Exploring the Effects of Diegetic and Non-diegetic Audiovisual Cues on Decision-making in Virtual Reality

Anıl Çamcı
University of Michigan
Department of Performing Arts Technology
acamci@umich.edu

ABSTRACT

The user experience of a virtual reality intrinsically depends upon how the underlying system relays information to the user. Auditory and visual cues that make up the user interface of a VR help users make decisions on how to proceed in a virtual scenario. These interfaces can be diegetic (i.e. presented as part of the VR) or non-diegetic (i.e. presented as an external layer superimposed onto the VR). In this paper, we explore how auditory and visual cues of diegetic and non-diegetic origins affect a user's decision-making process in VR. We present the results of a pilot study, where users are placed into virtual situations and are expected to make choices upon conflicting suggestions as to how to complete a given task. We analyze the quantitative data pertaining to user preferences for modality and diegetic-quality. We also discuss the narrative effects of the cue types based on a follow-up survey conducted with the users.

1. INTRODUCTION

Virtual realities are information-rich environments where multiple channels of communication can be formed between the system and the user. Narrative cues in the audio and visual domains are used to guide the users through their experiences in VR. These cues can be presented in the form of user interface elements superimposed onto the VR. They can also be built into the virtual environment itself as objects situated in the implied universe of the VR.

Similar distinctions between the elements of storytelling are used in film, theatre and games in service of diverse narrative goals. For instance, the conversation between the characters in a film can be contradicted by a narrator to warn the audience of a possible deception in the story. Such narrative devices have been utilized in video games with nonlinear gameplay, where users can make decisions upon conflicting situations that lead to branching storylines.

The form in which audio and visual cues are presented to the user in VR can be used for similar effects. In this paper, we explore these effects in the context of a VR game in which the users are expected to make decisions to exit a virtual room in light of conflicting audio and visual cues that are presented diegetically (within the room) and non-diegetically (as an external layer). We discuss the results of a study where the users were presented with various combinations of such cues. The use of concurrent conflicting cues allows us to investigate within-subject preferences based on modality and diegetic quality. Moreover, we explore the narrative functions that users impart to contradicting cues in terms of their origin and trustworthiness.

In our analysis, we look at decision types and timings, as well as modality and diegetic-quality pairings. Furthermore, we evaluate the qualitative responses gathered from a follow-up survey to highlight the ways in which the users interpreted the various cue types in their decisions and how these interpretations affected their narrative experience.

2. RELATED WORK

Modern virtual reality systems are constrained by hardware limitations to a much lesser extent than they were a decade ago. An increasing number of researchers are therefore able to focus on experiential qualities of VR. Accordingly, several design guidelines for VR have been proposed in the recent years [1,2]. We are arguably in the early days of formulating a VR theory akin to that of more established art forms; numerous researchers and practitioners work towards a deeper understanding of how we perceive modern virtual realities, and how we behave in them.

For instance, Naz et al. explore the links between affective qualities of a virtual space and its visual design parameters [3]. The researchers find that the hue and brightness of the colors used in a virtual room impacts the user's affective appraisal of the space in terms of how warm, spacious, intimate or exciting it is perceived to be. Dealing with a similar research question in a multimodal context, McArthur et al. argue that the spatial attention of the user in a virtual environment is coordinated across modalities that encode information differently, and that the prioritization of information in a virtual environment relies on the complex interactions between the individual modalities [4]. Accordingly, the current project adopts a cross-modal approach in its investigation of how the diegetic quality of cues in VR can affect the user experiences.

In 1995, Beroggi et al. hypothesized that VR can be used to support decision-making in emergency management, highlighting the potential of VR for training appli-

ISBN 978-84-09-08518-7 ISSN 2518-3672

Figure 1. An isometric view of the virtual room where the user study was conducted with its surrounding walls removed for this image. The user is tasked with picking a key and choosing a door to exit the room; while doing so, they are presented with conflicting suggestions in the form of diegetic and non-diegetic audio and visual cues.

cations [5]. Today, VR is viewed as a suitable platform for conducting behavioral studies including those that pertain to psychophysical operations [6,7] and decision-making [8–10]. Whereas previous projects have utilized VR simulations to study decision-making processes in life-critical situations [11,12], the current project adopts a user-experience approach to the study of decision-making in VR.

