Master Sound Design

Rooms, they speak (?)

Immersive storytelling with sound
(a foray into spatial sonic narratives)

ZHdK

Patrycja Pakiela

Zürich, Switzerland

2022.01.13

Mentor: Dr. phil. Daniel Hug

Table of Contents

Abstract

Telling stories is something akin to humankind. It is perhaps the single best descriptor for our species. Our culture, our systems and infrastructure and all its underlying technology, have amplified both the impact and the range of our stories by many orders of magnitude. We are drawn to recognize patterns, and find heuristics connected by a thread that can guide us through seemingly chaotic and random events. In this study, the use of sound spatialization and acoustic narrative techniques will be analyzed as a method for supporting and enhancing the immersive quality of sound-design and music-design in film. The 360 Video -Virtual Reality (360-VR) documentary film "Caves", by director Carlos Isabel García, was used to develop an auditive spatial storytelling approach. The idea of creating the soundscape for a documentary movie in VR, where the story takes place inside a cave system, felt like a technically complex but creative project, and the prospect of the evocative and mysterious setting, and the challenge to find a suitable language for that special character, was an idea I simply could not resist.

After providing a background of foundational literature, exploring the topic of immersion, documentary-style film and VR & Spatialization technology, I define a framework to question the design decisions and approaches taken for the project. I also provide an overview of the project, detailing the location, the story and outlining relevant information regarding the team, the process, and my workflow. Finally, a Scene Analysis dissects the selected "rooms", delving into a *Narrative* and *Spatial Analysis.* Analyzing a work in retrospective provides the advantage of 'foresight', but the main goal of this work is to challenge certain preconceptions about VR technology, and to hopefully discover some counter-intuitive possibilities while exploring further into the realm of perception and psyche in a fascinating underground world.

Acknowledgments

I would like to thank all the people who have helped me through the process of finding a way of telling sonic stories in spaces.

Many thanks to all the teachers at ZHdK, who introduced me to, and allowed me to explore different sonic worlds, leading me to the discovery of a fascinating field. I am grateful for their kind support, their guidance, and many interesting discussions. I would like to express my gratitude to Dr. phil. Daniel Hug for his patient mentoring and constructive feedback throughout my whole journey. A note of gratitude also goes for Dr. phil. Miriam Loertscher for the interesting conversation about narrative storytelling and documentary films and the abundant relevant recommendations for literature.

Thank you to Carlos Isabel Garcia for a great collaboration and interesting discussions while developing the sonic world of caves.

I am very thankful to Gregg Skerman, who shared with me multiple approaches on building sonic narratives for cinema, as well as showing an inspiring respect and openness while approaching every film-mix like an undiscovered cave which seeks for a way to express itself.

A big thank you to Matthias Knuser and Raumgleiter AG for letting me explore real virtuality with a great team, where storytelling and exploration play a key role in each project.

Lastly, a giant thankyou to Mauricio Rodríguez, who with his with analytical mindset and constructive criticism and open ears, gave me great feedback and support during the process of working on the project as well as the thesis.

1. Introduction

> *"As for me, I am tormented with an everlasting itch for things remote. I love to sail forbidden seas, and land on barbarous coasts." – Herman Melville*

When I started my journey as a sound designer, I found that sound elements, strung to each other and composed into layers, began evoking and suggesting meaning and could eventually tell a sort of story accompanying or challenging its visual counterpart. The discovery of this storytelling aspect of sound-design, was for me fascinating and it has become part of how I interpret, analyze, and conceptualize projects ever since. Diving into a world created out of the fragile intersection of multiple auditive parts, rendering an actual soundscape, feels close to my musical experience, and yet it is much more than that.

Ushering the audience into each project's narrative dimension, though ephemeral, was for me always an interesting "side effect" in my work as a musician. Intuitively, I was long ago already willing and curious to explore those topics of storytelling and immersiveness. Partly, it was this curiosity which brought me to study sound design and get in contact with "Immersive Technologies", particularly with Virtual Reality. In parallel with my studies, I have started working in a company which creates VR 3D solutions / scenes for architectural and real-estate projects. The possibility of finding a medium and a market for applied sound design has opened even further avenues for the development of narrative strategies with audio (audio-branding, functional-audio, atmospheric soundscapes, etc.).

Telling stories is something akin to humankind. It is perhaps the single best descriptor for our species. Our culture, our systems and infrastructure and all its underlying technology, have amplified both the impact and the range of our stories by many orders of magnitude. In my opinion, we are drawn to recognize patterns, and find heuristics connected by a thread that can guide us through seemingly chaotic and random events. We need stories to reassure us, and to make things accessible for others. I believe, stories are our medium. And their strategic use, beginning with persuasion, is becoming increasingly important in their role of identifying and communicating purpose and motivation.

Within the context of filmmaking, storytelling has evolved to incorporate a variety of elements, fusing our audio-visual perception with narratives that appeal to our memories, our imagination, and our intellect to convey an increasingly realistic and immersive sense of presence or experience. Moreover, some of these techniques become instrumental in creating a plausible context or an envelope, that sets the frame for stories and characters to develop and engage the audience. Sound has added an additional dimensionality to storytelling, but there are many approaches varying its importance, its strategy, and its character.

In this study, the use of sound spatialization and narrative soundscape techniques will be analyzed as a method for supporting and enhancing the immersive quality of sound-design and music in film. Concretely, the 360 Video - Virtual Reality (360-VR) documentary film "Caves", by director Carlos Isabel Garcia, was used to develop an auditive storytelling approach. The documentary will serve as a use-case to exemplify and explore the design decisions, the technology and the challenges and learnings resulting from the project, including its "screening". I will refer to the project interchangeably as a documentary-style film, since several of the scenes during production had to be prepared to guarantee safety and suitable lighting and recording conditions to achieve the desired aesthetic, mood and setting for the narrative. Spatial storytelling challenges classic linear narratives, introducing *dynamic attention focus*, *spatial awareness* and *context or localization*, but it also opens the doors to new forms of auditive enhanced narratives, supporting place-making, orientation, sense of scale and speed, and the simulation or reference to other spatial-temporal phenomena. Likewise, the narratorial capacity of sound in its ability to communicate emotion, mood and create references is something I will be exploring via applied examples.

I will provide a theoretical background as a foundation to the strategies developed for the acoustic design, and through a selection of the most representative scenes in the film, I will apply a conceptual

framework to analyze the elements used to affect the cognitive and emotional immersiveness via spatial and narrative sound. My goal is to compare the impact of two sound-design approaches for documentary-style film in the audience's immersiveness: a purely objective, high-fidelity description of space and close mapping to the story, and a more exploratory, detached, and subjective representation, loosely coupled with the actions. Furthermore, I am interested in exploring the interplay between these two strategies and in combining elements of both to achieve a tailored experience. In investigating these relationships, I intend to address the dualisms inherent in the medium of the selected use-case project: realism and authenticity vs. storytelling and design.

