tools is performance logging. As described by one of the interviewees if there are performance issues like frames dropped it could have an impact on VR sickness and may be interesting to research since devices can be compared with the performance data. Additionally, it may help to track down the problem a participant had in a study. One participant suggested that the VR analysis tool should be intuitive even for people with less experience in VR research. The participants were bachelor graduates or master students. VR analysis tools should aim to support both inexperienced VR programmers and students as well as people with more expertise in the field. Another interviewees requirement is fail-safe logging. Tools should be able to immediately write the tracked data or send it over a network and ensure that it is received. This is no big challenge for text and should also be considered for audio files.

A problem in the implemented tool was that some colors of the visualizations may match the background of a virtual environment. This was addressed by Kloiber et al. [26] by removing textures of the VR background. In general analysis tools should think about the virtual environment and how it can occlude or hide the projected visualizations. Furthermore, visualizations should be designed with mechanisms to either reduce or avoid overlap through e.g. clustering data points for a point cloud.

An interviewee suggested that there might be problems with finding heat map spots in the

[39] https://docs.unity3d.com/Packages/com.unity.xr.openxr@1.2/manual/index.html (last access 27.11.2021)

virtual environment if they are not marked in some way. Many participants also suggested some kind of bird view so the environment can be observed better. This reveals a general problem for in-situ analysis tools. Outlier or large amounts of data points could be hard to find especially in a large virtual environment. Such analysis tools should provide some way to find these spots or maybe even lead the analyst to them.

Which elements of the implemented logging and analysis tool are useful for a VR researcher? The sections in the questionnaire for "Analysis Tool Questions", "Logging Questions" and "General Questions" revealed which parts of the implemented analysis and logging tool are useful and where there might be a need for improvement (see appendix B).

Many participants find the recorded audio and generated audio events from the tool useful. While it may still depend on the study if the data is actually used or not it may motivate researchers if the audio data and the corresponding transcript is easily available.

The implemented events seem to cover a lot of the interviewees' needs except for eye-tracking. The elements of the line graph and point cloud were for most participants distinguishable. There is still room for improvement to increase clarity if a higher amount of participants needs to be displayed or if a color of the visualizations matches the background. Feedback for the proposed visualizations was mostly about customizability. The point cloud may benefit from using a different geometric form or labeling. The heat map may need a way to find outliers or spots with a high number of events. Many participants wanted to replay multiple animations, not just one entity. All participants found the in-situ audio replay functionality useful. One interviewee noted that the actual use of the function may depend on the research question. On the other hand, it may be beneficial to get more involved with the data and help revisit the initial context of the study. Interviewees also mentioned that it allows a researcher to check if for example, the speech recognition worked as intended. The answers show that the functionality may be beneficial in previously mentioned cases, but might not be used at all if the audio of the study is not relevant to the researchers' questions.

One of the "General Questions" aims to answer for what tasks participants would use the in-situ analysis. There was some overlap in the answers of the interviewees concerning the task of finding patterns and analysing data from audio. The question mostly reveals the variety of tasks that researchers may use the in-situ analysis tool for. Based on the answers on what information participants want to export out of the in-situ tool, it can be seen that logged data being easily accessible and usable without the tool is important. It also shows as many participants mentioned over many of the questions, that an alternative desktop exploration would make sense. Expert answers for the question concerned about in-situ analysis advantages and differences describe this well. Such tools help to analyse hard to view places and revisit the context of the study, but are also exhausting to use and may not replace a free fly through the world since the view is more restricted. These challenges provided by interviewees may be interesting to explore. This could be done for example with a mini-map of the environment or markings of significant spots where many events happened, as mentioned in the answers about visualization changes for this analysis tool (see section 5.4.3).

What are the benefits if analysis methods for visual focus, emotion recognition, sentiment recognition and speech analysis are added to a VR analysis tool? The interviews questions about sentiment, emotion and NLP analysis revealed benefits and challenges that experts in VR studies see (see section 5.4.4). All participants found the visualization of sentiment in collaborative as well as non-collaborative virtual environments useful. Interviewees explained that for collaborative studies it can function as a reminder of how participants felt or as an overview of where negative and positive sentiments happened. Compared to non-collaborative studies more data will probably be generated since participants talk more, as one of the interviewees mentioned.

