Gaze-Based Interaction for Interactive Storytelling in VR

Heiko Drewes(✉) , Evelyn Müller, Sylvia Rothe , and Heinrich Hussmann

LMU Munich, Munich, Germany
{heiko.drewes,sylvia.rothe,hussmann}@ifi.lmu.de,
Mueller.Evelyn@gmx.de

Abstract. Interactive stories in Virtual Reality need a way of making decisions that influence the further progress of the story. The decision making should be easy and should not disturb the feeling of presence in the virtual world. As many virtual reality glasses come with an integrated eye tracker, it suggests itself to use gaze for making the decisions. We created an interactive story and implemented three different gaze interaction methods, which we evaluated in a user study with 24 participants. The three interaction methods used a dwell-time mechanism, one with a gaze button, one with a texture change of the object looked at, and one method without any feedback. The texture change was the favorite of the users, however, the choice of the interaction method may also depend on the intended dramaturgy and aesthetic aspects of the story.

Keywords: Interactive story telling · Gaze interaction

1 Introduction

Virtual Reality (VR) is a highly immersive technology which aims on giving the users a strong feeling of presence in a virtual world. Interactive storytelling is a method of involving the listener into the story by giving her or him the possibility to make decisions which influence the ongoing of the story. Both, VR and interactive storytelling, have the goal to create a strong, preferably realistic, illusion of being part of a fictional world. Telling an interactive story in VR seems to be a perfect match.

VR and interactive storytelling need ways to interact with the fictional world, which do not destroy the illusion. Among all possible forms of interaction such as with a controller device, voice, or body movements, gaze seems to be a good candidate for this demand. Additionally, many of the newer head-mounted devices (HMDs) come along with an integrated eye tracker.

Therefore we explored how to utilize gaze for decisions in interactive stories told in a VR environment. Within a user study we tested three different interaction methods. All three methods used gaze, either in a more subtle way by interpreting at what participants looked, or actively by requiring them to look at

an interaction object for a certain time to trigger a decision. The first method, the *button method*, used labeled buttons, while the other methods used an object from the story world to look at. With the second method the interactive object changes its texture while the user is looking at it and consequently we called this method the *texture method*. The third method, the *invisible method*, interprets the user's gaze, without giving feedback by changing the texture.

Although our research aims for an application in cinematic VR (CVR), a story told in pre-rendered or prerecorded parts, we created an interactive story as animation rendered in real-time and implemented three gaze-based interaction methods. However, to keep the constraints of cinematic VR, the users' position was fixed in one location. We invited 24 participants to our study, where each participant tried all interaction methods and filled questionnaires after the trials.

The user study results show that the button method is easy to understand, but weakens the illusion. The invisible method does not affect the illusion, but is irritating because of loss of control. The texture method seems to be a good compromise between keeping the illusion and keeping control.

2   Related Work

Interactive stories existed already in the analog times, for example as role-playing games[1], there the story progress depends on a process of structured decision-making. In the digital world computer adventure games can be seen as an interactive story. Such adventure games typically have a very high degree of interactivity, which often depends on reaction time, and maybe more a game of skill than an interactive story.

Since the 1960s, filmmakers and researchers have been investigating how films can interact with cinema screens and different types of displays [15]. Interactive stories on the web uses game elements such as solving tasks, collecting points, and giving feedback [1]. The interaction is closely related to engagement: the audience changes their lean-back attitude to a lean-forward attitude. Vosmeer and Schouten introduced the term lean-in medium for CVR [23]: In lean-in media, the direction of view, and thus the seen part of the image, is freely chosen. As a result the user is more active than in classic lean-back media, but remains a spectator.

2.1   Story Structures

Different story structures can be used to create interactive stories [19,20]. The simplest of them, the linear structure, can also be implemented without any interaction and corresponds to the traditional film. However, a linear structure can also be implemented interactively. In contrast to a conventional film, the viewer's interaction determines the speed of the story progress. By expanding the interaction options to include additional interactive regions, non-linear storylines are possible. If the viewer looks at a certain region, the film continues with a connected scene.

---

[1] https://www.goodreads.com/genres/role-playing-games.

2.2   Interaction

The term interaction is used in different fields with different meanings. This paper looks at interactions between people and a digital story that is implemented for using a HMD. The selection of the image section by looking around is a very natural type of interaction: the user's head movements are analyzed by the sensors of the HMD and the corresponding image section is displayed, just as the user is used to from the real world.

Interaction techniques in VR are divided into navigation (change of location), selection (selection of objects), manipulation (change of objects) [2]. In CVR the user cannot walk, move or modulate objects. The main interactions in CVR are looking around and selecting areas. Most CVR films are currently linear, even if the additional spatial component is predestined for non-linear story structures. Interactive stories depending on the direction of view can be a natural interaction without restricting the freedom of the viewer. This requires selection techniques with which the viewer can intentionally or unintentionally select regions [16,17].

2.3   Selection Techniques in VR

Various studies have been carried out in recent years on the selection of objects in VR environments. Head and eye movements, gestures and sensor data can be used for selection processes. Most of these methods have been developed for VR environments with six degrees of freedom (DoF) and require careful examination of whether they are also suitable for CVR where only three DoFs exist. Some of these techniques focus on the accuracy of pointing or on performing tasks. Such criteria are often less important for CVR, since the focus is a pleasant film experience.

Nukarinen et al. [11] compared raycasting (using a beam emanating from the controller as a pointer) with two gaze-based selection techniques. In their experiments, the ray cast method followed the controller direction and pressing the controller button triggered the event. The two gaze-based methods differ in the release technique, one method being triggered by the controller button and the other by dwell time. In their study, the beam performed better than the gaze methods and the gaze with additional button better than the dwell time.

Kallioniemi et al. [9] examined different types of selection for hotspots: dwell time and immediate fade in. The immediate fade-in of the next scene started automatically when the head pointed in the direction of a hotspot. In their experiments, the fade-in with large symbols performed best. Qian and Teather [12] compared gaze- and head-based techniques. The head-based selection was faster, more accurate, and offered a better user experience.

