

Decoding the User Experience in Mobile Virtual Reality Narratives

Biswajit Sarker^(✉)

Uppsala University, Uppsala, Sweden
thisisbish@gmail.com

Abstract. Virtual reality narratives are slowly gaining popularity due to the rise of consumer oriented mobile VR headsets. Based on its potential, movie studios and other entertainment companies are investing heavily in this sector. However there is a gap in research that focuses on the user experience in these mobile VR narratives. This paper aims to fill this gap by presenting the findings of a study that investigated user experience in virtual reality narratives. Using qualitative research methods, this study analyzed users' experiences and explored what design factors contribute to their positive or negative experiences. Based on a thematic analysis of collected data, this study argues that audio-visual cues play an important role on how users perceive a VR narrative. Findings also suggest that the sweet spot of user experience lies between boredom and frustration. If the virtual world provides well-designed audio-visual cues to guide users' attention throughout the narrative then users experience immersion and spatial presence. On the contrary, lack of cues in a virtual environment fails to utilize the space available around the users resulting in their boredom. Moreover, excessive use of audio-visual cues makes users to keep switching their attention from one element to the next in fear of missing out something important, eventually resulting in their frustration and stress. Based on these findings, this paper suggests some guidelines to improve user experience in mobile VR narratives.

Keywords: Mobile VR · User experience · Qualitative methods

1 Introduction

To explore the space around us in real life, we use sensory cues in the environment that indicate the state of some property of the world that might be important to us. By following different sensory cues such as visual, auditory, haptic, olfactory and environmental cues; we perceive the world around us through active exploration [1]. But mobile virtual reality Head Mounted Devices (HMD) such as Google Cardboard and Samsung Gear VR do not offer positional tracking capabilities to track user's movement in a virtual environment [2]. As a result users cannot walk around to explore any virtual space with mobile HMDs. Furthermore, these devices do not have the capabilities to provide feedback through touch, smell or taste either. This means, all the interactions possibilities need to be conveyed through visual and audio cues in a mobile virtual reality narrative. This makes it a necessity for the designers to understand how these audio-visual cues affect user experiences in mobile VR narratives.

Currently virtual reality is seeing a breakthrough at the consumer market due to the release of HMDs such as Google Daydream View, Samsung Gear VR, Oculus Rift and HTC Vive to name a few [2]. These devices are expected to remove the barriers of screen-based narratives and put the users in the middle of the content. Usually these VR narratives include a storyline that has a specific beginning and ending. When the audience put on their VR headsets, they are transported into an immersive virtual environment and even though they can look around in any direction at any time, the designers of these narratives want them to pay attention to specific elements in that virtual world that are important to stitch the story together. These narratives are designed to give the audience the feeling of being a part of the story. But being inside an immersive world invites the audience to look around and since they are in an unknown environment, they are immediately filled with questions such as, where am I? What's going on around me? What should I do next? Then it becomes the responsibility of the designers to answer these questions quickly so that the audience can focus on the major elements of those narratives. This is where it becomes important [3] to find out, what drives the audience to pay attention to the right elements in virtual reality narratives.

This explorative study investigates, given a 360° virtual space that allows looking at in any direction while users can only focus on 90° to 110° at any given time [4]:

- How do people know where to look and how to proceed when presented with an immersive virtual narrative?
- How the audio-visual cues contribute to their experience?
- What design factors contribute to a positive experience and what design factors contribute to a negative experience for the user?

By building on existing theoretical accounts related to user experience in virtual reality and through a field based study using a popular consumer oriented device used by general consumers, this study analyzes users' experiences from their point of view. Based on a thematic analysis of collected data through observations, think aloud and semi-structured interviews from 10 participants between the ages of 22 to 29, this study argues that when users are experiencing an immersive virtual reality narrative their curiosity drive them to look for clues to follow in that virtual environment. If the virtual world provides well-designed audio-visual cues to guide their attention throughout the narrative then users feel immersion and spatial presence. On the contrary, lack of cues in a virtual environment keeps users looking straight ahead throughout the narrative, which results in their boredom. When it comes to lack of positional tracking in mobile VR, it can be asserted from the findings that users get confused and frustrated when their movements in the physical environment result in the whole virtual environment moving with them. This clash between real world perception and virtual world realization breaks their feeling of spatial presence in that VR environment. Finally, findings suggest excessive use of audio-visual cues forces users to keep switching their attention in multiple directions in fear of missing out something important, eventually resulting in their frustration and stress.