2. Background

2.1. Immersiveness

The term immersiveness in relevant literature, appears with somewhat inconsistent meanings. Often, it is used interchangeably with intentional concepts, or terms describing a motivation, such as: involvement, psychic presence, and engagement (Lombard & Ditton, 1997), while also allowing for more physical definitions of containment, bodily presence, or circumscription. The term is predominantly used in domains related to media such as video game studies, film studies, music studies and research dealing with linear and interactive works of literary fiction (Nilsson, Nordahl, & Serafin, 2016). Additionally, and particularly for the notion of presence, other academic disciplines have taken an interest and the topic is being explored in the context of communication, cognitive science, computer science, engineering, philosophy, and psychology (Nilsson, Nordahl, & Serafin, 2016). For this thesis, I will focus on immersion specifically in the context of Virtual Reality, Game Studies and Film, as these domains are overlapping in the selected case-study of the Project "Caves" and serve as anchors for the relevant issues and discussions in the topic.

> ***Immersion:*** *(Lat. immersio, dipping), the act of being plunged into a fluid, or being overwhelmed by anything; in astronomy, the disappearance of a heavenly body in the shadow of another, especially of a satellite in the shadow of its primary* - (Encyclopædia Britannica, 11th Edition, 1911).

For me, immersion prescribes a state of isolation and cancellation of one context, to achieve a sense of awareness, focus and presence in another context. One could argue that such a state of immersion can be reached by any sufficiently powerful and engaging medium, like music, literature, etc. Some mediums, like film, are predisposed or designed for offering a higher degree of immersive-ness by activating two or more of the most predominant human senses for which our bodies have been tuned for in modern times. Some newer technologies enable an even stronger immersive experience by engaging more of our senses in a seemingly natural way, and yet they are being are mapped onto an artificial milieu, which can be enhanced and extended to suit different needs and purposes.

In this context, two interpretations are predominant for immersive-ness. Firstly, the interpretation of immersion as a "psychological state of mind", describing the experience of a subject or user, and secondly, the interpretation of "immersive technology", describing the instruments and infrastructure which enables the state (Slater, 2003). According to Slater the technologies which deliver the level of development allowing the preservation of fidelity in relation to equivalent real-world modalities of motion and positioning are part of what can be described as immersive technologies.

As relevant as this context or plane of action created and sustained by immersiveness may seem, it would be in many ways devoid of meaning without producing a reason, an explanation, or simply put, a story

that takes place in it. In my opinion, immersiveness is an enhanced theater stage for storytelling. We can, in this context, understand immersion as one powerful supportive tool for narrative. The interdependence of **storytelling**, **sound** and **immersiveness** sets a foundation for an alternative to the standard visual-predominant narrative. This alternative has the potential to create highly engaging and rich experiences, as well as to contribute to the *"corrective efforts in dealing with the crisis of sensory culture in modern life"* (Fu, 2021). The current study aims to explore the concrete application of this narratological triad.

2.1.1. Immersive Technology

> *"Immersive technologies create distinct experiences by merging the physical world with a digital or simulated reality."* - (Bolter & Grusin, 1999)

Immersive technologies have a ranging impact and scale. They also vary in target, platform, and application. At their core, they must be generally comprised of software and hardware linking one or more users with digital content. There is therefore a significant overlap in the origins of immersive technologies with human-computer interaction (HCI) and its underlying principles. In essence, immersive technologies are interfaces between humans and digital experiences, which have managed to harness natural or close to natural movements and interactions to replace or minimize the need for interactive artifacts (such as a keyboard, a mouse, or a game controller). From a perspective of HCI, immersive technologies strive to create "imperceptible" interfaces.

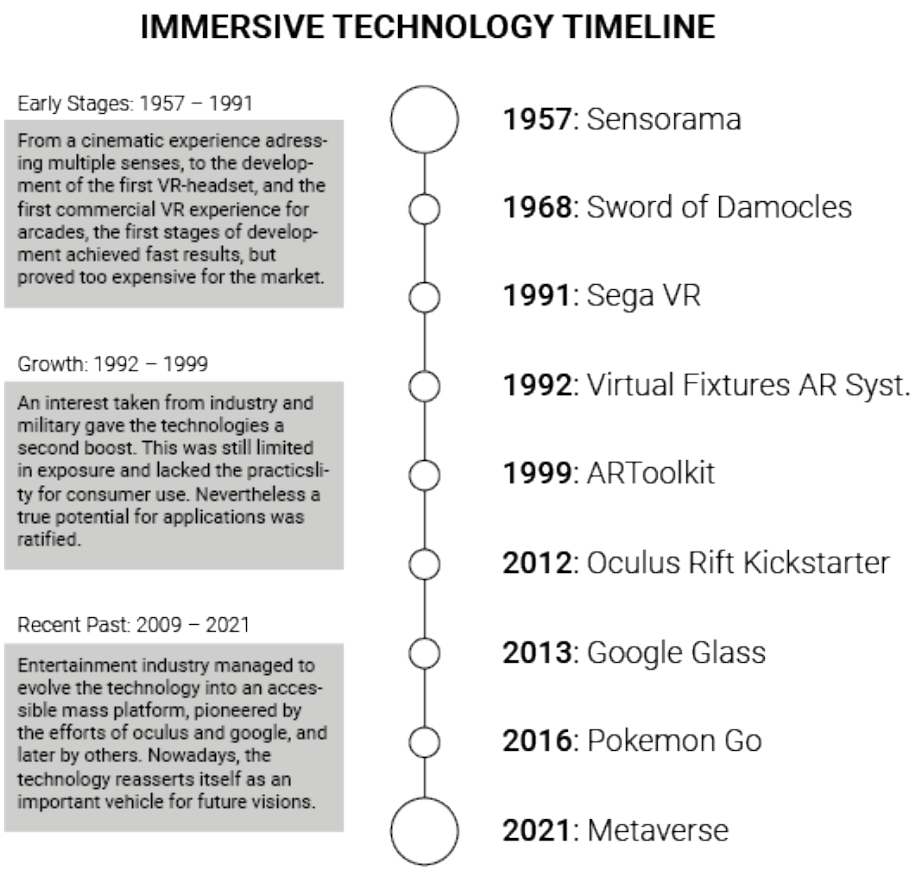


Figure 1, Source: Author, Data source: https://www.vistaequitypartners.com/insights/an-introduction-to-immersive-technologies/

Among the most prevalent immersive technologies today, is Virtual Reality (VR). Virtual Reality is a technology which, by an interplay of hardware (VR Headset, Infrared sensors) and Software (position head-tracking, mapping, and rendering), audiences are allowed to dive into a designed digital world and can participate and interact with that world. An evolution of the concepts driving the creation of VR and the software and hardware that enables it, has also yielded extended technologies such as Augmented Reality (AR) or Mixed Reality (MR) in which the actual physical context is actively used in combination with digital overlays to create a composed reality. Although the core technologies have been existing since the 1960's (*Figure 1.*) hardware costs and complexity have limited its wider adoption and use, however, the journey towards affordability and usability has been radically accelerating in the past few years.

Because of the specific technical (hardware and software) requirements for VR, AR, and MR, they are seldomly considered in isolation as a technology which can be applied onto other frameworks, rather they collectively define a platform and increasingly an ecosystem of technologies, tools, and content-creators. Some of the technologies that are central components of these platforms, have been borrowed and are applied to yet other uses. The 360-VR-Video is one good example of a common use of VR technology, in the creation of an in-between type of experience or medium.

