Applying diegetic cues to an interactive virtual reality experience

1st Thomas Beck
Dept. for Media Informatics
University of Munich (LMU)
Munich, Germany
tmsbeck@gmail.com

2nd Sylvia Rothe
Dept. for Media Informatics
University of Munich (LMU)
Munich, Germany
sylvia.rothe@ifi.lmu.de

***Abstract*—Gaze and attention guiding for application in Cinematic Virtual Reality (CVR) has been a research topic with many contributions in recent years. One promising approach coming out of this research, are hints or cues that are placed within the context of the scene that is currently being viewed in VR. These hints are called "diegetic cues". However, this approach, as many others, does not account for the possibility of interaction that modern VR technology brings with it. These interactions have the potential to move the viewer from a position of "impartial bystander" to one of "active participant" opening new possibilities for telling stories of various nature (e.g., fictional, or historical). This work, is the attempt to apply research from the field of attention guiding using diegetic cues to an interactive and narrative VR experience that was created using the Unreal Engine and newest VR-Technology. It was also attempted to use an alternative approach to narrative theory than the one being used in traditional filmmaking to explore the possibility of interactive storytelling. As a result, strong indications were found that the two diegetic cues that were implemented successfully enhance the experience of VR users in the interactive environment. The reception of the employed narrative approach also was generally positive.**

***Index Terms*—Virtual Reality, directing gaze, guiding attention, narrative theory, storytelling, spatial sound**

I. Introduction

With Cinematic Virtual Reality (CVR) filmmakers and researchers have started to bring the world of narrative and storytelling into the new medium of Virtual Reality (VR). A common definition of CVR found in research [1]–[3] is, that it is a branch of VR that focusses on virtual environments produced by omnidirectional camera systems, also called 360-videos. Computer generated imagery is sometimes excluded from this definition entirely [2] despite its heavy use in modern day movie productions and very successful computer-animated movies. Another definition found with members of the games-industry is that CVR "[...] essentially covers the many approaches where virtual reality content appropriates or employs filmmaking methods to deliver narrative experiences" [4] which is a more generalized view. For the purpose of this paper we prefer the latter definition as it includes computer-generated imagery as a possible source for a virtual environment. However, both definitions still treat the viewer as an observer or bystander instead of truly immersing them in the world of the story they are experiencing, which is very much the traditional view on narrative. An approach more suited to the full potential of the medium - including interaction - could move the viewer from observer to participant and open a new branch of immersive storytelling. As this approach faces similar problems as CVR in general, the solutions found so far might apply to it as well. However, traditional narrative theory does not account for interaction being a possibility and as such, the viewer, now participant might feel ignored by stories created using the traditional methods.

In this paper we propose combining existing gaze- and attention guiding methods from research in the field of CVR with storytelling as it can be found in live-action roleplaying games and narrative-driven videogames to solve the problems created by moving storytelling and interactions into the virtual realm simultaneously. In this approach, we will draw from the narrative structure of the popular tabletop RPG, Dungeons & Dragons 5th Edition and use audiovisual cues within the context of the scene, so-called diegetic cues, to guide the users attention towards important elements that can be used to progress the narrative while leaving the choice of narrative direction in the hands of the user. This approach differs from the classic definition of both CVR and games as it settles somewhere in between these two forms of narrative mediums. It provides interactivity and inclusivity like games but is more focused on the narrative thread created by an author like in movies.

To check the potential of our approach, we created a prototype project using Unreal Engine 4.25 and various assets provided within the limits of the Unreal Creators EULA. This project is made up of a scene set inside the chamber of a magic-user. The protagonist (the user) is trapped in this room for narrative reasons which are explained to the user by a narrator, speaking from the point of view of the character the user is about to take over. In an "Escape-Room"-like scenario the user then must solve a task with the objects available in the room. Cues are attached to objects of relevance and are supposed to guide the user towards them yet leave the decision of which solution to pick in the hands of the user. The potential of the approach is then evaluated by analyzing a user-study where participants were asked to run through two versions of

978-1-6654-3886-5/21/$31.00 ©2021 IEEE

the prototype described above. One with cues and one without.

The resulting data and analysis showed indications that the diegetic cues used in the prototype can guide users to objects of importance and, that especially novice users of VR found them to enhance their experience. We also found that the opposite seemed to be the case for more experienced users of VR. Additionally, we found that while the applied cues did not have a large effect on the feeling of presence and immersion, participants stated that they paid more attention to the real world in the version without cues than in the version with cues. The approach used for the narrative was also well received and many participants voiced that they would like to experience something similar in their leisure time as well, like visiting a cinema or arcade.

II. Foundations and Related Works

A. Narrative Theory

A story can be presented in a variety of ways and media. It can be told, written down, acted out in a theater, on the silver screen or television, played out in a videogame or even sung. The perspectives are as varied as the media as well. It can be told in the perspective of the protagonist (first-person) or in the one of a bystander (third-person), by an all-knowing narrator or one with limited knowledge, in a strictly linear or a more open non-linear, or even branching fashion, directly or through the design of the world surrounding the reader/viewer/player and in many other ways. While each media makes use of a different selection of techniques and narrative tools - for which we adapt the definition of Sylvester: "[narrative tools are] some device used to form a piece of a story in a player's mind" [5] - narrative in traditional media like theater, literature, cinema or TV has been well researched. It is with newer media such as games (both videogames and tabletop games) or VR where the traditional approach to narrative theory reaches its limits. This is in part due to the addition of interactions between the narrative authority and the ones perceiving the narrative. In the following we will have a look at how the traditional approach is defined, present its limits, and then see how tabletop- and videogames have worked around these limits, which we then will use for our project.