This paper follows the subsequent structure. The background section examines relevant literature and establishes a gap in research. Section three discusses the methodology used in this study. Section four presents the findings. Section five

discusses the findings in detail. And the final section includes conclusion, limitations and suggestions for future research.

2 Background

2.1 Immersion in VR Narratives

When it comes to analysis of experiences, one of the terms quite frequent in literature is flow [5], which describes the mental state in which a person is fully engaged in an activity by a feeling of intense involvement and energized focus. During a flow experience a person feels in control, loses sense of surroundings and his or her awareness is narrowed down to the activity itself. Sutcliffe [6] argues that experience in a virtual world can be explained in terms of flow. He asserts that when the virtual world is well designed, a person in that virtual world feels immersed in a strong sense of presence and the mediating virtual reality device and computer essentially disappears. Compared to screen based media like cinema or television, virtual reality as a media is much more flexible where audience can change their vantage point at any given moment. While more control in a virtual narrative experience provides greater feeling of involvement for the participants [7], this also takes the control away from the designers, who want the audience to follow a specific storyline. This presents a big challenge for the designers of virtual reality narratives.

2.2 Role of Audio-Visual Cues in Perception

Since virtual environments are usually representations of real world environments, it is reasonable to follow the ecological psychology approach proposed by Gibson [1], which describes how different structures in the external world guide people's everyday actions. According to Gibson's theory, we perceive the world around us through our actions. We turn our heads to direct our attention to different visual stimuli and we focus our attention to hear better and gather information about action possibilities available around us [1]. Many researchers consider this to be more relevant to HCI than classical cognitive theories [8–10]. When it comes to spatial cues available in the environment for perceiving the world around us, previous studies indicate most cues are linked to the visual modality; for example aerial perspective and relative brightness [1]. Along the same line spatial audio plays a big role in directing users' involuntary attention towards possibilities of action within the virtual environment [11].

2.3 Role of Audio-Visual Cues in Spatial Presence

Wirth et al. [12] describes spatial presence as a two-step process. On the first step the user draws upon available spatial cues to perceive the virtual environment as a plausible space. The virtual environment will more likely be perceived as a plausible space if these audio-visual cues are both rich in quality and have a logical consistency. On the second step, the user experiences herself as being located within that perceived space

by discovering possibilities of action within the virtual environment. Existing literature suggest [13, 14] a steady stream of highly detailed information flow supported by appropriate audio-visual spatial cues effectively builds the virtual environment as a plausible place and increases the experience of spatial presence for the users. However, an excessive use of spatial cues can cause sensory overload and produce fatigue for the users [15].

2.4 A Gap in Research

Even though research in virtual reality has been conducted on perception, immersion and spatial presence, little is known about how users decide where to look and how to proceed in an immersive VR narrative and what design factors contribute to that decision. There is also a gap in research that focuses on users' experience with a consumer oriented mobile virtual reality HMD which only offers applications that are purely narrative or have severely restricted interaction possibilities. Hence research is needed to explore how general consumers who have no experience or very little experience with mobile virtual reality applications perceive this new media. This study aimed to fill this gap in research.

3 Methodology

To follow a well-defined scientific research methodology for analyzing users' experience with virtual reality narratives and to theorize a set of propositions about those experiences, this study followed a theory informed inductive approach. Since qualitative studies help to uncover and interpret participants' understanding of the phenomenon that they are involved in [16], it was a good fit for an explorative study like this one where participants' behavior in a virtual reality narrative was being investigated.