2.4   Eye-Based Interaction

There are different ways to use the eyes in human-computer communication. Jacob [7,8] examined gaze-based interaction techniques using dwell-time. The user has to look at an interaction object for a certain time, the dwell-time.

With a short dwell-time this interaction method carries the risk of accidentally triggering events, e.g. when a viewer inspects an area for a long time. Jacob called this effect, the unwanted activation of commands, the Midas Touch effect [7,8].

Further gaze interaction methods like gaze gestures [4] or smooth pursuit eye movements [22] seem not suitable for gaze interaction in interactive storytelling. Both methods do not need calibration, but this advantage is not important for a personal HMD. Gaze gestures are far away from being intuitive, need instructions, and are much less natural than looking at something. Smooth pursuits are a natural eye movement, but need moving interaction objects. Moving objects are not a problem in a movie, however, reliable smooth pursuit detection requires a minimum speed of $5°$/s [3], which is relatively high. Our experiments with smooth pursuit detection in 2D videos were discouraging, mostly because there are not many moving targets in a 2D video as typically the camera follows moving objects. Although this might be different for CVR content we decided to stick on the dwell-time method for our study.

3   Prototype

3.1   Story

We created an interactive story to test different interaction methods for the decisions within the story. We implemented the story in Unity[2] with C# scripts, Blender, Photoshop, and Audacity[3]. As this research aims on cinematic virtual reality, the viewer of the story stays in a fixed location. Within a 3D animation solution it is easy to identify at which object the user looks. With a real video it would be necessary to provide a mapping of the interactive objects to regions over time.

The story plays on a spaceship where an one-eyed alien wants to do a coffee break and gets interrupted by an alarm caused by the cat (see Fig. 1). In the story progress the alien has to make decisions on how to deal with the problem.

Everything happens in the alien's relaxation room (see Fig. 2), where the character has access to two vending machines (coffee and strawberry milkshakes), various monitors, a maintenance hatch, an alarm, some objects for the cat and a cozy corner. Figure 2 also shows the positions of buttons, which were used for the button interaction method.

The story consists of six sections with a decision at the end of the first five sections and three possible ends - the alien finds the cat as the source of the problem (end 1), the alien calls the break-down service (end 2), the alien heads for a space ship garage (end 3). There are seven different paths through the story which lead to seven different comments (a–g) from the alien at the end (see Fig. 3).
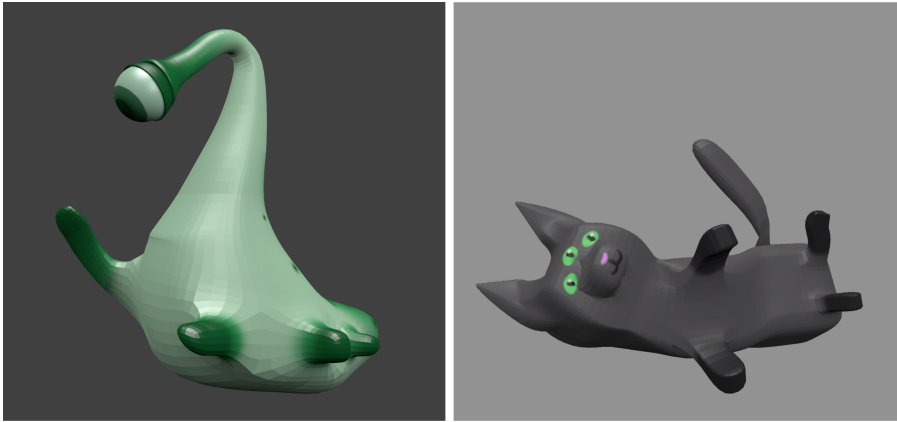

Fig. 1. An alien (left) is the main character of the interactive story. The cat (right) caused a malfunction of the spaceship.
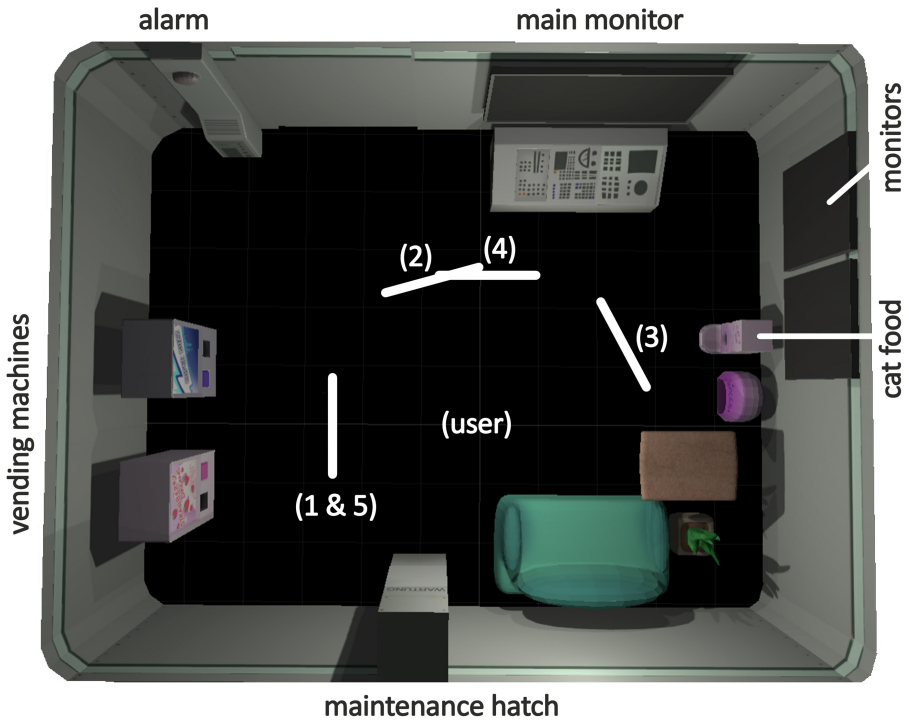
Fig. 2. Positions of objects, user and (if chosen) buttons for all decisions (1–5)