This is of course only true if the collaborative study allows verbal communication or non-verbal communication through chat. Another benefit that was mentioned was the possibility of an opinion forming when study participants talk with each other. This could be relevant for scenarios where subjects express their opinion for example about an element in the virtual environment but do not have to talk about their impressions in questionnaires after their study. Their conversation and sentiments might not be revealed unless the researcher goes through the audio material and investigates. This information might not be needed by every researcher, as one of the participants mentioned, but may be relevant if the negative or positive sentiment impacts the results of the study.

One of the experts proposed that there might be more diverse emotions expressed by study subjects in a non-collaborative setting. This theory might be worthwhile to explore further. Another interesting suggestion was the potential connection of presence and immersion to sentiment. Negative and positive spots in a virtual environment could reveal where sentiment was expressed that potentially impacted e.g. presence. This may also be an interesting topic and a potential use case for the visualization of sentiment. Another application mentioned by experts is the recognition of user problems and the quick categorization of utterances. Through the logging questions about audio (see section 5.4.2) it was found that participants of VR studies also talk even though there is no collaborative partner in the virtual environment. One scenario described subjects that talked with a virtual avatar. Another reported about users giving feedback and talking about the problems they have while the study is conducted. This may be an indicator that visualizing sentiment is useful even in a non-collaborative setting. Since there is no guarantee that participants will talk in a virtual reality study without explicitly telling them to vocalize their impressions or problems it may not show every opinion or questions that participants had. The actual willingness of subjects to talk without being asked to might be an interesting research direction and may be an indicator of the usefulness of visualising sentiment. It might be necessary to encourage subjects to talk about their perception, as one of the participants mentioned. This might be a disruptive factor and may not be possible for every VR study.

One of the participants mentioned for both collaborative and non-collaborative studies the benefit of combining the visualizations with qualitative data and other components that describe what happened. Other events like interactions and movement combined with the sentiment might help understand the situation better or act as a reminder.

The questions about sentiment visualization and how useful it is for research in general showed that some interviewees had doubts about accuracy. One participant mentioned the problem of irony. This is a problem for e.g. unsupervised approaches of sentiment analysis and could be addressed with contextual valence shifters [43][50]. Low amounts of spoken content can be checked manually but this gets more time intensive if it increases. It may be that errors such as irony are drowned out the more data is generated, but in case they are not there must be mechanisms in place to efficiently check for them or solve them.

All participants found the visualization of emotion in collaborative as well as non-collaborative virtual environments useful. Experts provided similar use cases as for sentiment. It may help as a reminder of how participants felt. Like with sentiment emotion visualization may help in combination with other events to understand a situation in the study. One of the participants suggested that it may help to find out if a user is comfortable in VR and any emotion that a study participant does not want to communicate. This may also be a potential benefit of visualising emotion compared to sentiment since spoken content is not a necessity. Furthermore, the possibility of observing comfort may be useful to e.g. explain a subjects worse performance because of stress.

An expert mentioned that emotion visualization might catch what participants of a VR study do not want to express. This requires emotion recognition that is not based on spoken content and audio (see section 2.4).

This papers expert study asked for a comparison of sentiment and emotion visualization. Experts had arguments for and against the accuracy of emotion analysis. Irony might be easier to

spot. Emotions are more fine-grained than the division into positive and negative sentiment. On the other side, VR experts expressed their doubts about the accuracy of the categorization. Additionally, there might be temporary emotions that might be included in a visualization and as a result give the wrong impression. The finer categorisation of emotion analysis may be a good advantage for example when a person is scared. The subject might also provide a negative sentiment, but it may not be possible to conclude through the situation what resulted in the negative sentiment.

Most participants found the visualization of spoken words in collaborative as well as non-collaborative virtual environments useful. In this case, not all participants were positive about the visualization of words. For collaborative studies participants mentioned that too much information might be shown and will clutter the screen. It has to be mentioned that the only provided example for the visualization of words was the conceptualized word cloud, which might have been a reason for participants to mention the potential of clutter. Overloading the researcher with information should be avoided. Specifically for words, there should be filter mechanisms in place. Another interviewee said that people might talk about topics that are irrelevant to the study. While this might be a problem it might also be an opportunity. Participants may be distracted by elements in the virtual environment and try to communicate with their collaboration partner or partners for example. Finding such occurrences might be beneficial for a researcher to explain potential problems that took place in a study. However, there should be a way to filter out or to search for specific words, in the case where there are words that are of no interest to the researcher.