3.1 Data Collection

Csikszentmihalyi and Robinson [17] argue that since experiences are subjective phenomenon that cannot be externally verified, a researcher has to rely on the testimonies given by the participants. They also downplay the validity of relying on physiological measures alone to collect data to explain users' experience [17]. Since most of the challenges and opportunities associated with users' experience in virtual reality were not directly observable, in depth semi-structured interviews were chosen as the main source of data collection for this study. Unlike surveys or questionnaires, in depth interviews are flexible, dynamic and those provide a more valid insight into the user's perception of reality [18]. In qualitative studies that use semi-structured interviews, the primary instrument of data collection is the researcher [19]. This is important in capturing the subject's point of view as argued by some researchers [20] who assert that due to the use of remote, inferential empirical materials; quantitative researchers seldom capture the users' point of view of an experience.

3.2 VR Applications Used in This Study

The following virtual reality applications were used during the study. “Oculus home” is the central interface through which other applications can be found and downloaded. The reason behind using existing VR applications was to collect data from professionally designed immersive narratives. Since the focus of the study was to investigate how users experience a virtual reality narrative and what role audio-visual cues play in those experiences; it was important that they were not using low quality prototypes that might not provide accurate data for analysis (Table 1).

Table 1. Summary of virtual reality applications used in this study.

| App | Type | 360° view | Spatial audio | Supported interaction |
|------------------|----------------|-----------|---------------|--|
| Oculus home | Interface | Yes | No | Head tracking in all directions Touchpad for selection |
| Muse Revolt | Music video | Yes | No | Head tracking in all directions |
| Song for someone | Music video | Yes | Yes | Head tracking in all directions |
| Invasion | Animated short | Yes | Yes | Head tracking in all directions |
| Rosebud | Animated short | Yes | No | Head tracking in all directions Scroll touchpad to zoom Tap touchpad to control camera angle |

3.3 Equipment and Setup

This study was conducted in a living room set up to ensure privacy of the participants and also to provide them with an environment where a mobile VR headset was most likely to be used. Participants were invited one person at a time to ensure they can act and talk freely. The HMD used in the research was a Samsung Gear VR coupled with a Samsung Galaxy S6 edge mobile phone. This HMD supports 3 degrees of Freedom (DOF) with a Field of view (FOV) of 96°. To avoid ambient noise an in-ear headphone was used during the experiments. To ensure anonymity, a list of randomized participant IDs were prepared and assigned to each participant. Participants were given a consent form that described the research study in a nutshell. It was made sure participants could exit the experiment at any time. They were informed that the follow up interviews would be audio recorded. Each participant had to read and sign the consent form before participating in the study.

3.4 Semi Structured Interviews

In total, 10 participants between the ages of 22 to 29 participated in the study. Five of the participants were male and five were female. All trials followed the same structure. First, the researcher introduced the HMD to the participants with a quick demonstration of how it works. The participant then put on the headset, potentially aided by the

researcher. Each of the participants was then instructed to explore the Oculus Home interface for a few minutes and then try out two of the applications selected randomly from the list above. All the participants were interviewed right after they had completed each of the applications. After the experiment, the participant was thanked for his or her time and was debriefed about the purpose of the study.

3.5 Data Analysis

All the recorded interviews were transcribed in details and coded using qualitative analysis software “NVivo”. After coding, the relevant data extracts were collated for analysis to find recurring themes from the dataset. Thematic analysis method was used to analyze the recorded data following the guidelines suggested by Braun and Clarke [21]. For each code, relevant data extracts were reviewed and compared against the whole data set to make sure the emerging themes make sense and no data extracts were being taken out of context from the interview transcript.

4 Results and Analysis

4.1 Specific Observations

After coding the interviews, think aloud data and observation notes; the entire data set was reviewed to identify themes relevant to positive or negative experiences of the users. During this phase of analysis, the use of audio-visual cues to attract and direct user’s attention stood out to be one of the most important design factors affecting the user’s experience during the narrative. The following table provides a short summary of the initial themes along with examples of their relevant data extracts. A detailed thematic analysis can be found here [22] (Table 2).