1) Traditional Approach: The traditional approach to narrative theory goes back to Aristotle and later Plato and was originally applied and observed in ancient Greek and Roman theater [6]. In its original form it can be split into the concepts of the telling of a story (Diegesis) and the showing of a story by characters (Mimesis) [6]. However, through the ages and the appearance of new forms of narrative media, which increasingly became mimetic in nature (e.g. Theater or Cinema), the term diegesis in modern film theory had been reinvented in the early 1950s to refer to the world in which a story takes place [6]. At the core of this concept the notion that narrative can only exist "[...] if it can be defined as artefact, essentially the output of the authoring process" [7] still stands. This goes back to the suggestion that the point in time of the conception or writing of the narration by the author and the time when it is presented to an audience is different. In cinema or television this happens by using the camera and a variety of angles and settings to get the vision of the author across to the viewer, making the camera an ideal observer which is controlled by the will of the author [7]. As such, this definition of narrative- and film theory stands in conflict to the immersive nature of VR and its possibilities of freedom, interactions, participations, and influences. It does not account for a viewer making use of any of these factors while listening to the author or following the plot [7]. Furthermore, since the control of the camera now lies in the hands of the viewer and no longer can serve as an instrument of the author, a secondary conflict is created. These conflicts are what makes VR more comparable to interactive media like improvisational theater or games rather than traditional film. Yet, in narrative virtual reality projects that exist today the traditional approach still is the one that is being used predominantly. They take place in either a virtual environment or a recorded real-life one, yet the user is limited to looking around. Moving around inside the scene is rarely intended, so is interaction, and there is no way to influence the path of the story that is being told. The viewer either takes on the perspective of a neutral 3rd person observer, similar to the concept of an invisible witness in classic film theory, or is put into the perspective of one of the participating characters, but without knowledge of their thoughts and feelings. In the case of 360-video, the viewer is also sometimes put into the position of a camera that flies through a sequence of shots and environments, much like a cameraman riding on a dolly track.

2) Collaborative Storytelling: An alternative approach can be found in live-action roleplaying games (LRPGs) like tabletop or pen-and-paper roleplaying games (TTRPG) or live-action roleplaying (LARP) and improvisational theater. Aylett and Louchart [7] suggest in their work that this approach, due to it's interactive nature might be better suited to the medium of VR as it takes audience participation into account. In LRPGs the role of the author is covered by the so-called Game Master (GM) or sometimes Dungeon Master (DM). While the name can differ between game systems, the role largely remains the same. It is described by the popular TTRPG *Dunegons & Dragons: 5th Edition* (D&D 5e) in the following way:

"The Dungeon Master (DM) is the creative force behind a D&D game. The DM creates a world for the other players to explore, and creates and runs ***adventures*** *[emphasis in the original] that drive the story. [...] As a storyteller, the DM helps the other players visualize what is happening around them, improvising when the adventurers do something or go somewhere unexpected. As an actor, the DM plays the roles of the monsters and supporting characters, breathing life into them. And as a referee, the DM interprets the rules and decides when to abide by them and when to change them."* [8].

Even though the GM is creating the world and adventures within it and as such is the inital author, the actual story of such an adventure is heavily influenced by the decisions the players make and communicate. This communication leads to the GM improvising and adapting results and paths on the spot,

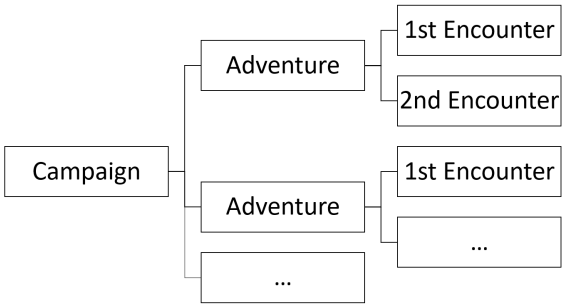

Fig. 1. The structure of a typical Dungeons & Dragons campaign. Own representation based on [8].

resulting in a unique experience every time the adventure is played. While this improvisational approach in its entirety is challenging to implement on a technical level now, one can still draw inspiration from the general structure of such an adventure and apply it to an episodic VR experience.

Speaking from personal experience the overall structure of a game like D&D 5e can be broken down into smaller story-telling chunks that are tied together by in-between activities and/or an overarching story. The smallest of these chunks - an encounter - can be a multitude of things: combat, exploration, social matters, traveling, shopping, puzzles, and many more. Multiple encounters joined together make up an adventure. Multiple adventures in turn make up a so-called campaign [8] (Fig. 1). To illustrate this structure with an example, we will take a look at the popular fantasy series *The Lord of the Rings* and assume the actions of each character within are their own and not predefined by an author. The trilogy as a whole could then be seen as a campaign with its ultimate goal being the destruction of the one ring and defeat of Sauron. Each movie on its own represents an adventure that is part of the larger, overarching narrative. And scenes within each movie in turn can be classified as encounters. The meeting between the wizard Gandalf and Bilbo at the start of the first movie is a social encounter, the fight between Frodo and the Ring Wraith later is a combat encounter. And so on.