This technology is essentially a recording of a full spherical visual spectrum from a particular point in space. While being reproduced, the audience has a view from the position of the camera used for recording but sees only a cropped "window" of the entire media at a given time. Software allows for interaction to control the active "window" (its direction). 360 Videos are often mistakenly called VR Films, simply because they can be experienced with VR technology (head tracking, vr-headset, and stereoscopic projection). However, the content of a 360 Video, can actually be reproduced by a variety of devices such as a smart phone or a computer and can therefore be interacted with in a variety of ways.

Our use-case is a documentary film that is designed to be experienced with VR hardware (360-VR-Video). We will therefore focus on this combination of technologies for the purposes of this study. To explore the implications for design decisions or approaches to storytelling, a clarification of the limitations and potentials of the selected technology is due. As mentioned before, a 360-VR-Video borrows hardware from traditional VR to gain an "imperceptible" or more intuitive interaction, with the purpose and effect of achieving a higher degree of immersiveness. Some core characteristics that can elucidate the differences and similarities between "pure" VR and 360-VR-Video are shown in *Figure 2*.

	VIRTUAL REALITY	360 VIDEO
FIELD OF VIEW	UNCONSTRAINED, INTERACTIVE VIA HEADSET	UNCONSTRAINED, INTERACTIVE VIA HEADSET
MOBILITY	IMMERSIVE WORLD WHICH IS TYPICALLY FREELY EXPLORABLE VIA PHYSICAL DISPLACEMENT OR TELEPORTING	360 DEGREE VIEW FROM CAMERA'S PERSPECTIVE, BUT LIMITED TO RECORDED / RENDERED POSITION
TIMELINE	NON-LINEAR: EVENTS PROGRESS EITHER INDEPENDENTLY (AGENCY) OR DEPENDENT AND REACTIVE TO THE USER EXPERIENCE	LINEAR: EVENS TRANSCURR ON A CURATED TIMELINE THAT IS FIXED
PLATFORMS	EXPERIENCE REQUIRES VR HARDWARE AND SOFTWARE (EITHER TETHERED OR UNTETHERED HEADSET INFLUENCES MOBILITY)	CAN BE REPRODUCED IN VR TECHNOLOGY, BUT ALSO FOR MOST TRADITIONAL VIDEO PLAYERS

source: https://filmora.wondershare.com/virtual-reality/difference-between-360-video-vr.html

Figure 2, Summary of Difference between technologies

Perhaps the biggest difference lies within the mobility rubric of the comparison, which is connected to what is defined as Degrees of Freedom (DOF) in digital environments. This concept often discussed in the development and analysis of Immersive Technologies, describes the possible axes of interactions within a three-dimensional space (Barnard, 2019). The first three degrees of freedom describe rotational movement on axes "x", "y" & "z", and the next three describe movement or displacement along those axes (forward/backwards, right/left, up/down).

In the case of 360-VR-Video content, we are allowed to choose where we look at (otherwise referred to as 'gaze' or 'field of view'), but we cannot manipulate our position nor our displacement within the virtual space, therefore unlike VR, the viewer cannot walk around in the virtual environment (VE), nor can they modify objects and characters within (Cummings & Bailenson, 2016). Without this mobility, the degrees of freedom, so characteristically powerful in creating a sense of presence in VR, are significantly reduced. Nevertheless, 360-VR-Film still conveys some of the "flair" of immersive technology. Viewing a film on a VR headset allows the viewer to experience it in a particular way by turning the viewer into "an active observer (compared to traditional video), or a passive participant compared to virtual reality (V*R)"* (Cummings & Bailenson, 2016).

Cummings and Bailenson (2016, p.274) further discuss the effectiveness of immersive technologies, singling out the following characteristics as primary for the enhancement of the immersive experience:

a) high fidelity simulations through multiple sensory modalities
b) fine mapping of user's virtual bodily actions to their physical body's counterparts
c) removal of the user from the external world through self-contained plots and narratives

Although the second recommendation of Cummings and Bailenson is out of scope for our selected technology, a potential remains for the role of sound or sonic narrative to likewise enhance and utilize the immersive-ness of the experience. With its basic, yet effective immersive qualities, 360-VR-Video has challenged filmmakers to look for new ways to approach the narrative aspect of movies. Thanks to this technology, the spatial medium is emerging as a new milieu, where new challenges and opportunities for sound design, music, and in general soundscapes, arise. This also is altering the existing collaboration processes, by allowing for more parallelized and holistic involvement of sound professionals in the project, ultimately achieving a more significant contribution to the story. I will be addressing some of these technological implications on the workflow, the collaboration process, and the design for our Use-Case project.

2.1.2. Psychological State of Mind

> *"Immersion is a metaphorical term derived from the physical experience of being submerged in water. We seek the same feeling from a psychologically immersive experience that we do from a plunge in the ocean or swimming pool: the sensation of being surrounded by a completely other reality, as different as water is from air, that takes over all of our attention, our whole perceptual apparatus."* - (Murray, 2017)

In my opinion, the technical background established in the previous segment, can be understood as a pre-condition or at least as a facilitating scaffolding, to the achievement of an immersive psychological state of mind. As Slater (2003) writes, achieving psychological immersion, is greatly affected, or enhanced by multisensory technological systems. The technology which has been developing since the 1960's in the form of Virtual Reality (VR) is increasingly enabling, the stimulation of user senses in an almost seamless integration with organic and intuitive body movements, and allowing that experience of immersion, of "being there" to happen. The main purpose of VR is to become an invisible or imperceptible medium: "virtual reality can disappear as an interface and give the viewer the same emotions that she would feel in the real world" (Bolter & Grusin, 1999).

According to (Witmer & Singer, 1998), *Immersion* is a feeling of being enveloped by, included in, and interacting with the virtual environment. There are two distinct aspects of that psychological state: "perceptual immersion", or the perception of being enveloped in the environment, and "involvement", or sensitivity to stimuli, focusing on inputs from the environment, the spaces and other narratorial elements (Witmer & Singer, 1998). Perceptual immersion is dutifully reinforced by what Bolter and Grusin (1999, p.14) refer to as "transparent immediacy" or the imperceptibility of media. The authors state that the awareness of the medium is intermittent, and that new media therefore tend to *"oscillate between immediacy and hypermediacy, between transparency and opacity*" (Bolter & Grusin, 1999). An absolute commitment of the audience or an unwavering belief of the representation is neither required nor aspired to, however, a minimal "contact point between the medium and what it represents" is needed to facilitate "involvement".

Virtual reality, as described in the previous section, enables a sense of presence by "fooling" the senses. A decisive vehicle for conveying presence which is simultaneously less immersive, is the "*responsive character of the environment*" (Bolter & Grusin, 1999) in VR. This responsiveness is a liminal characteristic in that it sets attention to the media by calling the audience to action (involvement) and switching from a pure spectator mode to an interactive one. Yet this physical action is transferred (ideally in a seamless synchronous fashion) into virtualized action, again shifting the media from a foreground to give way to focus on the content being represented and updated virtually. This cyclic process is referred to as "remediation" by Bolter and Grusin (1999).