The story has a branching structure with optional scenes. All links occur between narrative nodes and are therefore external according to Reyes [13]. There is no exploratorium in the sense of Vesterby et al. [21].
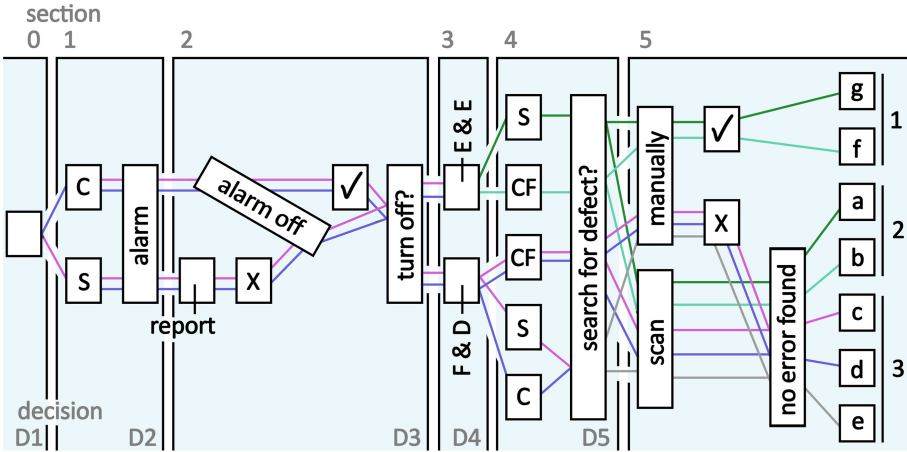
Fig. 3. Structure of the interactive story. S: Strawberry, C: Coffee, CF: Cat food, F & D: Food & Drinks, E & E: Engines and Energy

3.2 Interaction Methods

We decided to use three different gaze-based interaction methods. The first interaction method is a typical gaze button using the dwell time method and therefore called *button method*. For feedback the button fills with color while looking at it. While blinking with the eyes the filling stops for at most 150 ms, and after that, or when looking somewhere else, the button drains. A complete fill without interruption needs two seconds, which is quite long for a dwell time. A long dwell time helps to avoid the Midas Touch effect [7]. Ia a CVR context, and in contrast to eye typing, the users have to interact only few times and a long dwell time seems to be justified. Additionally, we placed a button label as a caption above the button. The appearance of the gaze button is depicted in Fig. 4.

The second interaction method works similar to the gaze button, however instead of looking at a button the users have to look at an object. We realized the feedback by a changing texture of the interactive object and therefore call the method *texture method*. In our prototype we changed the colors of the objects to blue or red (see Fig. 5). In contrast to the button method there is no labeling of the possibilities, and consequently it is not obvious to the users what the interaction objects are. A user has to look at an object and wait for a texture change to identify it as interactive.

The third interaction method is very subtle, as it does not offer any feedback, besides that the story progresses differently. Therefore we call it the *invisible method*. During a decision period this method measures for how long the user looks at each decision object. If they look at an object multiple times, all duration measurements are added up. The decision period ends, if either the gaze has been on relevant objects for at least three seconds, or if the elapsed time exceeded the maximum duration of five seconds. After that, the object with a higher gaze

Fig. 4. Button method: The gaze button fills from the bottom while looking at it, in this case the left button. When the button is filled completely the decision is done.

duration during the decision period gets selected. If the user did not look at any interaction object, the invisible method selects the object at which the user looked the most since the story started.

3.3 Hard- and Software

The hardware for the prototype consisted of a laptop (Acer Aspire A715-74G), headphones (VIVANCO 34877 COL 400) and a VR head-mounted display (FOVE[4]).

4 User Study

4.1 User Study Design

We decided to use a within-group design for the study so that every participant tested all of the three interaction methods and we could ask everybody about her or his opinion on every method. The participants watched the story three times and therefore had the opportunity to try different paths through the story.

We designed a questionnaire with questions on demographic data and possible debility of sight, experiences with virtual reality similar to the IFcVR questionnaire of Reyes et al. [14], questions on the interaction method, and a selection of questions from commonly accepted questionnaires. These questionnaires were the igroup presence questionnaire IPQ[5] with one question each for spatial presence, involvement, and experienced realism, the user experience

---

[4] https://fove-inc.com/product/.

[5] http://igroup.org/pq/ipq/download.php.
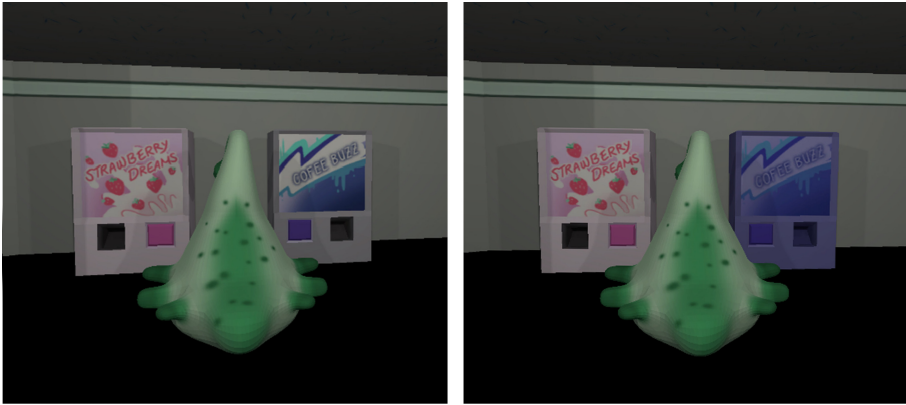
Fig. 5. Texture method: While looking at an object, the object's texture changes. Here it is the coffee vending machine in the right picture.

questionnaire UEQ-S[6], subjective task workload NASA-TLX [5,6], and simulator sickness questionnaire [10]. Additionally, inspired by Reyes et al. [14], we asked about autonomy, effectance, and believability. This part of the questionnaire had to be filled after each experienced interaction method and this was the reason why we used only a shortened version of the standard questionnaires.