Another expert mentioned for non-collaborative visualization of spoken words a scenario where participants started to talk with an avatar. The impression of the interviewee was that visualising such events might help to find out how involved a participant feels in the study. This aspect might be interesting even for collaborative studies since participants might not always be together and start interacting with the surrounding environment.

5.6 Limitation

There are some limitations of the study and analysis tool that are presented. A typing error for question 2e (see appendix B) was found in the miro board while conducting the study. The words "speech to text" were written as "text to speech". It is assumed that the question was understood correctly since each one was read and explained by the interviewer and because the question was about audio events, which were previously explained in the logging explanation (see section 5.2). The replay menu screen that was presented was missing a drop down menu for the selection of a study participant or transform. The menus functionality was explained and later presented in a video.

It should also be mentioned that the replay functionality of the tool for participants and trackable transforms was not finished. The limitations of only being able to select one participant and no specific task to replay and the missing association of trackable transform to participant were mentioned in the explanation video (see section 5.2). Any questions concerning the limitations were answered after the explanation of the analysis functionalities, but there might still be an effect on what and how many opinions and insights the participants might have given. Furthermore, the interview only showed pictures of the emotion heat map and the word cloud, since they were not implemented. This might have had a similar impact on the given answers.

There were no questions in the questionnaire that specifically discussed the topic of eye-tracking visualization. It was not included because of time constrains since the number of presented questions was already large. Additionally, some questions already gave the opportunity to express their need for such a functionality. It may still be beneficial to explore the benefits and challenges the experts see for this technologies visualization and use for research, similar to the questions about sentiment, emotion and NLP. The large question pool resulted in a high average interview length for a semi-structured interview, which is recommended to last a maxi-

mum of one hour because of interviewer and interviewee fatigue [2]. This might have affected the given answers for questions later in the interview even though none of the participants requested a premature end. The study was specifically designed without participants testing the tool for the reasons explained in section 5.1. This may have had an impact on the provided feedback since participants had only the presentation information that was provided.

6 Future Work

The expert interview of this study was conducted with students who had previously conducted a VR study. This revealed interesting requirements and feedback for this group of VR researchers giving a good basis for further research. It might be interesting to conduct interviews with researchers who have more experience in the field. An added benefit of including other researchers in further studies is that the findings generated by this thesis can be extended. The interviews for the analysis and logging tool implementation and concept were designed with a very broad set of questions to generate ideas and requirements from the invited experts. This presents a base for further, more specialized research to be conducted. Future studies could for example focus on the factors emotion, sentiment, spoken content and visual focus separately.

The practicality of the proof of concept prototype used to analyse VR research in this thesis could also be validated by further studies. Similar to MRAT [34] and MIRIA [9] a case study might be necessary for the tool of this thesis to show the applicability to a wide range of VR study domains. This tool was used in the work of Reinhardt [44] to generate more in-depth data that could be analysed in future studies. A case study using this data would be interesting but is beyond the scope of this master thesis.

Similar to Kloiber et al[26] it would be beneficial to research more into the field of how visual clutter can be removed. The number of participants in a VR study increases the number of elements that can be displayed for some visualizations. Some suggestions from the interview of this thesis were, for example, adding shadows, patterns instead of monochrome colors or gradient dotted lines for the clarity of line graphs.

One of the interviewed experts mentioned the possibility of study subjects expressing more diverse emotions in a non-collaborative setting. This theory might be worthwhile to explore further with the sentiment visualization of the analysis tool or future versions that implement the emotion visualization as well as other measures for triangulation like e.g. questionnaires.