The distinction between diegetic and non-diegetic sounds are already used in games as narrative devices and means of diversifying the modes of interaction for the user [13]. Non-diegetic sounds, for instance, are often used to cue the players into certain actions, or alert them of state changes [14]. Summers and Jesse argue that the use of diegetic and non-diegetic sounds in VR are integral to conveying elements of narrative [15].

In their study of similar factors in the visual domain, Salomoni et al. suggest that the appropriate use of diegetic and non-diegetic visual interfaces in VR can have a significant impact on the sense of presence in immersive simulations [16]. Accordingly, Nielsen et al. identify the diegetic nature of a cue as a primary dimension in their taxonomy of cues for guiding the user's attention in VR [17].

Exploring methods of guiding user attention in cinematic VR, Rothe and Hußmann perform an analysis of users' viewing directions when they are given diegetic sound and lighting cues. The researchers find that objects connected with sounds attract more attention and guide the viewing more effectively [18]. Accordingly, Mateer characterizes the act of locating diegetic sounds in cinematic VR as a "natural tendency" [19].

In a study that explores the effects of diegetic and non-diegetic user interface elements on game immersion, Iacovides et al. find that while removal of non-diegetic interfaces can increase the level of cognitive involvement for expert users, it does not have a notable influence on the experience of novice users [20]. Exploring similar effects in VR, Cliburn and Rilea evaluate the impact of signage on navigation speed in VR. Comparing cases in which the users were able to rely on either diegetic signs or non-diegetic maps shown as head-up displays, the researchers find that subjects who navigate the world based on the signs are significantly faster than those who use maps [21]. These studies indicate how the diegetic qualities of elements presented to the users in simulated experiences can affect performance and engagement.

3. DIEGESIS IN VR

The concept of diegesis is applied to a variety of narrative forms, such as film, theatre, and games, to explain how elements of narration are situated in reference to the implied universe of a story [22]. While diegetic elements are those that belong to this universe, non-diegetic elements are external to it.

Diegesis gains further significance in VR studies due to the inherent affordances of the medium, such as constant immersion, dynamic first-person view, and interactivity. Whereas the user assumes an outside perspective towards both diegetic and non-diegetic elements in most narrative forms, VR can situate the user as a diegetic actor in its implied universe. For instance, in film-making, the perspective from which the audience observes a narrative is pre-determined by the director. In VR, the user can not only shift their perspective but also interact with the narrative [23]. The diegetic disposition of audio and visual elements in a simulation can therefore have an impact on the user's experience based on whether they are internal or external to the virtual space that the user occupies [20].

3.1 Visual Diegesis

In VR, the visual objects situated in the virtual space are inherently diegetic. These objects are part of the 3D environment and maintain their positions relative to the world-space of the VR. Diegetic visual objects are often affected by the physical forces implemented in the virtual environment, and interact with other objects accordingly. The lighting and other occlusion effects can alter the visibility of these objects.

On the other hand, non-diegetic visual objects are most commonly presented in the form of visual overlays. These can be head-up displays that relay relevant information about the VR, or user interface (UI) elements such as menus and buttons. Non-diegetic visual elements persist over the user's field of view and are commonly positioned relative the user rather than the world-space of the VR. These objects are usually unaffected by lighting and occlusion. Some non-diegetic UI elements can be spatially mapped into the virtual environment while remaining external to the implied universe of the VR.

3.2 Auditory Diegesis

Similar to visual elements of the same nature, diegetic audio objects in VR belong to the virtual environment. These sounds often originate from visual elements in the scene and are subjected to localization cues based on the user's position. These sounds are also affected by room acoustics and occlusion.

Non-diegetic audio objects in VR are unaffected by the user's position, room acoustics or occlusion effects. Much like non-diegetic visual objects, these are presented in an additional auditory layer that persists over the virtual world. These objects can be alert sounds, voice messages relayed to the user from outside of the virtual space, or backing tracks similar to the non-diegetic score of a film.

3.3 Sample Scenario: Park Simulation

In a VR that simulates an outdoor park, examples of diegetic visual objects could be the trees and the benches in the park. A non-diegetic visual object would be the map of the park presented as a head-up display. Whereas the user can look away from a bench in the scene, the map would always remain in the user's field of view.

In the same simulation, an example of a diegetic audio object would be the sound of a bird chirping. Attached to the diegetic visual representation of a bird, this sound would display localization properties relative to the user's position as the bird flies around. When the user reaches their destination at the far end of the park, a non-diegetic bell sound could indicate the successful completion of a task. This sound would be detached from the visual objects in the scene and heard without localization cues.