4.2   Conducting the User Study

On arrival every participant got a printed information on the study and were told that their gaze will have influence on the ongoing of the story. Then the FOVE was calibrated on the participant and the calibration success was checked with a sample scene supplied by FOVE. After this the participants filled a consent form and the demographic part of the questionnaire.

In the next step the participants watched the story three times, each time with a different interaction method and a new calibration. We changed the six possible orders of interaction methods from participant to participant to eliminate possible learning effects. After each round the participants filled the questionnaire. Finally we asked the participants which interaction method they liked best.

5   Results

5.1   Demographic Data

We recruited 24 participants, 18 (75%) female and 6 (25%) male persons, in the age from 19 to 63 years (Mean: 28.6) for the user study. Eleven (45.8%) persons

---

[6] https://www.ueq-online.org/.

wore glasses or contact lenses. One third of the participants stated that they have no experiences with virtual reality, two persons were experienced in VR, and the rest stated that they have little experiences. Three quarters of the participants had no experience with eye trackers and the remaining quarter had only little experience.

5.2 Paths Through the Story

The average story time including the time for interaction was 4 min and 20 s. The story had three possible ends as explained in Sect. 3.1. Random decisions lead to end 3 with 50% and with 25% to end 1 or end 2. Figure 6 shows the distribution of reached ends for all three runs and Fig. 7 for the first run only.
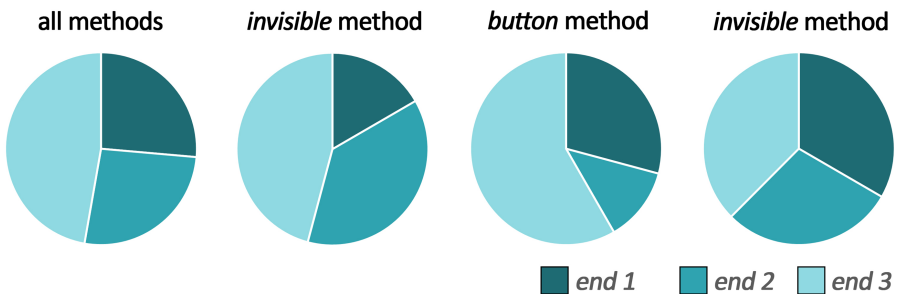

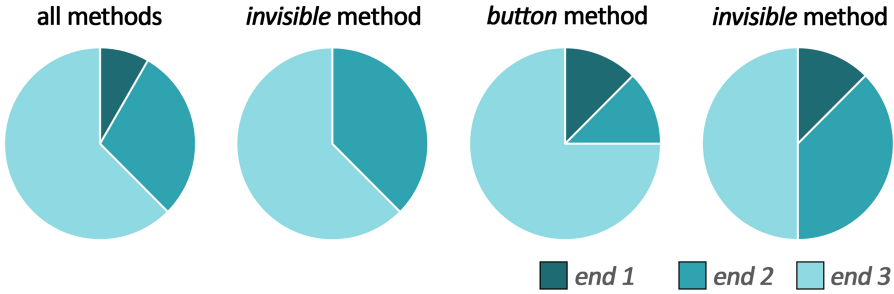**Fig. 6.** Reached story end over all participants for three runs.

Table 1 shows the mean decision times and standard deviations for each interaction method for all runs and Table 2 only for the first run. The mean decision times for the first run are longer than the mean decision times for all runs. This means there is a learning effect, however this might be by learning the story and not necessarily by learning the interaction method.

5.3 Questionnaire

The questionnaire used Likert scales for the answers. As this scale is non-parametric we used Friedman tests with a post-hoc Conover squared ranks test to calculate whether there are significant differences between the interaction methods (with a significance level $\alpha = 0.05$). The calculation where done with JASP[7].

Presence. Figure 8 shows the results for presence on a 7-point Likert scale. We could not find significant differences for presence. It seems that the interaction method does not affect the feeling of presence.

---

[7] https://jasp-stats.org/.

Fig. 7. Reached story end over all participants in the first run.

Table 1. Decision time mean and standard deviation in seconds for all runs depending on the method

| Method	Invisible	Buttons	Texture
Decision 1	3.79 (0.73)	6.83 (6.32)	5.82 (3.62)
Decision 2	4.32 (0.95)	8.23 (12.7)	6.48 (4.98)
Decision 3	3.90 (0.87)	4.87 (2.71)	4.95 (6.39)
Decision 4	4.47 (0.73)	9.68 (14.2)	4.56 (2.35)
Decision 5	4.59 (0.75)	12.4 (8.33)	7.44 (6.35)

Table 2. Decision time mean and standard deviation in seconds for the first run depending on the method

Method	Invisible	Buttons	Texture
Decision 1	3.87 (0.79)	8.89 (9.87)	7.02 (4.25)
Decision 2	4.82 (0.59)	6.65 (2.21)	11.5 (5.46)
Decision 3	4.16 (0.83)	4.96 (2.47)	3.37 (0.97)
Decision 4	4.60 (0.58)	6.96 (4.02)	3.89 (1.40)
Decision 5	5.01 (0.09)	16.5 (11.7)	9.14 (8.61)

User Experience. We evaluated the shortened version of the user experience questionnaire with the data analysis tools provided by the UEQ website[8]. Figure 9 shows the results on the 7-point Likert scales. In total the invisible method (Mean: 4.82—SD: 0.97) was significantly worse than the texture method (Mean: 5.69—SD: 0.81—$p_{Conover} < 0.001$) and the button method (Mean: 5.39—SD: 0.97—$p_{Conover} = 0.010$).