Table 2. Initial themes from the dataset with salient data extracts from the interviews.

| Data extract | Themes | Relevant application |
|---|--|----------------------------|
| <i>I knew where to look by the sound. And then I see something moving and looked that way I didn't see the rabbit coming out of it's cave but then I heard the sound to the left and I looked in the scenery and then I saw the rabbit</i> | Users' attention is directed by the audio-visual cues available in the virtual environment while they try to follow the VR narrative | Invasion, Song for Someone |
| <i>I thought (curse word) if I am here in virtual reality I am supposed to be able to turn around and see new things! You sort of want to look everywhere. It's like I am looking here and then .. Oh...It's over there!</i> | Users expect to see events all around them in a virtual environment since the application puts them in a 360-degree immersive world | All applications |

(continued)

Table 2. (continued)

| Data extract | Themes | Relevant application |
|---|--|----------------------------|
| <i>I looked around just because I wanted to see what's around me and I kept going I really did not feel the need to turn around I did that because I was just curious</i> | Out of curiosity users explore the VR environment in different directions | All applications |
| <i>Like if you start in the middle of the room may be you can back off and you have space to move ... I think that's the thingyou are just stuck in the middle and it's kinda creepy I don't know, I was trying to walk in the world just to see if something happens. But nothing happened and it felt weird!</i> | Due to the lack of positional tracking in mobile VR, user's movement is not tracked which breaks user's expectation resulting in a negative experience | All applications |
| <i>.... if you try to..ummm...just look around and forget about the reality, I think you are able to have this nice feeling It's cool that you can look everywhere. (laughter) Oh, This is really cool</i> | User has a positive experience if she feels immersed in the VR environment where things are happening all around | Invasion, Song for Someone |
| <i>No there wasn't any clue (to look around) Everything was happening in front of me more or less there was a clear menu just in front of you so there's no reason to look around</i> | If there's no possibilities of action in the environment or enough cues to attract user's attention in different directions, then user ends up looking straight ahead throughout the narrative | Rosebud, Oculus Home |
| <i>.... why not use the whole space ... I mean there's so much space around you for options and buttons and so on Well you see part of the room behind the menu, so I looked around a little bit. But nothing new was moving around</i> | User doesn't tend to look around when there are no clues available for the user to perceive the affordances offered by the environment | Oculus Home |
| <i>For this one it felt like... hmmm... how to say... ummm... more like a 3D video than virtual reality That was more like a TV screen I guess</i> | Without any audio or visual cues to state otherwise, users just focus on the field of view looking at an conceptual screen ignoring the rest of the space that is hidden from the view | Oculus Home, Rosebud |

(continued)

Table 2. (continued)

| Data extract | Themes | Relevant application |
|--|---|----------------------------|
| <i>Everything was happening in front of me and it was very boring I have to say Well, I am looking at a child watering a flower.... It's a bit dull!</i> | Due to the lack of use of the space surrounding the user, immersion is minimal and users end up having a boring experience in the VR environment | Rosebud |
| <i>I missed that the video was interacting with you... ummm... more like a 3D video than virtual reality where you are inside I felt like I was more watching from outside</i> | Spatial presence is minimal; user feels like looking at a giant screen from outside | Oculus Home, Rosebud |
| <i>There's a lot of stuff going on! It's hard to focus on one thing. There's a ... a lot of.... a lot of, OK I want to follow the flow of people but it's hard It's a bit difficult to know where to focus cause you can't see everything at once. You sort of want to look everywhere. It's like I am looking here and then ..Oh...It's over there!</i> | User has trouble keeping up with the narrative in Fear of Missing Out something important when there are too many audio-visual cues around | Muse Revolt |
| <i>I would say this video has the worst quality and the video where more things happening (laughter) may be too many things happening. I think you want to check everything and its overloaded and may be a little stressful</i> | User feels frustrated and stressed due to lack of control when audio-visual cues do not provide enough guidance | Muse Revolt |
| <i>Its good that I am in control and don't miss out or feel like missing out I kinda felt like I was in the same room with them. That was really nice It felt like I had a presence in the story cause the rabbit was kinda hiding behind me</i> | When user's attention is properly guided through well-designed audio-visual cues, user feels immersed in the environment, which results in spatial presence and enjoyment | Invasion, Song for Someone |

These initial themes were then reviewed to generate refined themes for further analysis. After naming, collating, defining and refining the specifics of each theme [22], the final thematic table was constructed to find patterns of answers for the research questions (Table 3).