In the tabletop world, these encounters play out a little different than on the silver screen, simply because what we just assumed in the previous example is now actually the case. An actual D&D encounter might look like this: The players/their characters found themselves locked inside a cave or room and must find a way out. The GM would illustrate and explain the visuals and feeling of the location, then await their input in form of various activities. Let's say, they find and pull a lever. Once the players have stated their course of action the GM will explain the chain of events that will take place because of these actions, before again giving the players the opportunity to react to the new situation. The GM then reacts to their new input and so on. Out of the resulting chain of decisions and reactions a unique and inclusive story is produced in collaboration between the GM and the players.

3) World- and Emergent Narrative: Since the presented experience in VR is based on the same technology used in videogames, it also makes sense to look at common narrative practices used in them. For this we will follow the categorization of [5] and in particular the categories of World Narrative and Emergent Story as they also play a role in the design of LRPGs experiences.

World Narrative describes the presence of narrative in the bounds of the world itself. It tells the story of the people, places, history, legends, and all other narrative elements that make up the setting [5]. This can happen through environment design on a larger scale, the placement of localities and documents or remnants of recent events. This has the advantage that the author retains control over what is being told, without the player being able to interfere with the event, while not limiting the player in their agency. Plot points, their progression and consequences are simply present in the world and can be discovered independently from one another. Each of which allows the player to draw a conclusion about what must have happened, even if they are discovered in a non-linear or incomplete fashion [5].

An Emergent Story on the other hand is described as a *"[...] story that is generated during play by the interaction of game mechanics and players"* [5]. It is the stories that transcend the virtual and find their way into the real world. It is the stories that players tell their friends and look back to even years later. Examples for this are stories told by players of older Massively Multiplayer Online games (MMO) and Dungeons & Dragons campaigns. They might not accurately represent the events that took place, but these events still created a lasting experience for those who experienced them. As such an experience is the goal of this project emergent narrative was also considered. However, we cannot forget, that it is not part of the story written by the author and as such most likely will not offer many lessons to learn from regarding how produce narrative, but it might still be worth it to keep this type in mind when designing stories for VR as the combination of the virtual and the real is a core part of VR.

Game Designers also have a lot of experience dealing with different story structures. While movies almost exclusively follow a linear approach, games are known to often dive into branching and converging storylines that allow for player-influence and decisions to varying degree [5]. Lessons from game narrative in this field could help to make stories told in VR more immersive and bring the viewer from the position of an invisible witness into the position of a true participant.

B. Interaction Methods

For realizing interaction methods in storytelling, natural techniques are needed which not disturb the experience. Interaction techniques can be classified in navigation (travel/wayfinding), selection, manipulation and system control [9]. One main interaction in storytelling is selecting and activating areas. Such a selecting process can be initiated consciously or unconsciously by the user or triggered by the system. Head or eye movements, gestures and even sensor data are conceivable for the selection process [10]–[12]. Input and output devices are important components for interactive VR

applications. In some cases, a controller device can support the experience by having a shape like a referenced object in the virtual environment [9]. Input devices can be classified into continuous-input devices (e.g. for tracking) and discrete-input devices (e.g. for pressing a button) [9]. Continuous-input devices are qualified for processing modulations or cursor movements. Discrete-input devices signal an event, such as to commit a selection. Often, input devices combine both. For example, a mouse movement is continuous and pressing the mouse button is discrete. The input device for story telling can be part of the virtual world. Tracking the head or the eyes are continuous-input methods which are suitable for pointing the viewing direction. For activating the selection, a discrete signal is needed. In our study the selection is activated by either a gesture input on a controller or the position of the participant in the virtual space.

C. Guidance Methods

Cinematic elements such as sounds, lights and movements draw the viewer's attention [13] in movies. In the literature [14]–[16] several methods for guiding the viewer are explored for non-VR environments, such as salient objects, sounds, lights, or moving cues. In VR, the viewing direction can be freely selected, so that important details may be outside the viewer's field of vision. Syrett et al. [17] have discovered that some viewers feel distracted by the freedom to choose the viewing direction. In their experiments, it happened that important parts of the storyline were missed. This can be unproblematic for some VR experiences: The user discovers a story world created by the author that does not require any additional guiding methods. In other story constructs it is important not to overlook certain details and the viewers should be guided in a discreet manner so that they can relax and enjoy the application. In such cases, guiding methods can improve the user experience [18].