The entirety of the physiological and psychological experience of virtual reality or interactive videos is composed of an interplay between intricate and complex mechanisms involving multiple if not all sensory inputs into a "unified perceptual world" that is mapped onto "surrounding environment dimensions, such as space and time" (Väljamae & Tajadura, 2007). According to Väljamae & Tajadura (2007), within this constellation of sensory inputs, sound has a dominant role in the construction of a credible "real" representation:

> *"Real world events are typically represented in different sensory modalities and sound usually plays a dominant role in the perception of the temporal structure of these events. For example, sonic realism was found significantly more important than the visual realism when representing object dynamics of multimodal virtual environments"* (Väljamae & Tajadura, 2007).

The relevance of sound in the realistic representation of "motion dynamics and embodiment" (Väljamae & Tajadura, 2007) and its capacity to improve immersive-ness are instrumental to the analysis and design decisions applied to the case-study project of "Caves". This central property of sound will be further discussed in the following section in terms of narrative and its potential applications for storytelling.

2.2 Storytelling

> *"(Interactive) Storytelling works by spatially engaging a user inside a mediated environment whose discovery through exploration advances a non-linear narrative, and where space is the essential communication medium. Inspired by 'immersive theater', it closely follows environmental storytelling in creating preconditions for immersive narrative experiences in four possible ways:*
>
> 1. *Ability of spatial stories to evoke pre-existing narrative associations*
> 2. *Provide a backdrop where narrative events unfold*
> 3. *Embed narrative information within their mise-en-scene*
> 4. *Provide resources for emergent narratives*
>
> *"* (Hameed & Perkis, 2018)

As mentioned before, the project "Caves" is a 360 VR Film which allows viewers to interact by being able to freely orient and focus their attention to content in 360 degrees (or more precisely in an almost complete spherical projection field). Given this controlled freedom, it is relevant to consider the experience of *immersive narration* in video games as a reference, since the concepts of "sound spatialization" and "narrative elements" can become useful to further understand the psychological and physiological underpinnings of the technology that enables it and the possibilities it opens for narrative purposes as suggested by the previous chapter.

Immersive narrative can be understood as the feeling of being completely absorbed by a fictive world, and nevertheless accepting the events and the story as real (Nilsson, Nordahl, & Serafin, 2016). According to the authors, the essential requirements to experience narrative immersion are an exhilarating plot, interesting characters, and dramatic situations. Ryan (2001), who describes immersion from the reader's perspective, states that any sort of narrative work can lead to narrative immersion; this immersion should be portrayed evenly by deep psychological involvement with the story, the space where the story takes place and its characters. The author divides narrative immersion into three categories: *temporal, spatial and emotional* immersion (Ryan, 2001).

- *Temporal immersion*: refers to the involvement of the reader in the story so that the reader desires to know how the story will evolve.
- *Spatial immersion*: is about involvement of the reader with the virtual, designed space. Immersion through reaction, feeling of and being in that space and being there while the events in the that space occur
- *Emotional immersion:* occurs when the reader gets emotionally involved in favor of the narrative protagonist or antagonist. (Ryan, 2003).

This constellation of space, time and emotion clearly represents a paramount contribution to the interpretation of a story and the involvement of the audience. It will also form part of the methodological framework I will use for studying the "rooms" and understanding the potential contributions of acoustic narrative elements in the Scene Analysis Section.

The influence of the narratorial structure and elements, represent a scaffolding and a valuable reference for conceptualizing the application of sound-design and music, providing anchor points for the story, the characters, and the space and either supporting or adding to the main story trail. It is also noteworthy, that narrative decisions follow diverse principles which demand different responses in the approach to sound-design and the corresponding acoustic narrative. For this purpose, we will dive into the specifics of documentary-style film, as it will help us set the boundaries and "rules" for the particular type of storytelling.

2.2.1. Suspension of disbelief in documentary-style film

Documentary- style films face an inherent challenge, as do any other non-fictional media, in surmounting a wall of "believability" or trust, and simultaneously bridging this with a voice and a vision which are inevitably nevertheless curated. A balance must be struck, between the cognitive and emotional involvement (Ji & Raney, 2016), one that facilitates a psychological state of immersive-ness. This gravitational pull from both the demand in realism and the emotional engagement is often defined in the interplay between Suspension of Disbelief (SOD) and Narrative / External Realism (Ji & Raney, 2016).

Ji & Raney (2006) have developed a model of reception for non-fictional films. In this model, the aspects of mediated, fictional, and real narrative are combined into a hierarchy that enables the reception of entertainment content. The study of the impact of the audience's involvement through the combination

of SOD and realism in documentary-style movies, and the established model will be used as a reference for the development of a method in my analysis on how to make design decisions.

Suspension of Disbelief describes the viewers' willingness to perceive and accept a presented story. Independently of whether the story is fictional or not in nature, SOD describes the tolerance to the presented world as a believable setting for the development of events, leading towards an emotional involvement; or towards the dissolution of "*a narrative's lack of realism or consistency"* (Ji & Raney, 2016).

This notion is utilized often to clarify why audiences *"care about people, things, and events that are clearly not real."* (Ji & Raney, 2016). It is believed by some scholars, that our 'default' receptive mode is typically positively geared towards SOD, and that there is an "*underlying assumption (…) that viewers actively stop evaluating the authenticity of fictional content once they decide to maximize enjoyment"* (Ji & Raney, 2016). I believe there are some intrinsic reasons for this behavior which can be found in Gestalt Theory or in the way our brains work. We are constantly processing streams of diverse signals and trying to derive sense out of them, distilling knowledge from mingled memories of experiences, and rapid heuristics which attempt to map the signals to known patterns. We are hard-wired to "believe", to "recognize", to "empathize" with other beings, with entities and stories because, once upon a time, our survival depended on our ability to detect patterns (especially dangerous ones).

Ji & Raney (2006) mention however, that studies have found fundamentally different audience standpoints between stories which are known to the audience to be representing reality vs fictional stories. Meaning that the expectation for realism and therefore for a certain degree of logic, rationality, and consistency in all aspects of the film is set up to match the claim of authenticity. Likewise, it is to be expected that a greater leniency will be granted from audiences adapting their expectations to a known fictional world:

> *"The researchers contend that viewers detect cues that indicate whether a film is raw reality or not. When it is, logical and rational processing (as well as reality testing) are employed; in turn, audiences are likely to have relatively fewer emotional reactions. However, if cues indicate that the action is scripted and acted, then viewers can accept the implicit invitation to play a "game of pretend" and to perceive what is on screen as a plausible reality for the purpose of being entertained. Emotional reactions will be relatively more salient in this case"* (Ji & Raney, 2016)

This means there is some degree of volatility in the balancing of the audience's standpoint and suggests that there are multiple factors which can and do influence the perception of reality and in turn the ease with which an audience will "dive" into the film. Among many possible factors, different audience groups might per-se represent varying pre-conditions (age, origin, or domain knowledge) which would favor a positive or negative disposition to SOD. In any case, this seems to be also not an immutable setting, as the evolution of a particular film's story or content, could surely also modify a predisposed expectation, and even reverse it.

A common example of this (and close to my experience) is with films portraying musicians performing. In such films, the actions are often shown by alternating between close-ups of a trained performer and more distant shots performed by an actor (normally musically untrained), which can lead to a sudden lack of believability for me based on my musical background. From the perspective of sound, this can also be the case when the acoustics of the visible room do not closely match the recorded performance. Again, this might not necessarily represent a "violation" of reality for a musically untrained audience.