Furthermore, future iterations of this logging and analysis tool might extend its functionality. The literature review and expert interviews reveal some extension areas. One such aspect is interaction. It is currently possible to display additional event information through the point cloud visualization cubes. Event interaction for other visualizations might be useful to explore. An example of that is the path interaction from Kloiber et al. [26] where additional information is displayed from the spot that is selected or the replay position can be adjusted. The visualizations heat map and line graph are designed to be easily extendable for future interaction functionality. Another highly requested functionality is eye-tracking capability. While this is dependent on the eye-tracking hardware and software of a study participant this is no problem for studies executed in a lab where all these elements are controlled. The current concept only spoke about visualizing the visual focus of a participant at one point in time. It might be necessary to show an overview of participants eye-tracking information to observe focus points over one or multiple tasks or time frames. From a technical point of view, it might be necessary to log participant position and eye-tracking data separately since they might require a different resolution of data points. The replay functionality was well received but needs to be extended for multiple participants. Interview participants were confident about the usefulness of sentiment and emotion visualization. Some were less convinced about showing spoken content because it could be too much information and clutter the screen. Future extensions should focus on not overusing screen space when implementing a visualization of spoken content in the virtual world. While these problems were named most experts still believe it has its application scenarios which makes it worthwhile to explore. Both the implementation of emotion and spoken content visualizations are interesting extensions for the tool of this thesis based on the interview results. Another important factor that was revealed through the interviews is that the in-situ analysis might be exhausting and that there might be a need for an option to transition to a desktop alternative. Additionally, this would satisfy the required overview of the visualizations that many interview participants spoke about.

7 Conclusion

This thesis presented an implementation of a VR logging and analysis tool and additional concept ideas. A specific focus of the related work that was researched for the design of the prototype was done about other analysis tools for AR and VR that were previously created. Furthermore, a review of research about sentiment analysis, emotion analysis, visual focus, spoken content and their visualization was conducted. Concept ideas and application was evaluated through an expert interview with eight students with previous experience in VR research over zoom. The prototype was shown in video format. Any concept ideas that were not implemented were presented. The results show the feedback, requirements and ideas that participants had about the presented tool and concept. A discussion of these results under the focus of the three research questions of this thesis is presented as well as ideas for future extensions and research work. The results show that there are benefits in adding functionalities for analysing visual focus, emotion recognition, sentiment recognition and speech analysis to a VR analysis tool. Overall this thesis is a further step into designing a VR logging and analysis tool to support researchers in understanding and visualizing human behavior in their studies.

Inhalt der beigelegten CD

1. Folder Master_Thesis: Contains LaTeX version and pdf of master thesis
2. Folder Logging_And_Analysis_Tool: Contains logging and analysis tool components as a unity package (LogAndAnalysisTool.unitypackage) and in a unity project (Test_Masterthesis_Project). VRAnalysisTool_LogData directory contains the mock data of the study
3. Folder Expert_Interviews: contains all expert interview data as well as well as files used for analysis. Contains Presentation from expert study.
4. Pictures_ and_Videos: Contains all pictures used in thesis and presentations. Contains video used in study.
5. Folder References: Contains references used in thesis
6. Antrittsvortrag Visualization of Human Behaviour in VR

References

[1] Francisca Adoma Acheampong, Chen Wenyu, and Henry Nunoo-Mensah. Text-based emotion detection: Advances, challenges, and opportunities. *Engineering Reports*, 2, 7 2020.

[2] William C. Adams. Conducting semi-structured interviews. *Handbook of Practical Program Evaluation: Fourth Edition*, pages 492–505, 10 2015.

[3] Shivam Agarwal, Jonas Auda, Stefan Schneegaß, and Fabian Beck. A design and application space for visualizing user sessions of virtual and mixed reality environments. The Eurographics Association, 2020.

[4] Samar Al-Saqqa, Heba Abdel-Nabi, and Arafat Awajan. A survey of textual emotion detection. pages 136–142. IEEE, 7 2018.

[5] Mohammed Almashraee, Dagmar Monett, and Adrian Paschke. Emotion level sentiment analysis: The affective opinion evaluation. 5 2016.

[6] Soheila Ashkezari-Toussi, Mohammad Kamel, and Hadi Sadoghi-Yazdi. Emotional maps based on social networks data to analyze cities emotional structure and measure their emotional similarity. *Cities*, 86:113–124, 3 2019.

[7] T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, and T. Ertl. Visualization of eye tracking data: A taxonomy and survey. *Computer Graphics Forum*, 36:260–284, 12 2017.

[8] John Blitzer, Mark Dredze, and Fernando Pereira. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. 1 2007.

[9] Wolfgang Büschel, Anke Lehmann, and Raimund Dachselt. Miria: A mixed reality toolkit for the in-situ visualization and analysis of spatio-temporal interaction data. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021.

[10] Lea Canales and Patricio Martínez-Barco. Emotion detection from text: A survey. pages 37–43. Association for Computational Linguistics, 2014.

[11] Viviane Clay, Peter König, and Sabine U. König. Eye tracking in virtual reality. *Journal of Eye Movement Research*, 12, 4 2019.