D. Diegetic Cues

Diegetic elements belong to the narrative world. The concept of diegesis is often used in film theory for music and other sounds. Diegetic music in a film is part of the story. It can be heard not only by the audience (like film music) but also by the characters. Examples are: music from a radio in a film or music from musicians who are film characters. A cue can draw attention to a target and can have different characteristics and positions. Posner [19] showed that viewers find a target more quickly if the cue is a feature of the target (e.g., a colored border). A cue that is not positioned on the target (e.g. an arrow pointing towards the target) takes more time to process. Posner introduced the terms exogenous and endogenous. Exogenous cues are stimulus-controlled and work automatically, for example a flash of light that attracts attention. Such cues cause a reflex-like orientation, are positioned on the target and can also be auditory or haptic [20]. They work as a bottom-up process. Since the reaction to such cues is reflex, they work quickly. However, if there is no interesting target information, the attention is short-lived. Endogenous cues are targeted and voluntary [20]. Often, they are based on a sign that indicates where to look and first require interpretation. Even if goal-directed guiding works more slowly, it improves the processing of the event [19] and the information can be maintained for longer periods of time. Yarbus [21] showed that the type of eye movement depends on the task. In his experiment, participants saw the same scene after being asked different questions. The eye movements differed significantly.

Following these insights, we consider the cues used in this paper diegetic cues.

1) Spatial Audio: Spatial Audio can be used for solving tasks in virtual and augmented reality. The use of spatial sound improves the results in search and navigation tasks [22], [23]. Van der Burg, Olivers, Bronkhorst and Theeuwes [24] showed that audio cues (pop) synchronized to a salient visual cue (pip) reduces the search time, even if the audio cue does not have any location information. Hoeg, Gerry, Thomsen, Nilsson and Serafin [25] expanded this experiment to virtual reality with sound cues from the same direction as the visual cue. They demonstrated that binaural cues lead to shorter search times, even though the visual cue was not always visible at the moment the audio cue was presented. In the experiments, the participants were given a search task in an abstract VR environment. Spatial sound also leads to a higher level of presence [26], [27] and increases the sense of place [28]. Similarly, Brown, Sheikh, Evans and Watson [29] connected several cues (motion, gestural and audio cues) to the main character of a scene. The head orientation was recorded and the percentage of people who had seen the target over time was evaluated. In their experiments, the cues with an audio component were proven to be more helpful than just visual cues, even if the sound was not fully spatialized. The results were displayed by diagrams showing the time for seeing the target. In histograms, tables and diagrams specific values were presented. In our study and in contrast to [24], [25], we move closer to a real cinematographic setting by using a realistic scene instead of abstract symbols and by not giving a concrete task to the participants but letting them choose freely what to do next.

The audio cues used in this paper were spatialized to make use of these findings.

2) Conspicuously Light Patches: The term "Conspicuously Light Patch" (CLP)[1] is borrowed from a trope of the same name that is often found in cartoons of the golden age of animation. Originally it describes an element in the background of a cartoon that is suspiciously different from its background, usually lighter in color or saturation, making it easy to guess as a viewer that this object is going to be important soon. As this term perfectly describes the method used to apply a diegetic cue to the books found in this project, it makes sense to adapt it for this purpose (Fig. 2). The change in color/hue was chosen as a cue, as it is a pre-attentive cue and as such has the potential to guide the user without them consciously noticing the presence of the cue in the first place.

[1]https://tvtropes.org/pmwiki/pmwiki.php/Main/ConspicuouslyLightPatch

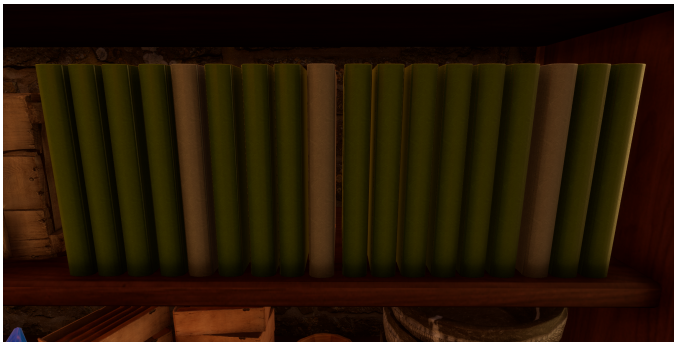

Fig. 2. A section of the bookshelf with cues activated. The lighter colored books contain the information needed to solve the task given in the scenario, the darker/normal colored books contain a placeholder text. If the player would step back, the highlighted books would return to the darker color.

This is important to uphold the suspense of disbelief and in turn the immersion of the user in the virtual environment. Another reason for this choice is that similar visual cues had already showed promising results in leading the users attention in previous research projects, e.g. [30].

E. Inspirations

Several research topics and media influenced the idea behind this project heavily. On the scientific side of things, the work by [7] and [30] inspired this approach to create an immersive, narrative VR Experience the most. From the work of [7] we adapt the proposal of letting the viewer actively take part in the experience, using TTRPGs as a role model. From [30] we adapt the implementation of diegetic cues as the predominant guiding method in the virtual environment as they show potential to keep the viewer as immersed as possible.

In terms of media, the biggest inspiration for this project is the popular TTRPG *"Dungeons & Dragons"* (D&D), one of the most popular LRPG-Systems on the market today. From it we adapt lessons around the structure and principles of storytelling present within this type of game. In this adaption, we take on the mantle of DM to a certain extend and are responsible for setting up the scene and basic narrative as an introduction to the situation the player will find themselves in. The narrative will be implemented statically in the test environment, so every participant will receive the exact same wording.