Ji & Raney discuss these limits even for an audience in which SOD is by standard tolerated, and therefore considered more lenient within the context of non-fiction:

> *"(…) such tolerance reaching its limit once violations of realism grow overly excessive. The researchers found that in the nonfictional media context, lack of realism—operationalized as narrative exaggeration—was associated with low SOD among viewers seeking to acquire*

> *knowledge from a film about a political candidate. However, lack of realism did not affect SOD among viewers seeking pleasure and amusement."* (Ji & Raney, 2016)

In the space of documentary films, we see that this fragile inclination of the audience's interpretation, is all more crucial, for the stakes of "realism" and "believability" are higher and can potentially affect the effectiveness of conveying a message and transporting information and facts. This interplay, is however, not only subtle but also more complex. The intentionality, style, flavor, and all specifics of the film will come also into play in achieving this balance. In the next section I will explore this topic further and provide some insights into existing models that put forth explanations and hint towards a strategy for this "juggling of realism".

2.2.2 Realism and Diegesis

> *"**Realism**, in the arts, the accurate, detailed, unembellished depiction of nature or of contemporary life. Realism rejects imaginative idealization in favour of a close observation of outward appearances."* – (Encyclopaedia, 2020)

In the previous section we discussed the decisive importance of realism or a believable version of reality for the effectiveness of a story, and how the lack of realism can easily break the "magic spell" in situations where the audience has expectations for high fidelity or accuracy. We also touched upon the dual nature that constitutes the perceived reality. Ji & Raney (2016) refer to this duality as ***external*** and ***narrative*** realism:

> "*In fictional* contexts*, **external** realism refers to the extent to which events and persons in a story are considered consistent with the actual world (e.g., with no prior explanation, a scene featuring a character leaping into the air and flying to work would be considered low in external realism), while **narrative** realism is defined as the degree to which events and persons are consistent and logical with respect to the narrative settings."* (Ji & Raney, 2016)

Thomas Görne (2017) offers an alternative description of these complementing realism facets, namely ***physical*** and ***perceptual***. The physical feature has the goal of "(…) *construct(ing) a detailed complex scene, where the audience can freely move the focus of attention. The goal of this approach is the **physical** realism"*, and the perceptual feature, aims for "*guiding the audience's focus of attention through a complex scene by highlighting the important (auditory) objects*" (Görne, 2017). There are several parallels between these two systems, but an underlying, more primal overlap could be established between a ***diegetic*** vs. a ***non-diegetic*** description of reality.

Still the question remains of what are the elements, layers and compositions that constitute this depicted reality. What are also the thresholds of these constructed realities, and what is the impact of their "realness" for their audiences? The idea that "*realism can be addressed with multiple elements such as visual and stylistic conventions, methods, and mechanics of storytelling, rhetorical rules, and commitment to the facts*" has been proposed by Ji & Raney (2006, p.129). The dimensionality of the depicted realism can therefore also positively affect the SOD, as it emulates a higher "resolution" of reality. Research in the area of entertainment psychology, has found that the perception of a rendered or presented reality can have "*emotional, cognitive, and behavioral impacts"* (Ji & Raney, 2016) on audiences. This indication, that realism contributes to an effective emotional and cognitive content perception, implicates that a particular "rendering" of reality can also be forged to coerce or exert specific and designed effects on an audience. In contrast, *"when content is judged to be unrealistic, viewers' transportation and identification will be inhibited, which leads to negative effects on*

enjoyment." (Ji & Raney, 2016). I will analyze this dynamic further in the sound-design examples of the use-case.

As we were discussing before, involvement and immersion, play a key role in the perception of the storytelling. Ji & Raney's paper states that: *"(...) it seems reasonable to expect that narrative realism might be positively associated with* ***emotional*** *and* ***cognitive*** *involvement, especially within a documentary-style fictional film context."* (Ji & Raney, 2016). On the emotional impact of sound on movies and the perception, Thomas Görne mentions the dependency on aspects further than the exclusive physical accuracy: "*As we consciously perceive only a small part of a complex scene, the key to realism is not the sheer amount of detail. Instead, it is most important that the essential elements of the scene are convincing and persuasive."* (Görne, 2017)

Contributing to this persuasion is the concept of diegetic and non-diegetic sound in film theory:

> *"The term diegesis refers to the relation of image or sound to the virtual world of the film. (…). A diegetic sound belongs to the world of the film, it exists as an acoustic signal in the virtual world. A non-diegetic sound can either be metadiegetic, i.e., subjective (e.g., an inner voice in the character's head, or an "semantically overloaded" or by other means alienized sound mimicking a character's perception), or it can be extradiegetic, i.e., completely outside the world of the film, audible only for the audience, not for the characters (e.g., typical film music or typical fancy "sound effects")."* (Görne, 2017)

In essence, the fluctuation and transitioning between diegetic and non-diegetic elements can be used to support, accentuate, and enhance aspects of the story, or to guarantee consistency across a style, thus varying and depending greatly on content and context. Görne writes that sound design which is based on entirely diegetic sounds can be called "*documentaristic"* or "*naturalistic"*, and it should be coherent with the world shown in the picture, additionally, its likely to be perceived as objective at the level of information: "*Its informational content – the information rate – is low due to redundancy of image and sound"* (Görne, 2017). On the contrary, a soundscape which is solely designed from meta or extra-diegetic sounds could acquire a s*urreal* or *mystical* character*:*

> *"(...) it presents mainly information that is not communicated within the image, the redundancy is low, and the information rate is high, thus it is challenging and probably stressful to perceive, and it is likely to be perceived as "subjective".* Görne (2017, p.25).

Görne suggests two additional levels of perception of sound-design, between the *naturalistic* and the *surrealistic* ones. The first is "*attention guiding"* sound-design, which is constructed from meaningful symbolic or metaphoric sounds, following the intention of perceptual sound design (to grasp the attention of the audience). *Supernatural* sound-design appeals to an even further mystical (or transcendental) plane, which can be achieved or activated with the use of sounds with dominant extradiegetic sounds. This is an alternative soundscape which might come into play, where the unknown non-diegetic sounds are a base for it. Görne finds a relationship between the level of information and diegesis in these four levels of intensity of sound-design:

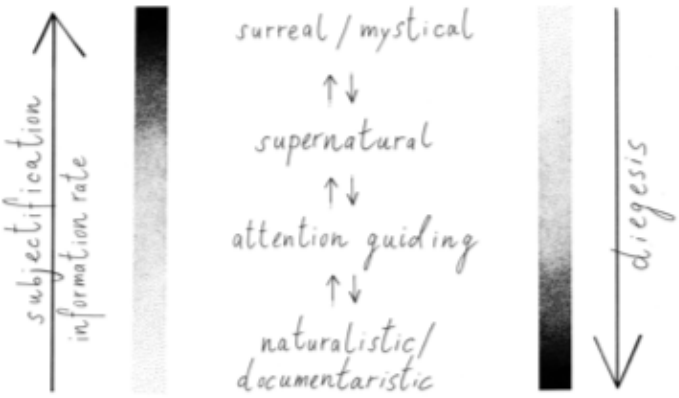


Figure 3: Görne: Four levels of sound design in dependence of diegesis (Bottom: diegetic, top: non- diegetic), subjectification (Bottom: objective, top: subjective) and information rate (bottom: low, top: high).