4. USER STUDY

We conducted a pilot study to evaluate how the modality and diegetic quality of cues in VR can affect the user's decision-making process. The study is designed as a game where the user is expected to make a series of decisions to exit a virtual room in presence of conflicting audio and visual cues of diegetic and non-diegetic nature.

4.1 Study Design

The game is designed with Unity and the HTC Vive System as a room-scale VR in a 5m by 5m open space. The virtual room seen in Fig. 1 is mapped onto this space. At one end of the room are two doors. At the other end is a table with two keys on it. The room is lit with a floor lamp placed in one of the corners. A loudspeaker is placed between the two doors above head-level. The purpose of the game is to pick a key and a door to exit the room.

The keys and the keyholes in the doors are the only interactable elements in the scene. The system registers a door selection once a key is inserted in one of the keyholes. When this selection is registered for the first time, the system resets the environment regardless of the selection, prompting the user to repeat the task. Once the system registers a door selection in the second attempt, the room is removed from the scene, effectively placing the user outside. This indicates that the user has successfully completed the task. We implemented the 2-attempt model to alleviate the effects of random decisions by encouraging the users to strategize over their decisions, and to monitor how users evaluate modal and diegetic pairings once a choice combination is perceived to be inaccurate.

Various combinations of audio and visual cues are presented in different configurations of the game. These cues include:

- **Diegetic visual (DV) cues**: a note on the table between the two keys, suggesting a key to pick, and a poster on the wall between the two doors, suggesting a door to choose;
- **Non-diegetic visual (NDV) cues**: head-up display messages that appear at the bottom of the screen when the user enters the collider around the table suggesting a key to pick, or when the user enters the collider around the doors, suggesting a door to choose;
- **Diegetic audio (DA) cues**: periodic announcements made through the loudspeaker in the room, suggesting a key to pick or a door to choose accompanied by classical music;
- **Non-diegetic audio (NDA) cues**: voice messages that are played-back when the user enters the collider around the table suggesting a key to pick, or when the user enters the collider around the doors, suggesting a door to choose.

The diegetic audio cues are spatialized binaurally using Google's Resonance Audio SDK for Unity. The cues are therefore subjected to room and distance effects, and are spatially mapped to the virtual speaker in the room. To better emulate the output of a loudspeaker, these sounds are given directionality characteristics so that they are heard more clearly when the user is within the dispersion field of the speaker. The cues are accompanied by classical music to give the user a constant sense of localized sound at the times the announcement is not being repeated. The music is ducked (i.e. attenuated) with a side-chain compressor when an announcement is being made in style of radio announcements. In configurations of the study where diegetic audio cues for both the keys and the doors are offered, the cues are compounded into a single announcement (e.g., "Pick the purple key and choose the door on the left").

The non-diegetic audio cues are played back without any localization or room effects. They are therefore detached from visual sources in the virtual room, and are perceived as originating from the user's current position. These cues are triggered each time the user enters one of the colliders surrounding the table or the doors. Both diegetic and non-diegetic audio cues are spoken by a neutral female voice.

Timestamps are generated when the user enters and exist the colliders surrounding the table and the doors. The times at which the user picks up a key, and inserts the key in a keyhole are also tracked. Additionally, the global time at which the game is started, and the time at which the game is reset for the second attempt are stored.

4.1.1 Scenarios

The study consists of 12 scenarios based on the different combinations of audio and visual diegetic and non-diegetic cues. These combinations are shown in Table 1. Between scenarios 1 and 2, 5 and 6, and 9 and 10, the cue types are swapped between the key and the door suggestions. In scenarios 3, 4, 7, 8, 11 and 12, the cue types are maintained across the two suggestions to control for within-modality and within-diegetic-quality conditions. Between scenarios 3 and 4, and 7 and 8, the diegetic quality of the cues are swapped while modalities are maintained, whereas between 11 and 12, the modalities of are swapped while diegetic qualities are maintained.