Another big inspiration are narrative-driven games such as *Life is Strange* or *Detroit: Become Human* in how they handle a branching story with the possibility of the player influencing the route the story takes. Another point of inspiration from games lies in the realm of world design and world narrative (e.g. *Nier: Automata*, *The Witcher 3: Wild Hunt*), as it plays a big role in getting subtle information about the world and setting across to the player. From games, we adapt the use of world narrative to provide information about a virtual world to the player through showing, rather through telling.

III. User Study

The prototype was implemented using Unreal Engine 4.25 using available assets under the Unreal Creators EULA (Free Asset Packs and Quixel Megascans) as well as custom made assets and materials. Once the prototype was finished, a user-study in the form of an experiment was conducted to explore whether the implemented diegetic cues would help guiding the users towards the solution or don't have any effect. For this, two versions of the scene representing the study of a wizard or sorcerer (Fig. 3) were built. One contained books and crystals with the cues active and the other contained objects without them active. In addition, the location and contents of the books containing information about the spell combinations and the crystals was changed in the version without cues, to prevent a participant of the study solving the puzzle by memory from a previous attempt. In addition to the scenes, a survey was developed and implemented.

A. Method

Due to an expected low participation rate caused by the ongoing COVID-19 pandemic and the restrictions for meetups and traveling that came with it, a within-subjects design was chosen. Each participant was led through the experiment following a fixed set of steps, so that there would be as little differences in the experience as possible. The participants were asked to answer some general questions, such as previous experience with VR technology and socio-demographic data before the experiment would start. They then were given the opportunity to familiarize themselves with the provided HMD, controllers and how to use them to interact with the different objects that they were about to encounter. Once the participants stated they were familiar with the controls, they went through both versions of the project. After the completion of each test run, the participants were asked to answer a section of the questionnaire containing questions specific to the scenes. The array of questions was the same for both scenes. After both versions were tested the participants were also asked to answer a final section of the questionnaire, containing questions about the overall experience. The order in which the participants experienced the scenes was swapped with each participant. This was done to ensure that any influences that would be

Fig. 3. The completed scene of the wizard study. Including the "Spirit", books, crystals and other setpieces to create an immersive experience.

captured from experiencing any of the two versions before the other, were equalized. In addition to the questionnaire, a form of logging was implemented into the project itself. These log entries contained data about the point in time an interaction took place (in seconds) since the task was started. This way, the time a participant needed to complete the task in each version could be measured without human interference (e.g., reaction time).

The experiment was conducted in multiple sessions in July 2020 under strict hygiene precautions. Another two participants were able to undertake the experiment in late August. A total of 20 participants took part in the study. Two participants were removed from the dataset before the evaluation, due to a previously unknown language-barrier. The remaining participants were between 22 and 65 years old (Avg. 39 years), 77.8% of participants stated to be male, 22.2% to be female. On average, the participants stated to be not very experienced with the use of VR-technology (Avg. 2.44 on a scale of 1 to 5, with 1 = no experience, 5 = expert-level experience).

B. Results

As a first step, we looked at how immersed or present the participants felt in the scenes. This was measured with two sets of questions: one covering the feeling of presence of the user inside the scene and one covering immersion more directly. The four items of the question regarding presence were selected from the igroup presence questionnaire (IPQ) [31] with slight alterations to their wording to fit the scale used in the survey. The five items regarding immersion were in part inspired by the survey used in the work of [32] and in part developed using personal experience and commonly voiced criticism of VR in its current state. The participants were asked to rate each item of both questions on a Likert-type scale with five levels, where 1 = strongly disagree and 5 = strongly agree.

For easier comparisons we formed an index for each scene with all items related to the perception and feeling of the participant of the virtual world. Before making a final decision however, we check these items for reliability using Cronbach's Alpha (ρ_T). For the first scene this test gives us a value of ρ_T = 0.799 and for the second scene a value of ρ_T = 0.869, indicating a good or very good reliability respectively and as such we went ahead with forming an arithmetic mean-index for these items of each scene.

Comparing the mean values (Avg.) and the standard deviation (σ) of both indices (Fig. 4), as well as of the remaining items of both question blocks we see that while the values of index we formed to represent the immersion and presence of the participants are almost equal (Avg. = 4.13; σ = 0.63 for the scene with cues; Avg. = 4.2; σ = 0.65 for the scene without cues), the participants passively felt more aware of the real world (Avg. = 2.94; σ = 1.11 for the scene with cues; Avg. = 3.44; σ = 1.1 for the scene without cues) and actively paid more attention to it (Avg. = 2.06; σ = 1.11 for the scene with cues; Avg. = 2.83; σ = 1.26 for the scene without cues) in the scene without cues. Objects or effects affecting the illusion