As mentioned before, the larger the number of diegetic sound elements present and the further down they are in the scale shown above, the more objective character is bestowed on the content and the closer it resembles the physical world. The less diegetic sound elements and qualities of the whole story, the more it is perceived as subjective and contains more meta-information. He states that in the transition from the lower level to a higher level, a suspense is being created, from naturalist sound design into supernatural, surreal, and mystical. We can assume that, upon descending the scale or transitioning to the lower levels, the displacement functions as an anchoring of the story, or a grounding of facts which can act as reference points for a plausible narrative. These references are often linking the visual elements to their real-world rules and to expected physical and spatial qualities.

A related interesting mechanism, aiding in the ability to orient oneself in a new complex or unknown environment, is called spatial-temporal congruency. This mechanism helps us connect synchronous auditive and visual signals to orientate in unknown spaces. An example of the application of this mechanism in film is in the use of “Foleys”, whose goal is to closely follow or mimic the likely and expected behavior of the physical world (however, they are often still presented to the audience at an exaggerated level or “closeness”). Evolutionarily, such mechanisms could have been part of a more primal set of skills to help an individual to orientate quickly in an unknown complex environment. They are therefore also embedded in our psyche and can be targeted and exploited for the sake of a narrative.

By exploring some of these topics of perceptual, psychological, and emotional nature, we have laid some foundational concepts and relations to develop a framework upon which to guide design decisions. The following section will describe the framework employed in the analysis of the use-case.

3. Case-Study: “Caves”

Scene from Caves @Carlos Isabel Garcia

The process of working on “Caves”, has been filled with lots of learnings and challenges. It has allowed me to establish a new workflow for my future 360 Projects, learn about new tools and explore the frontiers of storytelling with sound.

The idea of creating the sound design for a documentary movie in VR, where the story takes place in caves down under the earth, felt like a challenging and an undoubtfully creative project. Additionally, I have been asked to take care of the musical part as well, which given the prospect of the evocative and mysterious setting, and the challenge to find a suitable language for that special character, was an idea I simply could not resist.

In the Introduction, I have mentioned how working in VR Environments isn`t something new to me. Creating a soundscape for a linear 360 VR Film, however, was a new challenge for me. Most of the projects I have done in the context of my current employer (Raumgleiter AG, www.raumgleiter.com), were highly interactive, involving much more degrees of freedom, but within a more restrained expression.

Because of the high requirement of flexibility and interactivity (among other technical advantages), the workflow which I have developed for these past projects is based on Game-Engines and scripted Sound Events. However, due to the different platform selection for this project, a direct knowledge and process transfer was not feasible. The exercise of getting the two worlds of linear filmmaking and VR technology to work in tandem was very interesting and exciting. As a part of this work, I have conducted an interview with the director, to which I will be referring punctually as I explain the reasoning and the motivation behind some of the decisions taken. The complete Interview is in the Appendix.

3.1. Framework

Generalizing the models discussed in the previous sections, the analysis of selected scenes in the film “Caves” will be based on a simplified framework that considers the *Emotional Involvement* and the *Cognitive Involvement* as factors contributing to Immersive-ness. This framework is primarily based on the “Model of Entertainment Reception for Documentary-Style-Films” by (Ji & Raney, 2016), but it replaces the notion of *Suspension of Disbelief* and the categorical distinctions of realism (*Internal / External*), with the more fluid scale of *Subjectivity* and *Objectivity* of Görne (2017).

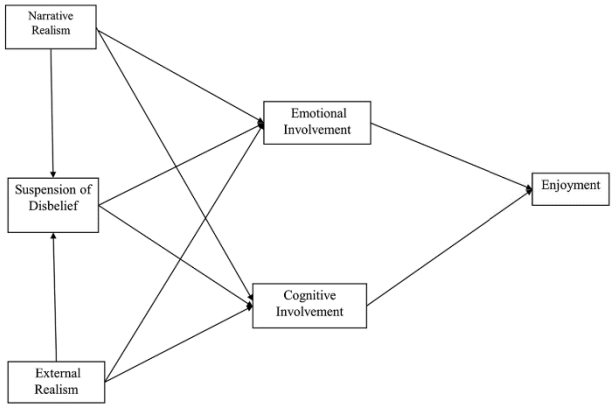


Figure 4: Ji & Raney (2006): Model of Entertainment Reception for Documentary-Style-Films

The following model will describe the intentions and characteristics of the elements used in each scene. The underlying premise for the model, is that realism is the basis for immersion, and that within the variations of qualitative and quantitative elements creating realism, there is a fluid space for designing a desired experience.

High - Emotional Involvement

Low - Cognitive Involvement

High - Cognitive Involvement

• High Subjectivity

Description:
Elements and compositions that fall in this area are highly suggestive yet ambiguous. These are evocative elements which might contain references to themes not present in the scene (either from a different time point, or altogether independent)

Application Scenario:
Opening and Closure

• High Subjectivity

Description:
Elements and compositions that fall in this area are both highly suggestive and simultaneously provide information relevant to current events. Having an emotional and cognitive loaded content has intense results

Application Scenario:
Climax or highlights

• High Objectivity

Description:
Elements and compositions that fall in this area are positioned in the background and are more contextual. This gives a higher protagonism and relevance to the visuals.

Application Scenario:
Transitions

• High Objectivity

Description:
Elements and compositions that fall in this area are rich in objective detail and allow the audience to focus on the story, an atmosphere or an on-going event.