	Purple Key	Green Key	Left Door	Right Door
1	DA	NDA	DV	NDV
2	DV	NDV	DA	NDA
3	DA	DV	DV	DA
4	NDA	NDV	NDA	NDV
5	DA	NDV	NDA	DV
6	DV	NDA	DA	NDV
7	DV	NDA	NDA	DV
8	NDV	DA	DA	NDV
9	NDV	NDA	DA	DV
10	DV	DA	NDA	NDV
11	NDV	DV	DV	NDV
12	DA	NDA	NDA	DA

Table 1. Distribution of diegetic audio (DA), non-diegetic audio (NDA), diegetic visual (DV), and non-diegetic visual (NDV) cues in each of the 12 scenarios. Each participant is placed into one of these scenarios.

4.2 Participants

24 participants (15 male, 9 female; mean age: 23) took part in the current study. 12 participants described themselves as experienced VR users. 5 participants reported having tried VR before, while 7 indicated that this was their first time trying VR. The participants were evenly distributed across the 12 scenarios listed in Table 1.

4.3 Procedure

The study takes approximately 15 minutes to complete with 3 minutes for instructions and preparations, 2-3 minutes for the VR portion, and 10 minutes for the follow-up survey. In the instructions, the following items are communicated to the user:

- You will be placed in a virtual room that is mapped to the physical space of the study area. In this room, you will find two keys and two doors: only one of the keys will open only one of the doors, and you will be given two attempts to find the right combination to exit the room. There are no hidden clues or puzzles that can guide you out.

- You might, however, encounter various combinations of audio and visual cues that guide you in your decisions; these cues might conflict with each other and it will eventually be up to you to make a decision. When you insert a key in a keyhole, the system will register this as a decision.

The use of the Vive Controller to interact with the keys is demonstrated to the user. The user is then asked to take a seat in the designated chair, and put on the head-mounted display. The room-scale VR experience is initiated with the user seated in the virtual room configured for one of the 12 scenarios. The virtual chair is precisely aligned with the one in the real world. Once in VR, the user can explore the room, and make decisions relevant to the task without a time constraint. When the first pair of decisions is made, the game gets reset with the same cues in place, indicating to the user that they did not make it out of the room. When the second pair of decisions is made, the user is placed outside, indicating that the game is successfully completed.

As a follow up to the VR portion of the study, the user is invited to respond to a qualitative survey in an interview format. In this survey, they are asked to identify the audio and visual cues they encountered in the scene, where they thought these cues might have originated from, and whether the cues were correlated in any way. They are then asked to describe their thought process in making each of their decisions in VR.

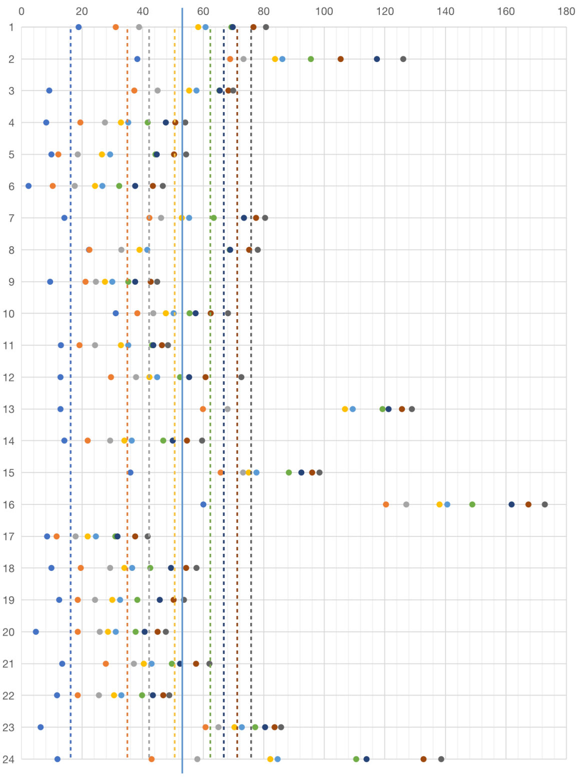

Figure 2. Action timings for each participant with the Y-axis indicating participant number and the X-axis indicating time in seconds. The consecutive dots from left to right indicate the times for: entry to the collider around the table, key pick-up, entry to the collider around the doors, key insertion in a keyhole, scene restart, second entry to the collider around the table, second key pick-up, second entry to the collider around the doors, second key insertion in a keyhole. The dashed lines indicate average times for each of these actions with the average scene restart time indicated with a solid blue line.

4.4 Results and Discussion

The users described their overall experience as responsive and convincing. None of the users reported having experienced discomfort or hard time interacting with the system. During the post-study interview, all users correctly identified the cues they encountered in VR.