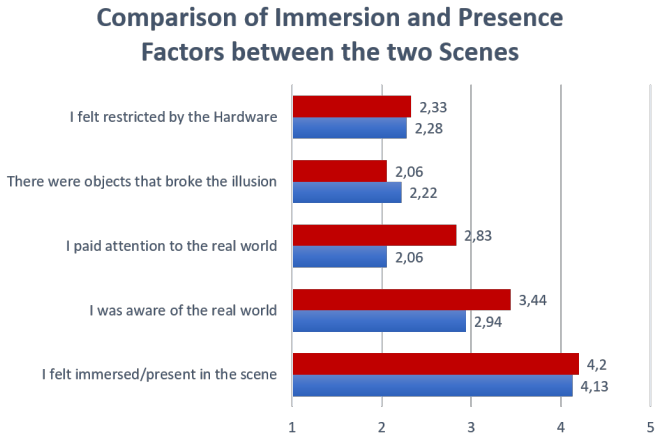

Fig. 4. The graph showes a direct comparison of the answers given between the scene with cues and without cues. All replies are in a scale of 1 = strongy disagree to 5 = strongly agree over a sample-set of N = 18. The bottom category represents an index created over multiple items regarding immersion/presence with Cronbach's Alpha values of ρ_T = 0.799 and ρ_T = 0.869 respectively.

were slightly more perceived in the scene with cues (Avg. = 2.22; σ = .31) than in the scene without cues (Avg. = 2.06; σ = 0.30). Hardware restrictions were perceived about equal in both scenes with average values of 2.28 (σ = 0.24) for the scene with cues and 2.33 (σ = 0.18) for the scene without cues.

Furthermore, the participants were asked to rate each scene right after experiencing it on a scale of 1 to 5 (1 = worst; 5 = best) as well as pick one scene they prefered over the other in the last section of the survey sheet. We looked at this rating and the sum of the preferences and set it into relation with the experience each participant stated to have with VR-technology so far.

While both scenes got favorable ratings in general, the scene without cues (Avg. = 4.5; σ = 0.17) came out slightly ahead of the scene with cues (Avg. = 4.44; σ = 0.19). However, ultimately 61.1% of participants preferred the experience with cues over the experience without them (38.9%). If we now consider the difference in experience with VR technology, we realise that those who stated to have no or only little experience preferred the version with cues. Once we get to participants who have at least casual experience with the technology the versions draw even, before the no-cues version is more liked with those who frequently use VR (Fig. 5).

Looking at the completion time, which was captured using a timer within the code of the project, we at first see a similar picture. The cue-less version was completed slightly faster (485.14 seconds on average) than the one with cues (492.74 seconds on average). Once we set it into relation to their experience however, we surprisingly notice that less experienced participants seem to have completed the version without cues up to one minute faster, while cues seem to have an impact of over 90 seconds on the most experienced participants (Fig. 6). The pivot again is placed with those who

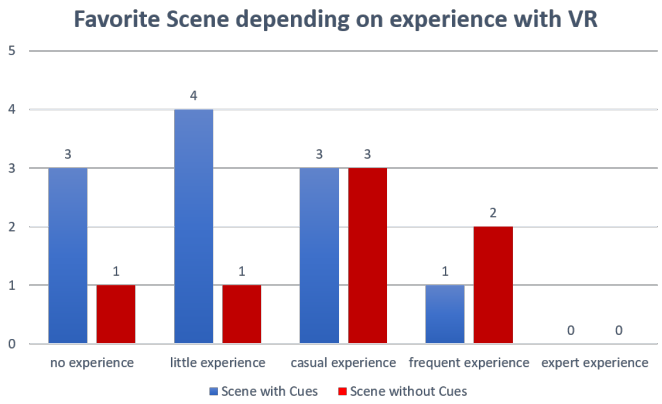

Fig. 5. The figure shows how many participants stated to prefer each of the versions, separated by their stated experience level with VR.

stated to have casual experience with VR.

C. Discussion

Regarding the initially posed questions whether diegetic cues enhance the user experience, there are two aspects. For rather inexperienced users of VR, the results show a clear preference towards the scene with diegetic cues, which was additionally confirmed by comments of the participants themselves. When it comes to more experienced users of VR however, they seem to prefer the version without cues, yet only two participants gave the cues as an explicit reason for their choice. This result can be explained by the Expertise Reversal Effect [33], which describes that the effectiveness of techniques depends on the knowledge of a person. In teaching, techniques which are effective for inexperienced learners can be less effective and even be negative for more experienced learners.

We can also see that the presence of diegetic cues in the way they were implemented in this project - as pre-attentive cues - did only have little impact on the presence. So, we have a strong indication that diegetic cues can be applied successfully to an interactive virtual reality experience in a similar way

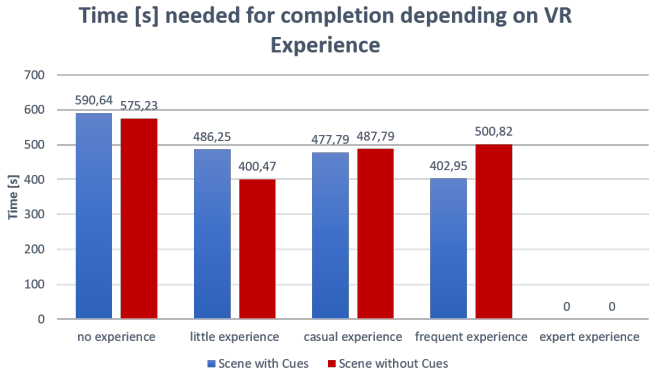

Fig. 6. The figure shows how long participants needed to complete each version in seconds, separated by their stated experience level with VR.

they were applied to omnidirectional videos in the work of [30] with similar guidance effects. It was also interesting, that the hue shift of the books seemed to have more success in guiding the players than the audio cue on the crystals. This goes contrary to the findings of [30] where the implemented audio-cue appeared to be the most successful. However, as we used pre-attentive cues this result might be based in the fact that not every participant actively recognized their presence.