Application Scenario:
Informative scenes, plot related

Low - Emotional Involvement

Figure 5: Source: Author

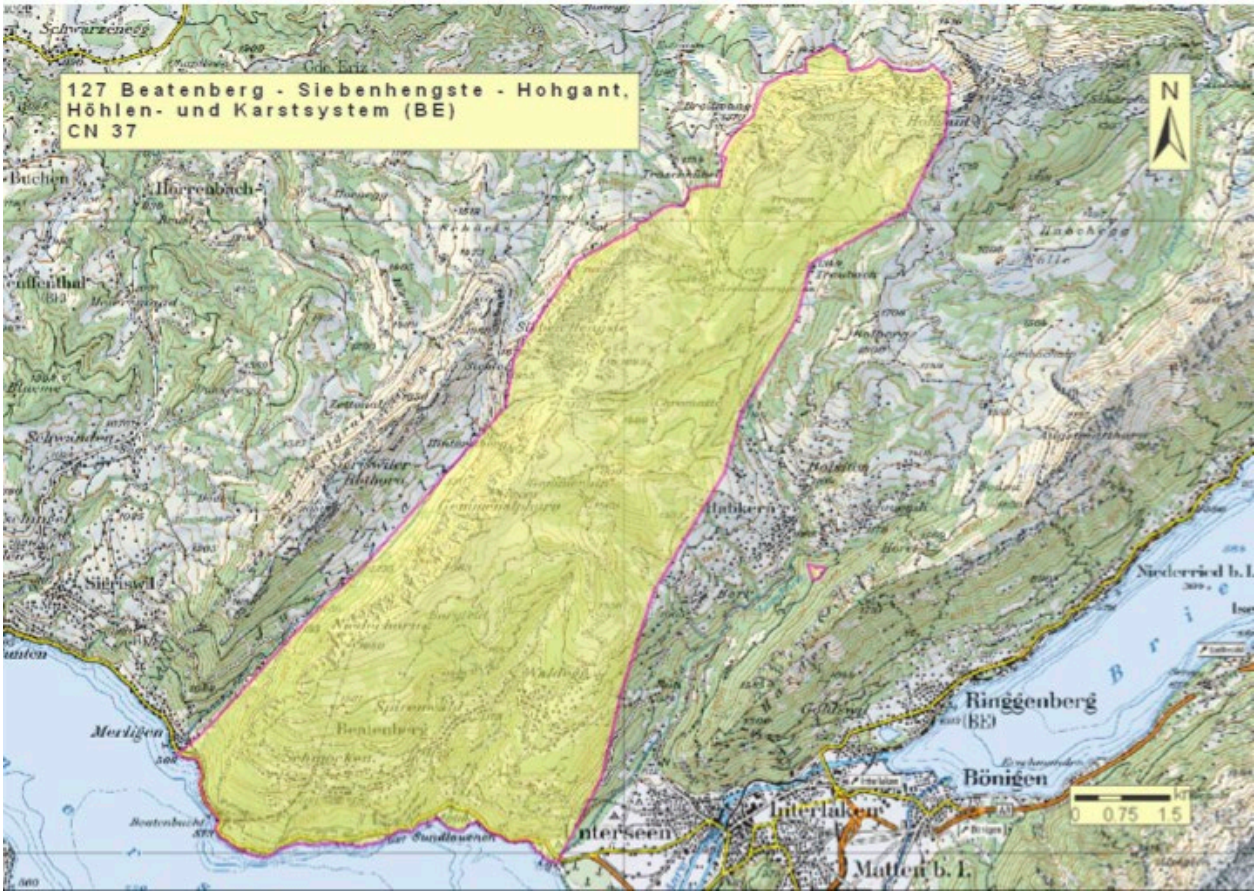
Additionally, to the insights provided by the model, I will be providing an overview of the project, detailing the location, the story and outlining the most relevant information regarding the team, the process, and my workflow. A Scene Analysis will dissect selected “rooms” by describing their context, and by delving into a *Narrative* and *Spatial Analysis,* that guide the sound-design and music through emotional and mood queues in the story as well as help draw inspiration and references based on the materials, the textures and other haptic inputs and spatial qualities.

Finally, I will be explaining the argumentation for the sound design and musical design, including my interpretation of the story from the director’s perspective, and its translation into sonic solutions, providing an overarching sonic “Story-Arc” with musical structures and sound design elements. I have approached the auditory level of the project holistically and attempted from very early stage on consolidating a homogeneous aesthetic, which is why I will be often referring to it as a “soundscape”.

The immersive pre-condition provided by the technology, “between film and game setting” where we can look but can’t move (allowing us to have 360 views and binaural sound), bridges a difficult gap of grasping the audience’s attention, and leaves us with the challenge and goal to convey a story.

3.2. The Place

Location: **Siebenhengste-Hohgant-Höhle**

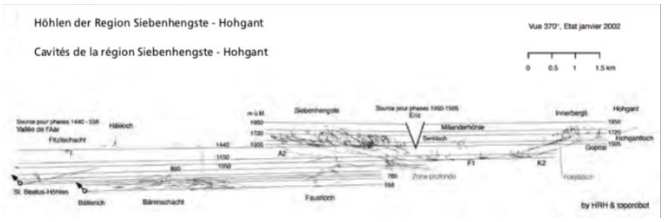


https://data.geo.admin.ch/ch.swisstopo.geologie-geotope/PDF/geotope-CH_127.pdf

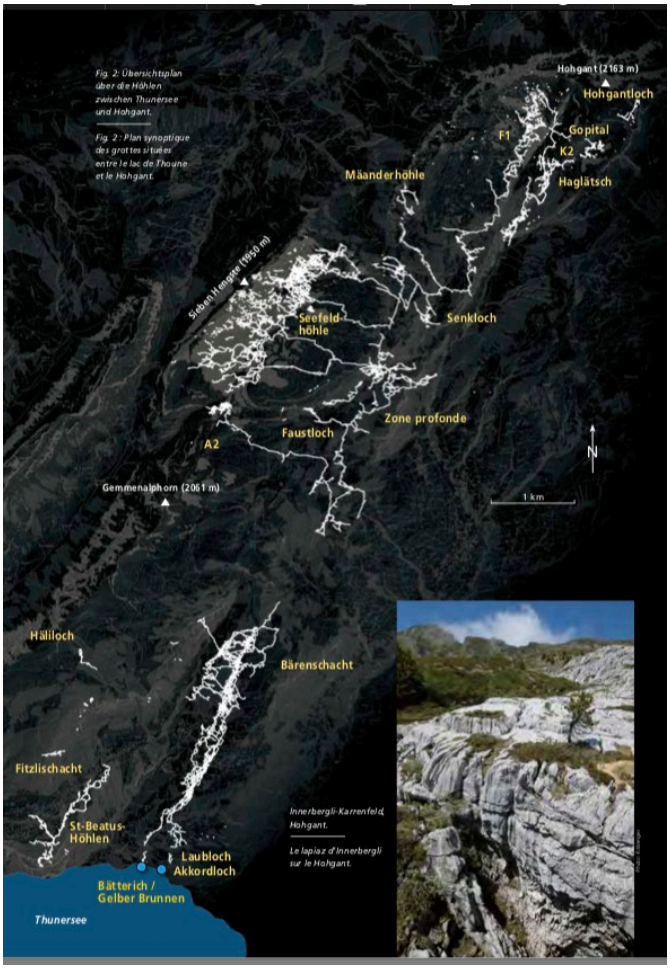
The *Siebenhengste* are a system of caves located in the Bernese Oberland, north of Lake Thun. It is in the *karst* area of the same name north of Interlaken. The cave system consists of various interconnected caves, and it is the second longest cave in Switzerland after the *Hölloch* in the *Muota* Valley. The measured length of the *Siebenhengste-Hohgant* System is over 164.5 km.

The vertical extension is 1340 m (Wikipedia, Siebenhengste-Hohgant-Höhle, 2021) Even though the passage length stretches over impressive length of 350 km, only a fraction of the "*karstsystem*" has been explored (Häuselmann, 2016).

In 1966, the cave system, was explored for the first time by the caving club *Jurassien*. In this exploration process, 3 out of the 42 entrances known today were discovered. Owing to the promising results and to the vast unexplored extent of areas identified, the international speleological community *Hohgant Region* was formed in 1979 to continue exploration of the system, until today. (Wikipedia, Siebenhengste-Hohgant-Höhle, 2021).



(http://www.isska.ch/refbase/pdf/Hauselmann_2016.pdf)



(http://www.isska.ch/refbase/pdf/Hauselmann_2016.pdf)

As it is possible to see in the satellite image, the extent of the intricate systems is vast and the connected systems of different caves in the topography crosses long horizontal and vertical distances.