Table 3. Refined themes from dataset with coded modes of user engagement.

| Initial themes | Refined themes | Final theme |
|--|---|--|
| Users expect to see events all around them in a virtual environment since the application puts them in a 360-degree immersive world | Out of curiosity and due to familiarity with a 360-degree real life environment, users expect to see events happening all around them | Users look for clues to follow and they experience immersion when their attention is directed to different elements of the VR narrative all around them through audio-visual cues |
| Out of curiosity users explore the VR environment in different directions | | |
| Users' attention is directed by the audio-visual cues available in the virtual environment while they try to follow the VR narrative | Users feel immersion if their attention is directed through audio-visual cues available around them in the virtual environment | |
| User has a positive experience if she feels immersed in the VR environment where things are happening all around | | |
| If there's no possibilities of action in the environment or enough cues to attract user's attention in different directions, then user ends up looking straight ahead throughout the narrative | Users keep looking straight ahead at a conceptual screen when there are no audio-visual cues available around them to attract their attention | When there are no audio-visual cues available to attract and direct their attention, users get into a "Screen" mode of experience where they keep looking straight ahead at a conceptual screen resulting in minimum immersion, boredom and lack of spatial presence |
| User doesn't tend to look around when there are no clues available for the user to perceive the affordances offered by the environment | | |
| Without any audio or visual cues to indicate otherwise, users just focus on their field of view looking at an conceptual screen ignoring the rest of the space that is hidden from the view | Due to lack of interactivity around them, users feel minimum immersion and spatial presence and the experience becomes boring pretty quickly | |
| Due to the lack of use of the space surrounding the user, immersion is minimal and users end up having a boring experience in the VR environment | | |
| Spatial presence is minimal; user feels like looking at a giant screen from outside | | |

(continued)

Table 3. (continued)

| Initial themes | Refined themes | Final theme |
|---|--|---|
| User has trouble keeping up with the narrative in Fear of Missing Out something important when there are too many audio-visual cues around | Too many audio-visual cues to direct users’ attention make them to use extra efforts in Fear of Missing Out something important in the narrative | Excessive use of audio-visual cues put the users in a “ Fear of Missing Out (FOMO) ” mode of experience that results in frustration and stress |
| User feels frustrated and stressed due to lack of control when audio-visual cues do not provide enough guidance | Users feel like they are not in control, which results in their frustration and stress | |
| When user’s attention is properly guided through well-designed audio-visual cues, user feels immersed in the environment, which results in spatial presence and enjoyment | Well-designed audio-visual cues guide users’ attention in different directions all around them throughout the narrative | Well-designed audio-visual cues put users in a “ Guided ” mode of experience throughout the narrative resulting in their immersion, spatial presence and overall enjoyment |
| | Users feel immersion, spatial presence and enjoyment when attention is guided to different elements of the narrative happening all around them | |
| Due to the lack of positional tracking in mobile VR, user’s movement is not tracked, which breaks user’s familiarity with physical reality resulting in a negative experience | Users have a negative experience when their movement doesn’t match with the movement in the VR narrative | Lack of positional tracking in mobile VR requires additional time for the user to get used to the virtual environment |

4.2 Analysis of Results

For all the applications when participants were placed inside the virtual narrative, they explored the space around them out of curiosity and looked for anything that grabbed their attention. When they found something that caught their attention they kept looking in that direction way until their attention was directed to some other element in the environment by a visual or audio cue. In all the applications used during the study, due to the lack of positional tracking in Mobile VR, when the participants moved during the narrative, the whole VR environment moved with them. It came as a surprise for the participants since they were expecting their movement to be tracked inside the virtual environment. It took a little bit of time for the participants to get used to this conflict between expectation and reality, but once they got used to their movement, the participants had no further difficulties with following the narratives.

When participants were exploring the *Oculus Home* Interface, the participants felt like there was a big screen in front of them and they kept looking straight ahead towards that conceptual screen throughout the experience since no audio-visual cues directed their attention towards any other element on the surrounding space. To match with the participants' experience, this mode of engagement has been coded as the "screen mode".