The approach to storytelling used, drawing from influences of TTRPGs as suggested by [7], was met with positive responses as well. Some of the participants mentioned that they wished for even more interactions and agency over the conclusion of the scene. While an argument could be made that such an approach is more suited for games than for projects centering on telling one specific story (e.g. a movie) it should also be remembered that there are many games that are very much driven by the story they tell rather than their gameplay. Yet, as we did not directly compare the approach used in this work to a more traditional cinematic approach, we can only take this result as an indication that research in this direction might be a sensible step in the future.

D. Limitations and Future Work

The presented results came from a study that had to be performed during the peak of the COVID-19 pandemic, which made finding a working study-design as well as a venue and participants rather difficult. An alternative solution was found and a small-scale user study could take place. This alternative approach, while providing at least some data, was limited by time constraints per participant (around 40 minutes study and a buffer to account for disinfection and ventilation of the venue) as many participants had to be led through the experience on each of the days the study took place on. Distributing the study onto even more day was also not possible, as finding a venue that was large enough to set up room scale VR of a sufficient size and uphold all hygiene requirements was difficult. Considering this, the data gained can be used as an indication whether the approach used in the project is promising and cannot provide a definitive, empirical answer to the questions posed initially.

To find a clearer answer, the experiment should be repeated over a longer period once the restrictions due to COVID-19 have been eased up once again. Aside from repeating the experiment, this approach should be extended to a virtual reality experience with more than one scene. Additionally, other types of diegetic cues should be investigated, such as moving lights or objects, as well as looking at the effect of non-diegetic cues in such a setting and whether they influence the immersion of the users more or less than the diegetic cues used here. The presented approach can be extended with neural networks for creating an AI that can adapt and change the path of a preset story depending on the decisions and reactions of the player within each scene of it more dynamically.

IV. Conclusion

In this paper we proposed to move away from using traditional narrative theory and adapt a different style of storytelling with the aim to support interactions of viewers with the virtual environment while still being able to tell a story. Essentially moving them from the role of an "impartial bystander" to one of "active participant". To be able to retain some control over the narrative as the author, we furthermore proposed the use of diegetic cues as a method to guide the viewers attention and gaze towards important objects and locations.

Even though they are to be treated as indicators, the results of our study show that diegetic cues in an interactive, narrative VR experience have the potential to provide a tool for storytellers to guide the viewer or player towards the intended narrative without restricting the freedom VR provides. The combination of this method with a non-traditional approach to storytelling has the potential to create new, immersive, and interactive ways to experience works of fiction or historic events, among other experiences. We believe that this potential could not only be used for entertainment purposes, but also serve as a new way to provide accurate, interesting, and immersive education about a wide variety of topics. Now, and even more so with the growing possibilities research and technology will provide for this field in the future.

References

[1] J. S. Pillai, A. Ismail, and H. P. Charles, "Grammar of vr storytelling," in *Proceedings of the Virtual Reality International Conference - Laval Virtual 2017*, ser. ICPS: ACM international conference proceeding series, Association for Computing Machinery, Ed. New York, NY, USA: ACM, 2017, pp. 1–4.

[2] M. C. Reyes and G. Dettori, "Combining interactive fiction with cinematic virtual reality," in *Proceedings of the 9th International Conference on Digital and Interactive Arts*, Arantes, Ed. New York, NY, USA: ACM / Association for Computing Machinery, 2019, pp. 1–8.

[3] J. Mateer, "Directing for cinematic virtual reality: how the traditional film director's craft applies to immersive environments and notions of presence," *Journal of Media Practice*, vol. 18, no. 1, pp. 14–25, 2017.

[4] "What is ar, vr, mr, xr, 360?" 2020. [Online]. Available: https://unity3d.com/what-is-xr-glossary

[5] T. Sylvester, *Designing games: A guide to engineering experiences*. Sebastopol, CA: O'Reilly Media, 2013.

[6] D. Bordwell, *Narration in the fiction film*. Madison: Univ. of Wisconsin Press, 1985.

[7] R. Aylett and S. Louchart, "Towards a narrative theory of virtual reality," *Virtual Reality*, vol. 7, no. 1, pp. 2–9, 2003.

[8] M. Mearls, J. Crawford, C. Perkins, J. Wyatt, R. J. Schwalb, R. Thompson, and P. Lee, *Dungeon master's guide*, 5th ed., ser. Dungeons & dragons. Renton, WA: Wizards of the Coast, 2014.

[9] D. A. Bowman, E. Kruijff, J. J. LaViola, and I. Poupyrev, "An introduction to 3-d user interface design," *Presence: Teleoperators and Virtual Environments*, vol. 10, no. 1, pp. 96–108, 2001.