The timings of the actions performed by each user are shown in Fig. 2. The average time to complete the study was 75.8 seconds (SD = 34.2). First-time users have completed the study in 56 seconds on average (SD = 8.7). Those with prior experience with VR completed the study in 84 seconds on average (SD = 37.6). The average action times shown in Fig. 2 indicate that the users spent the most amount of time during the initial exploration of the VR, and while making the decision to pick up a key for the first time. The increased speed in choosing a door observed with some users can be attributed to the fact that certain scenarios expose the users to diegetic cues about doors as soon as the game is started.

In repeated conditions (i.e. scenarios 3, 4, 7, 8, 11, 12 seen in Table 1), 9 out of 12 users followed the same cue type (e.g., DA in Scenario 3) for the first set of decisions, and followed the other cue type (e.g., DV in Scenario 3) in their second attempt. All 3 users who followed mixed cue types in their first attempt (e.g., DA and DV in Scenario 3) indicated that they thought the system was trying the deceive them. While one of these users followed a matching cue type in their second attempt, the other 2 users continued to follow mixed cue types in their second attempts. These users expressed that even though they imagined different instigators behind the audio and visual cues, they thought all of these characters were trying to trick them.

	DA	**NDA**	**DV**	**NDV**
Key Decision	25%	75%	75%	25%
Door Decision	25%	75%	25%	75%
	DA	**DV**	**NDA**	**NDV**
Key Decision	25%	75%	50%	50%
Door Decision	75%	25%	50%	50%
	DA	**NDV**	**DV**	**NDA**
Key Decision	100%	0%	50%	50%
Door Decision	100%	0%	75%	25%

Table 2. Frequency distribution (%) of decisions made in the first attempts for each conflicting cue pairs where 25% represents a decision by 1 participant.

Some users interpreted the diegetic visuals as clues left behind by game characters who had previously gotten out of the room, and the non-diegetic visuals as originating from an authority figure who designed the game. Other users have also expressed that the sense of an external authority was a decisive factor.

Table 2 shows the decision rates for each of the 6 cue combinations (i.e. conflicting concurrent cues) that were presented in relation to the key and the door decisions. A consistent preference is not observed in the conditions where diegetic qualities are matched (i.e. DA-DV, and NDA-NDV). In the audio cue combinations (i.e. DA-NDA), a preference towards non-diegetic cues can be seen. A strong preference towards diegetic audio cues is observed in the DA-NDV combination.

Once the initial decisions were found out to be wrong after the first attempt, the users followed a variety of approaches in their second attempts: 7 users chose the same key but a different door, 6 chose a different key but the same door, and 11 changed both of their decisions. Some of these decisions were reported to be based on maintaining a diegetic consistency in the first attempt and a modal consistency in the second. Table 3 shows the instances where the users have maintained a pairing in terms of modality or diegetic quality between their decisions about the key and the door. For instance, User 18 followed the diegetic (i.e. DV and DA) cues for both decisions in their first attempt, while they followed the visual (i.e. DV and NDV) cues in their second attempt. Overall, diegetic-quality pairings were more common in the first attempts, whereas visual pairings were a prominent choice in the second attempts.

The users' perception of whether the cues were addressed

Scenario	1st Attempt				2nd Attempt			
DA-NDA; DV-NDV	1	-	13	ND	1	D	13	-
DV-NDV; DA-NDA	2	-	14	-	2	ND	14	-
DA-DV; DV-DA	3	-	15	-	3	V	15	-
NDA-NDV; NDA-NDV	4	V	16	A	4	-	16	V
DA-NDV; NDA-DV	5	D	17	D	5	V	17	V
DV-NDA; DA-NDV	6	A	18	D	6	V	18	V
DV-NDA; NDA-DV	7	DV	19	ND	7	NDA	19	-
NDV-DA; DA-NDV	8	DA	20	NDA	8	-	20	-
NDV-NDA; DA-DV	9	V	21	-	9	A	21	V
DV-DA; NDA-NDV	10	-	22	-	10	A	22	V
NDV-DV; DV-NDV	11	-	23	D	11	-	23	ND
DA-NDA; NDA-DA	12	-	24	D	12	D	24	ND

Table 3. Modality and diegetic-quality pairings formed by the participants between their key and door choices in the first and second attempts. Scenarios highlighted in gray consist of repeated conditions where modality or diegetic-quality pairings may not be possible in the first place.

to them privately or in the form of public messages had an impact on how they evaluated the suggestions. For instance, a user placed in a DA-DV scenario thought the "silent" note on the table was more trustworthy than the announcement that others could potentially hear. Conversely, in every condition where an NDV cue was presented in conflict with a DA cue, the users preferred the latter, stating that the NDV cue seemed "robotic", "inhuman" and "unidentified".