The UEQ's data analysis tools distinguish pragmatic and hedonic qualities, The first four questions in our shortened questionnaire were pragmatic and the last four questions were hedonic. All interaction methods got positive values

---

[8] https://www.ueq-online.org.
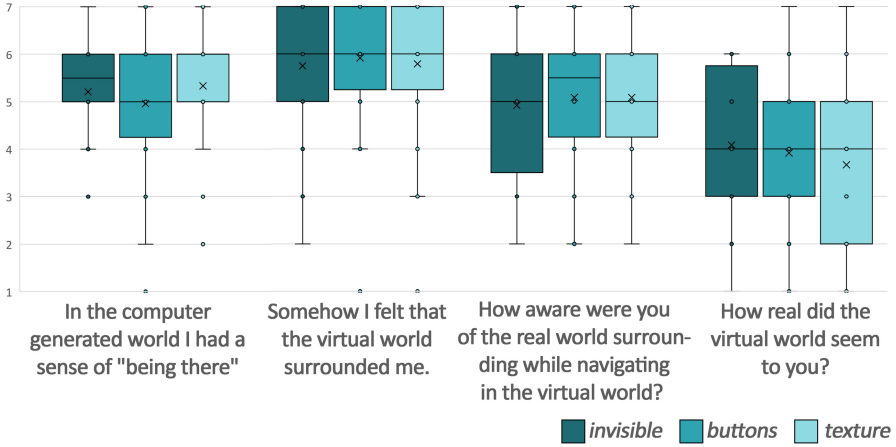
Fig. 8. Results of the presence questionnaire (1: negative, 7: positive).

for both qualities, except the pragmatic quality of the invisible method, which evaluated to neutral.

Workload. We used the NASA-TLX [5,6] to measure the participants' workload. Figure 10 shows the results. The evaluation with a Friedman and a Conover Post Hoc test on a 5% significance level did not reveal any significant difference for the three interaction methods, except for frustration. The users' frustration for the invisible method (Mean: 35.21—SD: 27.63) was significantly bigger for the texture method (Mean: 18.96—SD: 17.26—$p_{Conover} = 0.032$) and for the button method (Mean: 15.0—SD: 14.14—$p_{Conover} = 0.001$).

Further Questions. Further questions on a Likert scale from 1 to 7 aimed on how users experienced the study. Figure 11 summarizes the results. Again we used a Friedman and a Conover post hoc test for the analysis.

There were no significant differences for the three methods for the question whether the story was comprehensible, and also for the question whether the users would like to watch further stories using this interaction method.

For the question whether the participants felt they had control over the story we found significant worse judgment for the invisible method (Mean: 4.13—SD: 2.01) against the button method (Mean: 5.83—SD: 1.57—$p_{Conover} = 0.003$) and the texture method (Mean: 5.96—SD: 1.21—$p_{Conover} = 0.002$).

Also for the question on whether the system correctly registered the users' decisions the invisible method (Mean: 5.0—SD: 2.31) was significantly judged worse than the button method (Mean: 6.63—SD: 1.22—$p_{Conover} = 0.001$) and the texture method (Mean: 6.63—SD: 1.25—$p_{Conover} < 0.001$).
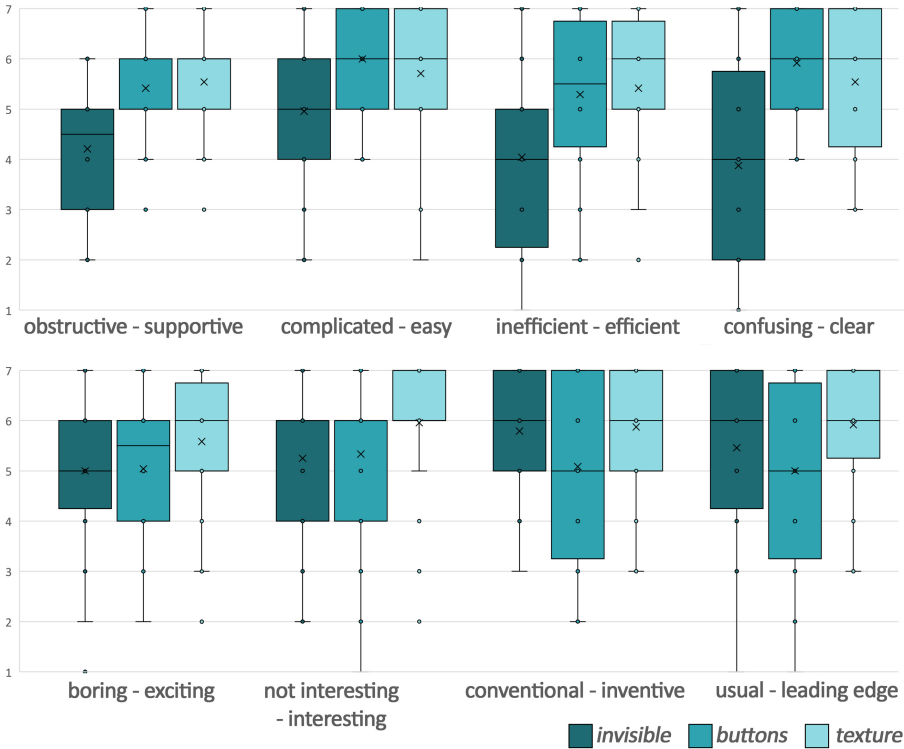
Fig. 9. Results of the user experience questionnaire (1: negative, 7: positive)

The question whether the users are always aware of all possible choices showed a significance for the button method (Mean: 6.46—SD: 1.29—$p_{Conover}$ = 0.002) against the invisible method (Mean: 4.88—SD: 2.13) but not for the texture method (Mean: 6.13—SD: 1.30).

Finally the users perceived the texture method (Mean: 6.42—SD: 0.76—$p_{Conover}$ < 0.001) and the button method (Mean: 6.41—SD: 0.81—$p_{Conover}$ < 0.001) as more intuitive than the invisible method (Mean: 4.29—SD: 1.95).