Participants also experienced this "screen" mode of engagement in *Rosebud*. Once they were placed inside the narrative, the participants looked all around out of curiosity but the only element that caught their attention was the asteroid in front of them. Once they started focusing in front, no other audio-visual cues directed their attention to any other direction in the narrative. The participants expressed that there was not much going on around them and they got bored pretty quickly. It is interesting to note that even though *Rosebud* offered the largest possibilities for interaction by letting the audience change camera angles, the participants still got stuck in the "screen" mode since everything was happening in one direction, and they were unable to affect the storyline even by changing camera angles.

In the case of *Muse Revolt*, participants experienced a different mode of experience. When they got into the immersive world of this VR music video, multiple visual elements started attracting their attention at the same time. First they saw the band performing on the stage, but their attention quickly got directed to the groups of people running all around them. While they were trying to follow the groups to find out what's going on, their attention got directed again by several police cars coming into the scene. While several visual cues tried to catch their attention at the same time, lack of directional audio cues made it even harder for the participants to decide what element of the narrative to focus on, which made them to try and follow too many random cues in Fear Of Missing Out (FOMO) something important around them. Due to this excessive use of visual cues and lack of any sort of guidance, eventually they got frustrated and expressed that there was simply too much going on throughout the narrative. To match with the participants' experience, this mode of engagement has been coded as the "FOMO mode".

In the case of *Invasion* and *Song for Someone*, once the participants entered the VR environment, visual and directional audio cues directed their attention to the first element they needed to focus on. From that point onwards their focus was guided throughout the narrative from one element to the next. The audio-visual cues were well designed to make sure multiple cues were not asking for attention at the same time. The participants felt guided throughout the experience and they followed the audio-visual cues all around them without much effort. Since the narratives were gradually unfolding all around them, the participants felt immersed in those narratives. They also expressed the feeling of "being there" in those virtual environments. To match with the participants' experience, this mode of engagement has been coded as the "guided mode".

The following table lists the modes participants engaged in throughout different narratives (Table 4):

Table 4. Modes the users engaged in with different applications.

| Narrative | Mode of engagement | User experience |
|----------------------------|--------------------|-----------------------------|
| Rosebud | Screen | Boredom, Lack of Presence |
| Muse Revolt | FOMO | Frustration, Stress |
| Invasion, Song for Someone | Guided | Immersion, Spatial Presence |

One interesting finding from this study is the mismatch between user's real world perception and virtual world realization due to the lack of positional tracking in mobile VR. This results in users requiring some additional time to get used to the VR environment.

5 Discussion

From the inductive analysis of user experiences with mobile virtual reality narratives used in this study it can be hypothesized that, users have an overall positive experience when there are well-designed audio-visual cues available throughout the experience that put them in a "guided" mode which follows the narrative flow. In this scenario users feel like they are in control, they know where to look and how to follow the cues and they are not missing out anything important. They also feel immersed in that virtual environment which gives them a feeling of spatial presence. It can also be suggested that excessive use of audio-visual cues or poorly designed cues put the users in a mode of engagement where they try their best to follow the cues in Fear Of Missing Out (FOMO) something important and end up feeling stressed and frustrated with the overall experience. They feel there is too much going on and they have no control over the experience. Finally, it can be implied that, lack of audio-visual cues throughout the VR narrative puts the users in a mode of engagement where they end up looking in one direction at a conceptual screen, which breaks the immersion and stops them from experiencing the feeling of "being there" or spatial presence. Users get bored in this mode and eventually end up with a negative overall experience.

By comparing the results from the analysis with the previous studies presented in the theoretical framework section, we can see that some of the findings are supported by previous literature.

5.1 Role of Audio-Visual Cues in Perception and Immersion

In guided mode users experience several components of psychological flow state [5] where they feel immersed in the virtual environment, lose track of the surrounding real environment, feel in control and their focus get directed to follow the storyline of the narrative through available audio-visual cues. This results in enjoyment and a sense of spatial presence. This finding matches with Sutcliffe [6] in terms of flow experience.