Furthermore, NDV was the least preferred cue type for the first key choice overall with only 3 users following NDV cues for this choice. One of these users expressed having trusted the DV cue first but decided that the NDV cue felt like a more recently updated source. Another user reported not having paid attention to the NDV cue in their decision. The third user mentioned having decided to go with a modal parity; therefore, the DV cue about the doors motivated their preference towards the NDV cue with the keys.

Some users expressed a preference for diegetic cues, mentioning that these were "actual sources" in the room, whereas non-diegetic sources were "surreal" or "unidentified". On the other hand, after following a diegetic cue for the key choice, some of these users maintained the modality of this cue in their door choices, even though this meant following a non-diegetic cue.

5 out of 18 users who were placed in a scenario that consisted of a diegetic audio cue expressed uncertainty about whether the diegetic sound was coming from the speaker in the scene. However, these users described the cue with such terms as "intercom", "announcement", "PA sound", and "in the space", which indicate an association of the sound with the environment. As in most modern applications of binaural audio, a standardized head-related transfer function (HRTF) was used for audio spatialization in the current study. Standardized (i.e. non-individualized) HRTFs can be prone to localization errors due to disparities between the head model used in the function and the user's anatomical features [24]. This might explain the cases in which the users were not certain about the spatial correspondence between a diegetic sound and the speaker model in the scene.

There was a notable preference towards NDA cues when presented in conflict with the DA cues. The terms used to describe the NDA cues were "Voice of God", "narration", "exogenous", "in the head" and "disembodied". These descriptors indicate an interpretation of NDA cues as originating from an external source. While this non-diegetic quality was a cause of disinclination to NDV cues, it served a reassuring role for some users when presented in the audio domain.

5. CONCLUSION

VR is transforming our relationship with arts and entertainment: whereas the spectators have often been situated as consumers of such spectacles, VR offers experiences where they can be users and even performers. This distinction brings about many new and interesting challenges for VR researchers and content creators. An increasing number of studies from a wide range of fields, such as human-computer interaction, user experience design, psychology and serious games, are now dealing with such challenges.

The preliminary results gathered from the current study indicate that the diegetic quality of a cue in VR can have a noticeable effect on the decisions made by the users when presented with conflicting cues to complete a task. We believe that these results can inform the design of interactive VR experiences, such as those afforded by games and VR films.

While some users followed modal parities in their decisions, others paired cues in terms of diegetic quality. A strong preference for diegetic audio cues over non-diegetic visual cues was observed. Another notable preference was towards non-diegetic audio cues over diegetic audio cues. Despite these correlations, a preference was not observed between non-diegetic cues of different modalities.

We plan to expand our sample size in future iterations of the study in order to improve the statistical significance of our results. We also plan to investigate the narrative function of cues in isolation rather than in conflicting situations to better understand between-subject preferences towards modality and diegetic qualities. When paired with the qualitative feedback from the users, the current results show that the diegetic quality of a cue can serve as a conceptual indicator that affects the user's narrative interpretation of a VR experience and how they group elements of the VR. This affect was observed when users commonly associated the cues with such concepts as trust, authority, deception, and privacy.

We hope to further explore such conceptual implications of the modal and diegetic qualities of the elements in VR that can facilitate the design of compelling user experiences. Moreover, we plan to incorporate Witmer and Singer's presence questionnaire [25] in our survey to control for the effects of such elements on perceived immersion.

ISBN 978-84-09-08518-7 ISSN 2518-3672

6. REFERENCES

[1] R. Yao, T. Heath, A. Davies, T. Forsyth, N. Mitchell, and P. Hoberman, "Oculus VR best practices guide," 2014.

[2] Google, "Designing for google cardboard," https://designguidelines.withgoogle.com/cardboard/, Accessed on 2018-10-28.