We evaluated the results for the first run with a Kruskal-Wallis test with Dunn post hoc test as this test works for unpaired data. Already for the data from the first run only we found significant differences for the question on whether the system correctly registered the users' decisions, where the invisible method (Mean: 4.88—SD: 2.37) was rated lower than the button method (Mean: 6.88—SD: 0.33—p_{Dunn} = 0.009) and the texture method (Mean: 7.00—SD: 0—p_{Dunn} = 0.007) and for the question how intuitive are the methods, where the texture method (Mean: 6.75—SD: 0.43—p_{Dunn} = 0.021) was judged significantly more intuitive than the invisible method (A: 4.875—SD: 1.83).
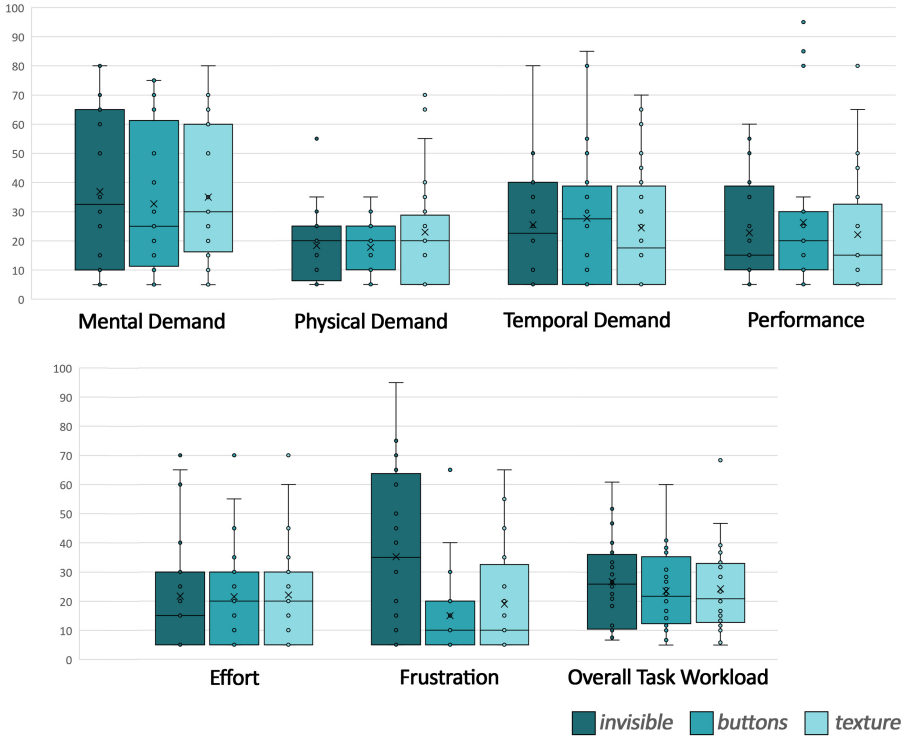
Fig. 10. Results of the NASA-TLX questionnaire for the three interaction methods.

User Opinion. In the last questionnaire after the users tried all interaction methods we asked the users which method they liked best, which method was least interrupting, and which method was the easiest interaction. Figure 12 shows the results.

6 Discussion and Limitations

We found several differences between the methods. The button method is easiest to understand for the users. It reveals all possible choices for a decision, has explanatory texts, and it is clear that the story waits for a decision. The texture method has no explanatory texts and the story has to communicate that it expects a decision. The possible options have to be discovered by the user by searching for objects which give feedback. The invisible method is confusing for the user.

The result that the form of interaction does not significantly influence the feeling of presence was surprising in the first moment. The reason is that there is a difference between the feeling of being present in a virtual world and the illusion of being in a real world. The appearance of a button levitating in space
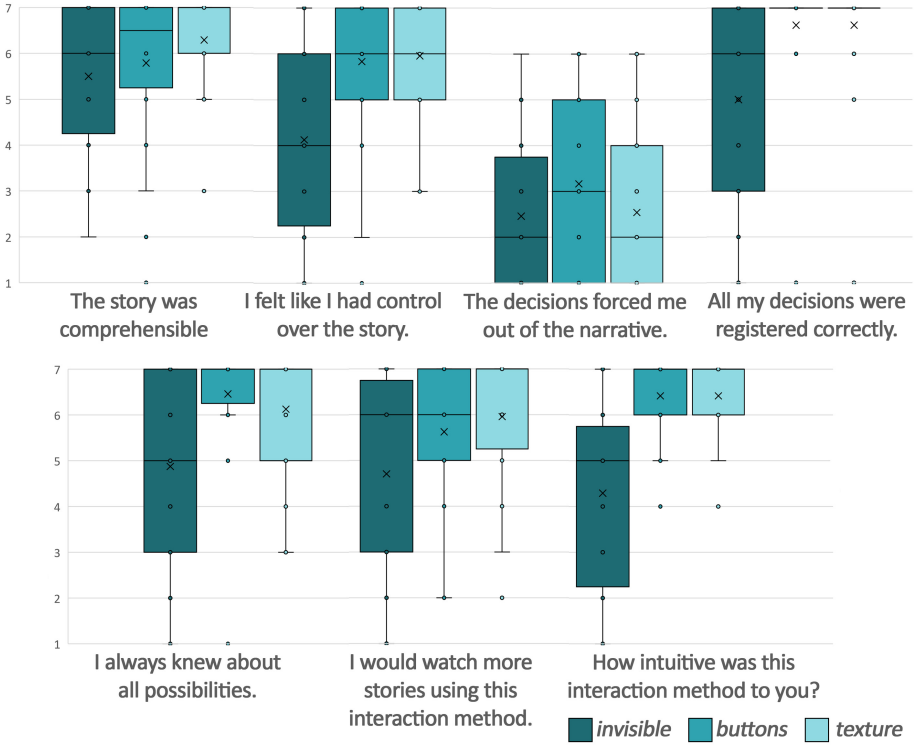
Fig. 11. Results of further questions on how users experienced the study.
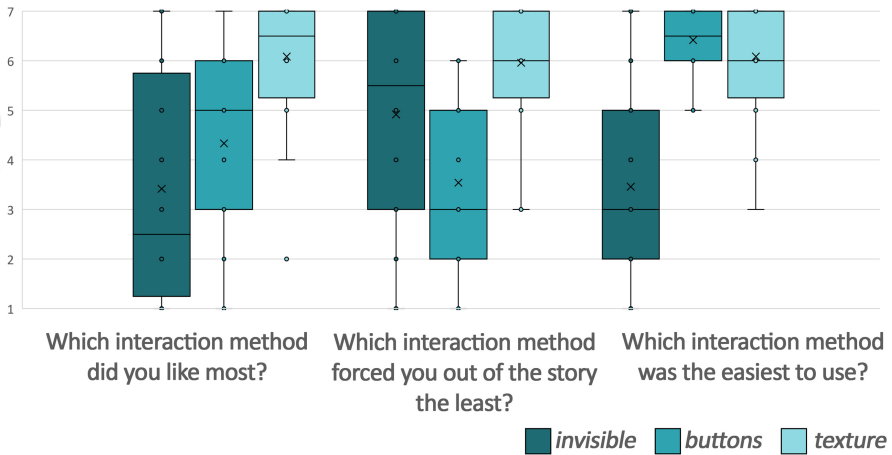