The role of audio-visual cues in attracting and directing users' attention in a virtual environment agree with the theory of ecological perception [1], which states that we turn our heads to direct our attention to different visual stimuli and we focus to hear

better and gather information about action possibilities around us. The use of spatial audio to direct users' attention in different directions throughout a narrative matches with the findings of involuntary attention allocation in a virtual environment studied by Hendrix and Barfield [11].

5.2 Role of Audio-Visual Cues in Spatial Presence

When the users' experience was directed by appropriate audio-visual cues, multiple users expressed the experience of being spatially present. This fits the findings from existing literature [13, 14] that emphasize on the use of appropriate audio-visual spatial cues to increase the chance of users feeling spatial present in a virtual environment. It is also important to point out the use of directional audio cues in both of the narratives where users experienced immersion and spatial presence.

In "FOMO" mode, users ended up having a frustrating experience because they felt like they were not in control. They got stressed thinking that they might be missing something important in the storyline and too many things were going on at the same time, which in many cases broke their immersion. This agrees with the findings from Wirth et al. [12] who argue that a virtual environment will more likely be perceived as a plausible space if the used audio-visual cues have a logical consistency. In "FOMO" mode, the inconsistencies with the audio-visual cues confuse the users, which eventually block them from experiencing spatial presence in most cases. The negative experience of the users also matches the findings from de Rijk et al. [15] who argue an excessive use of spatial cues can cause sensory overload producing fatigue for the users.

6 Conclusions

Noticing the exponential rise of virtual reality applications in 2016, the goal of this study was to explore user experience in mobile virtual reality narratives and to investigate what role audio-visual cues play in users' positive or negative experiences. By using a consumer oriented mobile HMD and some popular VR applications, this study also examined whether the results relate to the findings from existing literature, where the research was conducted mostly in controlled environments using proprietary devices.

We know from the language of cinema that motion, color and contrast work really well as visual cues to direct audience's attention where needed. But when those audiences are placed inside the media in virtual reality, there is always a chance of having their back turned to important elements. To avoid this situation a designer can use visual cues inside the field of view of the users and audio cues outside the FOV to make the users turn their heads to face the elements important for the narrative. While only four narratives cannot be used to generalize the finding, it can be taken as a basis for further investigation into the effect of audio-visual cues on user experience in virtual reality applications.

The main findings of this study can be summarized as the following:

6.1 Audio-Visual Cues Make or Break an Experience

One of the most important findings of this study is how different audio-visual cues attract and direct user's attention throughout a VR experience. It is clear from the analysis of the recorded data that a virtual reality narrative needs to have well designed audio-visual cues to guide users' attention in a virtual environment to increase immersion that results in a positive overall experience. It is also important to keep in mind the usefulness of spatial audio cues that direct user's attention to elements of the VR experience not visible in user's field of view.

6.2 The Sweet Spot Lies Between Boredom and Frustration

Another important finding is how the amount of available audio-visual cues affects user experience in a virtual narrative. It is clear from the findings of this study that excessive audio-visual cues put the users in "FOMO" mode resulting in their frustration and a negative overall experience. While too many cues bring frustration, lack of cues brings boredom since the users expect to see events happening all around them in an immersive VR environment. Only a limited number of well-designed audio-visual cues hits the sweet spot and guides the audience throughout the narrative without requiring much effort from them.

6.3 It Takes a Little Extra Time to Get Used to Mobile VR

When it comes to lack of positional tracking in mobile VR, it is clear from the findings that users get confused and frustrated when their movements in the physical environment result in the whole virtual environment moving with them. This clash between real world perception and virtual world realization breaks their feeling of spatial presence in that VR environment. Fortunately, once they get used to the limited tracking capabilities of the HMD, users can easily get back into the flow of the narrative, especially when well-designed audio-visual cues guide them throughout the experience. Based on this finding it can be suggested that users should always be given a little extra time in the beginning to get used to their movements in a mobile virtual environment.

6.4 Limitations and Future Work

Due to the small sample size and usage of inductive methods, no claims can be made towards the generalizability of the findings. While the data collection and analysis were done thoroughly and carefully, the results need to be tested in a controlled study to verify the findings. In particular, it must be established in future studies if the same three modes of engagement would re-emerge with new users and new applications.