[3] A. Naz, R. Kopper, R. P. McMahan, and M. Nadin, "Emotional qualities of vr space," in *Proceedings of IEEE VR Conference 2017*. IEEE, 2017, pp. 3–11.

[4] A. McArthur, R. Stewart, and M. Sandler, "Sounds too true to be good: diegetic infidelity–the case for sound in virtual reality," *Journal of Media Practice*, vol. 18, no. 1, pp. 26–40, 2017.

[5] G. E. Beroggi, L. Waisel, and W. A. Wallace, "Employing virtual reality to support decision making in emergency management," *Safety science*, vol. 20, no. 1, pp. 79–88, 1995.

[6] R. Skarbez, S. Neyret, F. P. Brooks, M. Slater, and M. C. Whitton, "A psychophysical experiment regarding components of the plausibility illusion," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 4, pp. 1369–1378, 2017.

[7] T. Ye, S. Qi, J. Kubricht, Y. Zhu, H. Lu, and S.-C. Zhu, "The martian: Examining human physical judgments across virtual gravity fields," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 4, pp. 1399–1408, 2017.

[8] M. Minderer, C. D. Harvey, F. Donato, and E. I. Moser, "Neuroscience: virtual reality explored," *Nature*, vol. 533, no. 7603, p. 324, 2016.

[9] J. Smith, "Immersive virtual environment technology to supplement environmental perception, preference and behavior research: a review with applications," *International journal of environmental research and public health*, vol. 12, no. 9, pp. 11 486–11 505, 2015.

[10] F. La Paglia, C. La Cascia, R. Rizzo, G. Riva, and D. La Barbera, "Decision making and cognitive behavioral flexibility in a ocd sample: a study in a virtual environment," in *Annual Review of Cybertherapy and Telemedicine 2015: Virtual Reality in Healthcare: Medical Simulation and Experiential Interface*. IOS Press, 2015.

[11] F. Kaneko, Y. Ikemoto, and M. Fukumoto, "Evacuation simulator for analysis of evacuees' decision in a ship under casualty," in *Proceedings of IEEE VR Conference 1999*. IEEE, 1999, p. 80.

[12] U. Ju, J. Kang, and C. Wallraven, "Personality differences predict decision-making in an accident situation in virtual driving," in *Proceedings of IEEE VR Conference 2016*. IEEE, 2016, pp. 77–82.

[13] I. Ekman, "Meaningful noise: Understanding sound effects in computer games," in *Proceedings of Digital Arts and Cultures*, 2005.

[14] A. Järvinen, "Gran stylissimo: The audiovisual elements and styles in computer and video games." in *Proceedings of Computer Games and Digital Cultures Conference*, 2002, pp. 113–138.

[15] C. Summers and M. Jesse, "Creating immersive and aesthetic auditory spaces in virtual reality," in *3rd IEEE VR Workshop on Sonic Interactions for Virtual Environments (SIVE) 2017*. IEEE, 2017, pp. 1–6.

[16] P. Salomoni, C. Prandi, M. Roccetti, L. Casanova, L. Marchetti, and G. Marfia, "Diegetic user interfaces for virtual environments with hmds: a user experience study with oculus rift," *Journal on Multimodal User Interfaces*, vol. 11, no. 2, pp. 173–184, 2017.

[17] L. T. Nielsen, M. B. Møller, S. D. Hartmeyer, T. Ljung, N. C. Nilsson, R. Nordahl, and S. Serafin, "Missing the point: an exploration of how to guide users' attention during cinematic virtual reality," in *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*. ACM, 2016, pp. 229–232.

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

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

[20] I. Iacovides, A. Cox, R. Kennedy, P. Cairns, and C. Jennett, "Removing the hud: the impact of non-diegetic game elements and expertise on player involvement," in *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*. ACM, 2015, pp. 13–22.

[21] D. C. Cliburn and S. L. Rilea, "Showing users the way: Signs in virtual worlds," in *Proceedings of IEEE VR Conference 2008*. IEEE, 2008, pp. 129–132.

[22] G. Genette, *Narrative discourse: An essay in method*. Cornell University Press, 1983.

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

[24] S. Zhao, R. Rogowski, R. Johnson, and D. L. Jones, "3d binaural audio capture and reproduction using a miniature microphone array," in *Proceedings of Digital Audio Effects (DAFx) Conference 2012*, 2012, pp. 151–154.

[25] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence*, vol. 7, no. 3, pp. 225–240, 1998.