Fig. 12. User opinion on the three interaction methods.

contradicts experiences from the real world, but does not affect the perception of the three-dimensional space. The user still feels to be present in a virtual world, but not in a real world.

The button method interrupts the story and destroys the feeling of being in a real world. To mitigate the effect the button design should match the style of the story, what is very common in 3D computer games, or could even be a part of the storyworld (diegetic). In the user study's story, buttons could be on the spaceship's control console. However, depending on the story, such solution is not always possible, especially as it is not possible to move inside cinematic VR.

The invisible method does not disturb the story at all and consequently keeps the illusion, but for the price that the user may not even be aware of her or his decision. Within a user study, where the users have the feeling that they have to fulfill a task, the invisible method creates a feeling of failure. However, if the story maker sells the story as a sequence of riddles, the perception will be different. It will lead also to a more exploratory way of looking at the scenes.

The texture method seems to be a good compromise for keeping the illusion and at the same time enables control over decisions. However, a changing texture is something that does not happen in reality and therefore still flaws the illusion. A form of feedback which does not contradict real world experiences is preferable. For example an actress or actor could turn towards the observer when looking long enough at her or him, or an indicator light could start blinking when looking at the coffee vending machine. However, such solutions for feedback depends heavily on the story and are hard to generalize.

The texture method is only one representative for a more general class of interaction methods. In this class the interaction takes place by an diegetic object that belongs to the story, which gives feedback. There are many possible ways to provide this feedback. For example instead of changing the texture it is also possible to let the object glow or give it a halo. The answer to the question which feedback works best depends on the story and the context.

Our results are in accordance with the research of Vesterby et al. [21] for stories presented on a 2D screen. However, in the context of VR the loss of the illusion weights much more. It can be seen as a limitation of our study, that the story was a rendered animation and not a real video. A real video creates a more realistic illusion and the button method may destruct the illusion even more than in an animation.

7   Future Work

The interaction objects in our study did not move. It is unclear whether moving objects or subjects are suitable for interaction. On one hand following a character leaving the scene with the gaze is natural, signals interest in this character, and could be taken as a decision. On the other hand such a situation limits the time to make the decision. Alternatively, the movement of an interactive object could stay within the scene of a 360° movie but leave the field of view. Following the object by turning the head to keep the object in the view could be the required

action for decision making. It is also possible to have a moving interaction object which stays in the field of view and the viewer makes the decision by following the moving object. Questions on the speed range and the predictability of the movement trajectory and their influence on the user experience are not answered yet.

New HMDs have integrated trackers which report the exact position of all fingers. This allows to control an interactive story with hand gestures such as pointing with the index finger. Such gestures are as natural as gaze movements and open further possibilities which needs further research.

With volumetric video [18] a new media technology is already on the horizon. With this technology the borders between 3D computer games and interactive stories dissolve. In a volumetric video the watching person can move within the presented world, which is not possible in cinematic VR. This has consequences for interaction design. For the button method it means that the buttons have to be placed within the proximity of the spectator. For the texture and the invisible method the diegetic interaction object changes the (angular) size depending on the user's position. It may be necessary to be close to the object to interact with it and make a decision for the story that way. Additionally, it may make a difference from which direction, means from the front or the backside, a user looks at the interaction object.

8   Conclusion

The work analyzes the behavior of viewers and shows different aspects that should be taken into account in interactive stories. We compared three gaze-based methods, which do not require any additional devices for input. For the investigated scenario, the viewers preferred to be informed about the decision possibilities. They favored methods in which they can actively influence the decision-making process. If possible, the feedback should be embedded in the story, e.g. by changing the texture, which was used in this work. The opportunities for interaction that remain hidden from the user and are triggered due to their gaze behavior, without the viewer being actively involved in this decision, led to lower acceptance.

This work is a first step to find methods to facilitate the realization of interactive stories in VR that support the user in exploring the story in such a way that the story is not disturbed.