[10] M. Hassib, M. Pfeiffer, S. Schneegass, M. Rohs, and F. Alt, "Emotion actuator: Embodied emotional feedback through electroencephalography and electrical muscle stimulation," in *Proceedings of the 2017 chi conference on human factors in computing systems*, 2017, pp. 6133–6146.

[11] Y. Y. Qian and R. J. Teather, "The eyes don't have it: an empirical comparison of head-based and eye-based selection in virtual reality," in *Proceedings of the 5th Symposium on Spatial User Interaction*, 2017, pp. 91–98.

[12] J. P. Hansen, V. Rajanna, I. S. MacKenzie, and P. Bækgaard, "A fitts' law study of click and dwell interaction by gaze, head and mouse with a head-mounted display," in *Proceedings of the Workshop on Communication by Gaze Interaction*, 2018, pp. 1–5.

[13] D. Arijon, *Grammar of the film language*. Silman-James Press, 1991.

[14] S. Coren, L. M. Ward, and J. T. Enns, *Sensation and perception*, 5th ed. Fort Worth: Harcourt Brace, 1999.

[15] E. B. Goldstein, *Sensation and perception*, 8th ed. Belmont, Calif.: Wadsworth Cengage Learning, 2010. [Online]. Available: http://www.loc.gov/catdir/enhancements/fy1303/2008940684-b.html

[16] E. E. Veas, E. Mendez, S. K. Feiner, and D. Schmalstieg, "Directing attention and influencing memory with visual saliency modulation," in *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, D. Tan, G. Fitzpatrick, C. Gutwin, B. Begole, and W. A. Kellogg, Eds. New York, New York, USA: ACM Press, 2011, p. 1471.

[17] H. Syrett, L. Calvi, and M. van Gisbergen, "The oculus rift film experience: A case study on understanding films in a head mounted display," in *Intelligent Technologies for Interactive Entertainment*, ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, R. Poppe, J.-J. Meyer, R. Veltkamp, and M. Dastani, Eds. Cham: Springer International Publishing, 2017, vol. 178, pp. 197–208.

[18] A. Sheikh, A. Brown, Z. Watson, and M. Evans, "Directing attention in 360-degree video," *IET Conference Proceedings*, 2016.

[19] M. I. Posner, "Orienting of attention," *The Quarterly journal of experimental psychology*, vol. 32, no. 1, pp. 3–25, 1980.

[20] L. Ward, "Attention," *Scholarpedia*, vol. 3, no. 10, p. 1538, 2008.

[21] A. L. Yarbus, *Eye Movements and Vision*. Boston, MA: Springer US, 1967.

[22] D. Rumiński, "An experimental study of spatial sound usefulness in searching and navigating through ar environments," *Virtual Reality*, vol. 19, no. 3-4, pp. 223–233, 2015.

[23] R. Gunther, R. Kazman, and C. MacGregor, "Using 3d sound as a navigational aid in virtual environments," *Behaviour & Information Technology*, vol. 23, no. 6, pp. 435–446, 2004.

[24] E. van der Burg, C. N. L. Olivers, A. W. Bronkhorst, and J. Theeuwes, "Pip and pop: nonspatial auditory signals improve spatial visual search," *Journal of experimental psychology. Human perception and performance*, vol. 34, no. 5, pp. 1053–1065, 2008.

[25] E. R. Hoeg, L. J. Gerry, L. Thomsen, N. C. Nilsson, and S. Serafin, "Binaural sound reduces reaction time in a virtual reality search task," in *2017 IEEE 3rd VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*, S. Serafin, Ed. Piscataway, NJ: IEEE, 2017, pp. 1–4.

[26] S. Poeschl, K. Wall, and N. Doering, "Integration of spatial sound in immersive virtual environments an experimental study on effects of spatial sound on presence," in *2013 IEEE Virtual Reality (VR)*. IEEE, 18.03.2013 - 20.03.2013, pp. 129–130.

[27] P. Larsson, D. Västfjäll, P. Olsson, and M. Kleiner, "When what you see is what you hear: Auditory-visual integration and presence in virtual environments," *CyberPsychology Behavior*, 2005.

[28] S. Serafin and G. Serafin, "Sound design to enhance presence in photorealistic virtual reality," in *ICAD*, 2004.

[29] A. Brown, A. Sheikh, M. Evans, and Z. Watson, "Directing attention in 360-degree video," in *IBC 2016 Conference*. Institution of Engineering and Technology, 8-12 Sept. 2016, pp. 29 (9 .)–29 (9 .).

[30] S. Rothe and H. Hußmann, "Guiding the viewer in cinematic virtual reality by diegetic cues," in *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Berlin, Heidelberg: Springer, 2018, pp. 101–117.

[31] igroup - project consortium, "igroup presence questionnaire (ipq)," 2020. [Online]. Available: http://www.igroup.org/pq/ipq/index.php

[32] C. Jennett, A. L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, and A. Walton, "Measuring and defining the experience of immersion in games," *International Journal of Human-Computer Studies*, vol. 66, no. 9, pp. 641–661, 2008.

[33] J. Sweller, P. Ayres, and S. Kalyuga, "The expertise reversal effect," in *Cognitive Load Theory*, J. Sweller, P. Ayres, and S. Kalyuga, Eds. New York, NY: Springer New York, 2011, vol. 25, pp. 155–170.