Another limitation is the age group of the participants. The findings might be different if the participants were from an older generation who are more hesitant towards new technology or if the participants were children who are more curious in a new

environment. There are also cognitive differences in perception among different age groups, which might affect the different modes of engagements proposed in this study.

Acknowledgments. The author would like to express his gratitude to Professor. Annika Waern for her invaluable guidance throughout the study.

References

1. Gibson, J.J.: *The Ecological Approach to Visual Perception: Classic Edition*. Psychology Press (2014)
2. Stein, S.: Everyone wanted a piece of virtual reality at this year's CES (2016). <https://www.cnet.com/news/everyone-wanted-a-piece-of-virtual-reality-at-this-years-ces/>
3. Child, B.: Steven Spielberg warns VR technology could be 'dangerous' for film-making (2016). <https://www.theguardian.com/film/2016/may/19/steven-spielberg-warns-vr-technology-dangerous-for-film-making>
4. Smith, S.: VR Headset Mega Guide: Features and Release Dates (2016). <http://www.tomsguide.com/us/vr-headset-guide,news-20644.html>
5. Csikszentmihalyi, M.: *Flow: The Psychology of Optimal Experience*. Harper Collins, New York (1990)
6. Sutcliffe, A., Gault, B., Shin, J.E.: Presence, memory and interaction in virtual environments. *Int. J. Hum Comput Stud.* **62**(3), 307–327 (2005)
7. Sherman, W.R., Craig, A.B.: *Understanding Virtual Reality: Interface, Application, and Design*. Elsevier (2002)
8. Gaver, W.W.: Technology affordances. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 79–84. ACM (1991)
9. Norman, D.A.: *The Design of Everyday Things: Revised and Expanded Edition*. Basic Books (2013)
10. Rasmussen, J., Rouse, W.B.: *Human Detection and Diagnosis of System Failures*, vol. 15. Springer, Heidelberg (2013)
11. Hendrix, C., Barfield, W.: The sense of presence within auditory virtual environments. *Presence: Teleoperators Virtual Environ.* **5**(3), 290–301 (1996)
12. Wirth, W., Hartmann, T., Böcking, S., Vorderer, P., Klimmt, C., Schramm, H., Saari, T., Laarni, J., Ravaja, N., Gouveia, F.R., Biocca, F.: A process model of the formation of spatial presence experiences. *Media Psychol.* **9**(3), 493–525 (2007)
13. Steuer, J.: Defining virtual reality: dimensions determining telepresence. *J. Commun.* **42**(4), 73–93 (1992)
14. Biocca, F.: The cyborg's dilemma: Progressive embodiment in virtual environments. *Hum. Factors Inf. Technol.* **13**, 113–144 (1999)
15. de Rijk, A.E., Schreurs, K.M., Bensing, J.M.: Complaints of fatigue: related to too much as well as too little external stimulation? *J. Behav. Med.* **22**(6), 549–573 (1999)
16. Merriam, S.B., Tisdell, E.J.: *Qualitative Research: A Guide to Design and Implementation*. John Wiley & Sons (2015)
17. Villeneuve, P., Csikszentmihalyi, M., Robinson, R.: The art of seeing: an interpretation of the aesthetic encounter. *J. Aesthetic Educ.* **27**, 120 (1993)
18. Minichiello, V., Aroni, R., Hays, T.: *In-depth Interviewing: Principles, Techniques, Analysis*. Pearson Education Australia (2008)

19. Miles, M.B., Huberman, A.M.: *Qualitative Data Analysis: An Expanded Sourcebook*, 2nd edn. Sage Publications, Thousand Oaks (1994)
20. Denzin, N.K., Lincoln, Y.S.: *The Sage Handbook of Qualitative Research*, pp. 1–20. Sage Publications, Thousand Oaks (2011)
21. Braun, V., Clarke, V.: Using thematic analysis in psychology. *Qual. Res. Psychol.* **3**(2), 77–101 (2006)
22. Sarker, B.: *Show me the sign!: The role of audio-visual cues in user experience of mobile virtual reality narratives*. Master's thesis. Uppsala University, Uppsala (2016). <http://uu.diva-portal.org/smash/record.jsf?pid=diva2:1044065>