References

1. Argyriou, L., Economou, D., Bouki, V., Doumanis, I.: Engaging immersive video consumers: challenges regarding 360-degree gamified video applications. In: 2016 15th International Conference on Ubiquitous Computing and Communications and 2016 International Symposium on Cyberspace and Security (IUCC-CSS), pp. 145–152 (2016). https://doi.org/10.1109/IUCC-CSS.2016.028

2. Bowman, D.A., Kruijff, E., LaViola, J.J., Poupyrev, I.: An introduction to 3-d user interface design. Presence Teleoperators Virtual Environ. **10**, 96–108 (2001). https://doi.org/10.1162/105474601750182342. http://www.mitpressjournals.org/doi/10.1162/105474601750182342

3. Drewes, H., Khamis, M., Alt, F.: Smooth pursuit target speeds and trajectories. In: Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia, MUM 2018, pp. 139–146. ACM, New York (2018). https://doi.org/10.1145/3282894.3282913. http://doi.acm.org/10.1145/3282894.3282913

4. Drewes, H., Schmidt, A.: Interacting with the computer using gaze gestures. In: Baranauskas, C., Palanque, P., Abascal, J., Barbosa, S.D.J. (eds.) INTERACT 2007. LNCS, vol. 4663, pp. 475–488. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-74800-7_43   http://dl.acm.org/citation.cfm?id=1778331.1778385

5. Hart, S.G.: Nasa-Task Load Index (NASA-TLX); 20 years later. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 50, no. 9, pp. 904–908 (2006). https://doi.org/10.1177/154193120605000909

6. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) Advances in Psychology: Human Mental Workload, vol. 52, pp. 139–183. North-Holland (1988). https://doi.org/10.1016/S0166-4115(08)62386-9. http://www.sciencedirect.com/science/article/pii/S0166411508623869

7. Jacob, R.J.K.: What you look at is what you get: eye movement-based interaction techniques. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 1990, pp. 11–18. Association for Computing Machinery, New York (1990). https://doi.org/10.1145/97243.97246. https://doi.org/10.1145/97243.97246

8. Jacob, R.J.K.: The use of eye movements in human-computer interaction techniques: what you look at is what you get. ACM Trans. Inf. Syst. **9**, 152–169 (1991). https://doi.org/10.1145/123078.128728. http://portal.acm.org/citation.cfm?doid=123078.128728

9. Kallioniemi, P., et al.: Hotspot interaction in omnidirectional videos using head-mounted displays. In: Proceedings of the 22nd International Academic Mindtrek Conference, Mindtrek 2018, pp. 126–134. Association for Computing Machinery, New York (2018). https://doi.org/10.1145/3275116.3275148. https://doi.org/10.1145/3275116.3275148

10. Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G.: Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int. J. Aviat. Psychol. **3**(3), 203–220 (1993). https://doi.org/10.1207/s15327108ijap0303_3

11. Nukarinen, T., Kangas, J., Rantala, J., Koskinen, O., Raisamo, R.: Evaluating ray casting and two gaze-based pointing techniques for object selection in virtual reality. In: Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology, VRST 2018. Association for Computing Machinery, New York (2018). https://doi.org/10.1145/3281505.3283382. https://doi.org/10.1145/3281505.3283382

12. Qian, Y.Y., Teather, R.J.: The eyes don't have it: an empirical comparison of head-based and eye-based selection in virtual reality. In: Simeone, A.L. (ed.) Proceedings of the 5th Symposium on Spatial User Interaction, Brighton, UK, pp. 91–98. ACM (2017). https://doi.org/10.1145/3131277.3132182

13. Reyes, M.C.: Screenwriting framework for an interactive virtual reality film. In: Online proceedings of the 3rd Immersive Learning Research Network, Coimbra, Portugal (2017). https://doi.org/10.3217/978-3-85125-530-0-15

14. Reyes, M.C.: Measuring user experience on interactive fiction in cinematic virtual reality. In: Rouse, R., Koenitz, H., Haahr, M. (eds.) ICIDS 2018. LNCS, vol. 11318, pp. 295–307. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-04028-4_33

15. Roth, C.: Experiencing interactive storytelling. Ph.D. thesis, Vrije Universiteit Amsterdam, January 2015. http://dare.ubvu.vu.nl/handle/1871/53840

16. Rothe, S., Hussmann, H.: Spaceline: a concept for interaction in cinematic virtual reality. In: Cardona-Rivera, R.E., Sullivan, A., Young, R.M. (eds.) ICIDS 2019. LNCS, vol. 11869, pp. 115–119. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-33894-7_12

17. Rothe, S., Pothmann, P., Drewes, H., Hussmann, H.: Interaction techniques for cinematic virtual reality. In: Teather, R., Itoh, Y., Gabbard, J. (eds.) Proceedings, 26th IEEE Conference on Virtual Reality and 3D User Interfaces, Osaka, Japan, pp. 1733–1737. IEEE (2019). https://doi.org/10.1109/VR.2019.8798189

18. Schreer, O., et al.: Lessons learned during one year of commercial volumetric video production. SMPTE Motion Imaging J. **129**(9), 31–37 (2020). https://doi.org/10.5594/JMI.2020.3010399

19. Tanak, N.: Interactive Cinema. MediaLAB Amsterdam (2015)

20. Verdugo, R., Nussbaum, M., Corro, P., Nuñnez, P., Navarrete, P.: Interactive films and coconstruction. ACM Trans. Multimedia Comput. Commun. Appl. **7**, 1–24 (2011). https://doi.org/10.1145/2043612.2043617. http://dl.acm.org/citation.cfm?doid=2043612.2043617

21. Vesterby, T., Voss, J.C., Hansen, J.P., Glenstrup, A.J., Hansen, D.W., Rudolph, M.: Gaze-guided viewing of interactive movies. Digital Creativity **16**(4), 193–204 (2005). https://doi.org/10.1080/14626260500476523

22. Vidal, M., Bulling, A., Gellersen, H.: Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In: Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp 2013, pp. 439–448. ACM, New York (2013). https://doi.org/10.1145/2493432.2493477. http://doi.acm.org/10.1145/2493432.2493477

23. Vosmeer, M., Schouten, B.: Interactive cinema: engagement and interaction. In: Mitchell, A., Fernández-Vara, C., Thue, D. (eds.) ICIDS 2014. LNCS, vol. 8832, pp. 140–147. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-12337-0_14