[12] Christopher Collins. Docuburst: Document content visualization using language structure. *Proceedings of IEEE Symposium on Information Visualization, Poster Compendium*, 2006.

[13] Jan Deriu, Maurice Gonzenbach, Fatih Uzdilli, Aurelien Lucchi, Valeria De Luca, and Martin Jaggi. Swisscheese at semeval-2016 task 4: Sentiment classification using an ensemble of convolutional neural networks with distant supervision. pages 1124–1128. Association for Computational Linguistics, 2016.

[14] Susana Dinis, Emília Duarte, Paulo Noriega, Luís Teixeira, Elisângela Vilar, and Francisco Rebelo. Evaluating emotional responses to the interior design of a hospital room: A study using virtual reality. volume 8014 LNCS, pages 475–483. Springer, Berlin, Heidelberg, 2013.

[15] Thorsten Dresing and Thorsten Pehl. *Praxisbuch Interview, Transkription & Analyse Anleitungen und Regelsysteme fur qualitativ Forschende*. dr dresing & pehl GmbH, 2012.

[16] Paul Ekman. Basic emotions. *Handbook of cognition and emotion*, 98:45–60, 1999.

[17] Anna Felnhofer, Oswald D. Kothgassner, Mareike Schmidt, Anna Katharina Heinzle, Leon Beutl, Helmut Hlavacs, and Ilse Kryspin-Exner. Is virtual reality emotionally arousing? investigating five emotion inducing virtual park scenarios. *International Journal of Human Computer Studies*, 82:48–56, 5 2015.

[18] Oliver Guhr, Anne-Kathrin Schumann, Frank Bahrmann, and Hans-Joachim Böhme. Training a broad-coverage german sentiment classification model for dialog systems. pages 1620–1625. European Language Resources Association, 5 2020.

[19] Pavol Harar, Radim Burget, and Malay Kishore Dutta. Speech emotion recognition with deep learning. pages 137–140. Institute of Electrical and Electronics Engineers Inc., 9 2017.

[20] Florian Heimerl, Steffen Lohmann, Simon Lange, and Thomas Ertl. Word cloud explorer: Text analytics based on word clouds. pages 1833–1842. IEEE Computer Society, 2014.

[21] Mengdie Hu, Krist Wongsuphasawat, and John Stasko. Visualizing social media content with sententree. *IEEE Transactions on Visualization and Computer Graphics*, 23:621–630, 1 2017.

[22] Xin Kang and Fuji Ren. Sampling latent emotions and topics in a hierarchical bayesian network. pages 37–42. IEEE, 11 2011.

[23] Anne Kao and Steve R Poteet. *Natural Language Processing and Text Mining*. Springer Science & Business Media, 2007.

[24] Lakshmish Kaushik, Abhijeet Sangwan, and John H. L. Hansen. Automatic sentiment extraction from youtube videos. pages 239–244. IEEE, 12 2013.

[25] Diksha Khurana, Aditya Koli, Kiran Khatter, and Sukhdev Singh. Natural language processing: State of the art, current trends and challenges. 8 2017.

[26] Simon Kloiber, Volker Settgast, Christoph Schinko, Martin Weinzerl, Johannes Fritz, Tobias Schreck, and Reinhold Preiner. Immersive analysis of user motion in vr applications. *The Visual Computer 2020 36:10*, 36:1937–1949, 8 2020.

[27] Kostiantyn Kucher, Carita Paradis, and Andreas Kerren. The state of the art in sentiment visualization. *Computer Graphics Forum*, 37:71–96, 2 2018.

[28] Peter J. Lang. The emotion probe: Studies of motivation and attention. *American Psychologist*, 50:372–385, 5 1995.

[29] Jia Zheng Lim, James Mountstephens, and Jason Teo. Emotion recognition using eye-tracking: Taxonomy, review and current challenges. *Sensors*, 20:2384, 4 2020.

[30] Bing Liu and Lei Zhang. A survey of opinion mining and sentiment analysis, 8 2012.

[31] François Mairesse, Joseph Polifroni, and Giuseppe Di Fabbrizio. Can prosody inform sentiment analysis? experiments on short spoken reviews. pages 5093–5096, 2012.

[32] Marlen, Birgit Mack, and Ortwin Niederberger Renn. *Fokusgruppen in der empirischen Sozialwissenschaft*. VS Verlag für Sozialwissenschaften, 2012.

[33] Myriam Munezero, Calkin Suero Montero, Erkki Sutinen, and John Pajunen. Are they different? affect, feeling, emotion, sentiment, and opinion detection in text. *IEEE Transactions on Affective Computing*, 5:101–111, 4 2014.

[34] Michael Nebeling, Maximilian Speicher, Xizi Wang, Shwetha Rajaram, Brian D. Hall, Zijian Xie, Alexander R.E. Raistrick, Michelle Aebersold, Edward G. Happ, Jiayin Wang, Yanan Sun, Lotus Zhang, Leah E. Ramsier, and Rhea Kulkarni. Mrat: The mixed reality analytics toolkit. pages 1–12. Association for Computing Machinery, 4 2020.

[35] Marcus Nyström, Richard Andersson, Kenneth Holmqvist, and Joost van de Weijer. The influence of calibration method and eye physiology on eyetracking data quality. *Behavior Research Methods*, 45:272–288, 3 2013.

[36] Andrew Ortony, Gerald L Clore, and Allan Collins. *The cognitive structure of emotions*. Cambridge university press, 1990.

[37] Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and Trendső in Information Retrieval*, 2:1–135, 2008.

[38] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. pages 79–86. Association for Computational Linguistics, 2002.

[39] Jose Costa Pereira, Jordi Luque, and Xavier Anguera. Sentiment retrieval on web reviews using spontaneous natural speech. pages 4583–4587. IEEE, 5 2014.

[40] Thies Pfeiffer. Measuring and visualizing attention in space with 3d attention volumes. pages 29–36. ACM Press, 2012.

[41] Thies Pfeiffer and Cem Memili. Model-based real-time visualization of realistic three-dimensional heat maps for mobile eye tracking and eye tracking in virtual reality. volume 14, pages 95–102. Association for Computing Machinery, 3 2016.

[42] Robert Plutchik. A general psychoevolutionary theory of emotion, 1980.

[43] Livia Polanyi and Annie Zaenen. Contextual valence shifters, 2006.

[44] Karlheinz Reinhardt. *Visualization of Movement in Social Virtual Reality (MSc thesis in progress)*. Ludwig-Maximilians-Universität München, 2021.

[45] Robert Remus, Uwe Quasthoff, and Gerhard Heyer. Sentiws - a publicly available german-language resource for sentiment analysis. 1 2010.

[46] Giuseppe Riva, Fabrizia Mantovani, Claret Samantha Capideville, Alessandra Preziosa, Francesca Morganti, Daniela Villani, Andrea Gaggioli, Cristina Botella, and Mariano Alcañiz. Affective interactions using virtual reality: The link between presence and emotions. *Cyberpsychology and Behavior*, 10:45–56, 2 2007.

[47] James A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178, 1980.

[48] James A Russell and Albert Mehrabian. Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 11:273–294, 9 1977.

[49] Hari Singh and Jaswinder Singh. Human eye tracking and related issues: A review. *International Journal of Scientific and Research Publications*, 2, 2012.

[50] Mohammad Soleymani, David Garcia, Brendan Jou, Björn Schuller, Shih Fu Chang, and Maja Pantic. A survey of multimodal sentiment analysis. *Image and Vision Computing*, 65:3–14, 9 2017.

[51] Xiangyan Sun, Yanghua Xiao, Haixun Wang, and Wei Wang. On conceptual labeling of a bag of words. pages 1326–1332, 2015.

[52] Santawat Thanyadit, Parinya Punpongsanon, and Ting Chuen Pong. Demonstration of observar: Visualization system for observing virtual reality users using augmented reality. pages 1–2. Association for Computing Machinery, Inc, 12 2020.

[53] Florence Ying Wang, Arnaud Sallaberry, Karsten Klein, Masahiro Takatsuka, and Mathieu Roche. Senticompass: Interactive visualization for exploring and comparing the sentiments of time-varying twitter data. pages 129–133. IEEE, 4 2015.

[54] Hwanmoo Yong, Jisuk Lee, and Jongeun Choi. Emotion recognition in gamers wearing head-mounted display. pages 1251–1252. Institute of Electrical and Electronics Engineers Inc., 3 2019.

[55] Dong Yu and Li Deng. *Automatic Speech Recognition*. Springer London, 2015.

[56] Cem Yuksel, Scott Schaefer, and John Keyser. On the parameterization of catmull-rom curves. page 47. ACM Press, 2009.

[57] Amir Zadeh. Micro-opinion sentiment intensity analysis and summarization in online videos. pages 587–591. Association for Computing Machinery, Inc, 11 2015.

[58] Jian Zhao, Liang Gou, Fei Wang, and Michelle Zhou. Pearl: An interactive visual analytic tool for understanding personal emotion style derived from social media. pages 203–212. IEEE, 10 2014.

[59] Rui Zhao and Kezhi Mao. Fuzzy bag-of-words model for document representation. *IEEE Transactions on Fuzzy Systems*, 26:794–804, 4 2018.

A Expert Interview Pre-questionnaire

1. Identification key (Use your last name. This is just used to identify which persons interview questions belong to which pre-questionnaire. The names will not me used in my master thesis)

2. What is your highest level of education and on which subject? (e.g. Master Human Computer Interaction. If you are still studying, write e.g. studying Master Human Computer Interaction.)

3. What is/are your field(s) of expertise in VR research? (e.g. VR locomotion, VR eyetracking etc.)

4. What is/are your most recent VR study/studies about? (e.g. VR locomotion, VR eyetracking etc.)

B Expert Interview Questions

1. What functions would you expect from an analysis tool for vr studies?
2. Logging Questions:
   (a) Are the proposed logging events sufficient to record the data you need? If not: Which events would you add/remove/change?
   (b) Are the proposed logging events sufficient to record the data you need? If not: What Data do you need?
   (c) Do you prefer saving logging data locally? Why? Why not?
   (d) Do you prefer saving logging data on a remote server? Why? Why not?
   (e) Is the recording of audio useful for your research? Why? Why not?
   (f) Would you use the Audio Events generated from speech to text? Why? Why not?
   (g) Additional ideas/remarks/requirements for logging.
   (h) Additional ideas/remarks/requirements for logging tool.
3. Analysis Questions
   (a) Analysis Tool Questions:
      i. Were you able to distinguish the line graph paths? If not, why?
      ii. Did you recognize the type of events? If not, why?
      iii. Is there something you would add/change/remove about the line graph representation?
      iv. Is there something you would add/change/remove about the point cloud representation?
      v. Is there something you would add/change/remove about the heat map representation?
      vi. Is there something you would add/change/remove about the replay functionality?
      vii. Are there other visualizations you need or would find useful? Event Interaction:
      viii. Event Interaction: Which interaction would you like to add/change/remove?
      ix. Was the replay functionality of voice recordings helpful? Why? Why not?
      x. Would you like to edit events inside or outside of vr? e.g. when a voice event failed.
      xi. User Interface: What filter categories would you like to add/change/remove?
      xii. Was there something you did not understand in the User Interface?
   (b) Sentiment, Emotion, NLP Questions:
      i. Could the visualization of sentiment be useful for collaborative virtual reality studies? Why? Why not?
      ii. Could the visualization of sentiment be useful for non-collaborative virtual reality studies? Why? Why not?
      iii. Is visualization of sentiment for vr studies useful for research? Why? If not, what has to change for it to be useful?
      iv. Could the visualization of emotion be useful for collaborative virtual reality studies? Why? Why not?
      v. Could the visualization of emotion be useful for non-collaborative virtual reality studies? Why? Why not?

vi. Would the addition of emotion visualization help in understanding human behaviour? If yes how? If no why not?
vii. Can you see benefits or problems compared to sentiment visualization?
viii. Could the visualization of spoken words be useful for collaborative virtual reality studies? Why? Why not?
ix. Could the visualization of spoken words be useful for non-collaborative virtual reality studies? Why? Why not?
x. Are there other aspects of Natural Language Processing which could be useful to visualize?

(c) General Questions:

i. Are there advantages and disadvantages in this tools in-situ visualization compared to a desktop analysis?
ii. For what tasks would you use this in-situ analysis tool?
iii. What information would you like to export outside of the in situ visualisation?
iv. What analysis tools do you use when conducting a VR study?
v. What logging tools or methods do you use when conducting a VR study?
vi. Do you have any other general feedback about the logging tool?
vii. Do you have any other general feedback about the analysis tool?
viii. Are there other things you would like to add to, change or remove from the tool?
ix. Other remarks.