The study of "Karst" is typically associated with interests of geological, geomorphological, and hydrological nature. The topography and structure of Karst is created by the dissolution of rocks by underground water systems. (Wikipedia, Karst, 2022)

(*Siebenhengste-Hohgant* Scene from Caves ©Carlos Isabel Garcia)

3.3. The Story

> *"Mankind has landed on the moon and flies around in space. But under our feet there are about a million kilometers of cave systems – only one per cent of which has been explored."*

This is the opening statement for the 19-Minute long 360° VR Movie. The motivation and fascination for the underground world is clear from frame one. The director, Carlos Isabel García, follows a group of experienced speleologists in their journey underground. The director accounts his motivation for the movie by his fascination in people who place themself in borderline situations and how they handle themselves under those extreme conditions.

Throughout different interviews, I have heard him saying "it's just a walk through the caves" with a laugh on his face, as he describes his project. But the physical and mental effort of the crew and the protagonists while exploring these enigmatic, underground rooms, is very real and far from "walkable". Lea Odermatt, and siblings Nora Sanz and Diego Sanz, are close friends, who's trust in each other is crucial, according to the director, as every move inside the caves, under many meters of rock and far from contact can result in a disaster.

In the daylight, we meet the three protagonists walking through the Swiss fields, watching them before their journey into the earth begins, and before their journey becomes ours. Experienced with Virtual Reality Technology, the film brings both the physical as well as the emotional plane of the story to the foreground and places the audience amid these dark, quiet, and imposing mysterious spaces, awaiting to be explored and uncovered.

3.4. The Project

Protagonists:	Lea Odermatt, Diego Sanz, Nora Sanz
Director, camera and audio	Carlos Isabel García
Screenplay and editing	Thäis Odermatt, Carlos Isabel Garcia
Music, Sound Design & Mix	Patrycja Pakiela, Raumgleiter

Timing:
Three weeks audio- postproduction:

1-day foley- recordings,
1-day instrumental recordings,
1.5-day ADR recordings,
Sound-editing,
Sound –design,
Musical Composition,

The moment I joined the project, “Caves” had already been shot and it was in post-production phase, even very close to picture lock. I was supposed to get the edited picture with sound within a week. I could already start the conceptual work and research in terms of looking at the reference documentary movies about caves and reading about the topic. I needed to know how the underworld sound is like, since I have never experienced such journey on my own. Together with the rough-cut from the director, I have received a Sound-Protocol (see Appendix 2) with notes for audio where the director shared his indications for the soundscape.

The project “Caves” was a great challenge for me, as it required me to dive into the world of 360 film post-production, to learn new workflows and most importantly, to explore new narrative ways to tell a story in the mystical world of caves with 3d sound.

Workflow

Mix for Oculus Quest, 2nd order Ambisonics.
To create an auditive setting for potential immersive experience of the movie, which already has the advantage of the VR Technology, it seemed like a straightforward decision to work with spatial sound. An ambisonics mix was my goal, a fully 3d spherical audio with which I could place sound source objects on certain positions. I knew that the effect of spatial sound greatly enhances the full experience.

Additionally, the ambisonics mix was not restricted to any playback system and can be accordingly decoded for any platform needed.

While the director was finalizing the edit, I have started finding the best possible combination of tools for the most efficient workflow to create my final ambisonics mix. In early discussions, the format of the mix I am supposed to deliver in the final stage has been clarified: Second Order Ambisonics Mix (SOA, Ambix). The final playback platform would be the Oculus Quest, and the mix will be decoded binaurally. (Later on, a requirement has been added to export Youtube First Order Ambisonics /FOA mix).

As I was supposed to create a linear 3D export with non-interactive environment, my choice of tool, was standard DAW for Film Postproduction: Pro Tools Ultimate. There are a few differences between standard the Sound Postproduction Workflow and the workflow for 360 Video I would like to mention. The main difference is that positioning sound elements and the adjustment of their loudness must happen at very early stage of the project, which, in the classical workflow for film, is normally done later in the mixing process. Additionally, the testing of the interactive scenes becomes a more frequent task in 360 films due to issues in certain combinations of presets and elements in the final platform.

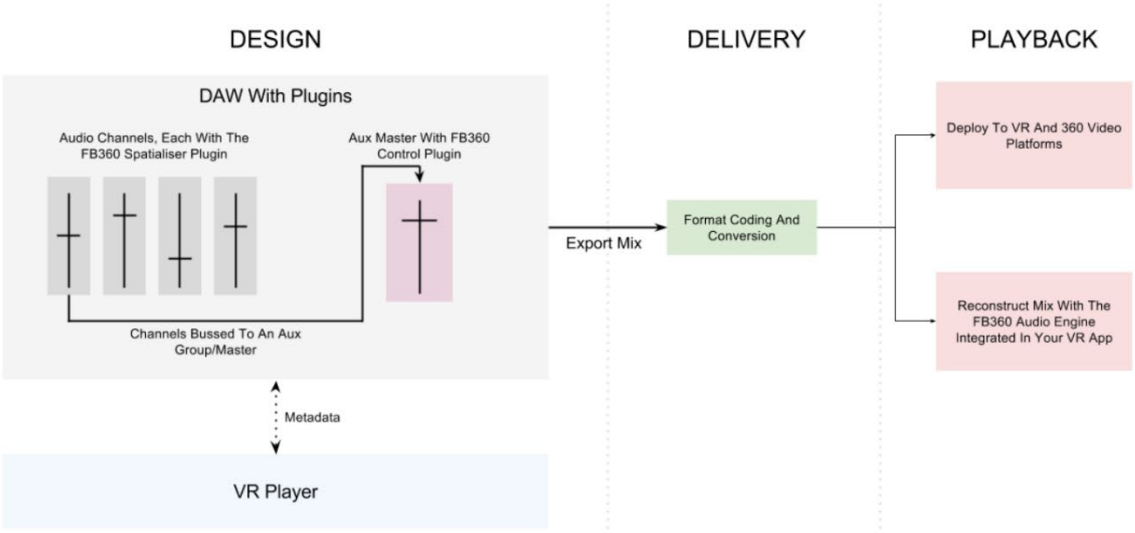
Pipeline

Concept	→	|	Implementation	→	| Delivery
Editing		|	Mixing		|
Recording	←	|	Positioning		|
Creation (SFXAssets)		|	Testing		|

I have tried several plugins available on the market at the beginning of the work-process. My priority in the ambisonics mix was to achieve a crisp and well sounding head-tracked ambisonic sound, decoded binaurally. That is why a plug-in using Real Spherical Harmonic Representation of Head Related Transfer Function (HRTF)-Algorithm was a crucial choice. In the interactive projects, I have experienced the use of many plugins for Game Engines delivering semi satisfying spatialization in the scenes, and in my opinion, that spatial aspect of sound in perceiving 3d audio is one of main qualities improving immersive-ness in VR.

There is no “one stop shop solution” which would provide a satisfying result. I was either confronted with test-export bugs, crashes of Pro Tools, weak positioning, or video player error, which added pressure in getting my template for the project ready. These limitations restricted further experimentation. From the available plug-ins on the market, for mixing in 360, I have tried combo-variations of workflows, including plug-ins like Gaudio, DearVR, 360Pan, BWaves 360. Through testing and some experience with the plugin, well documented website (FB360), and its CPU-friendliness, I have decided upon Pro Tools Ultimate and integrated in it Pro Tools Facebooks Spatial Audio Workstation (FB360). I additionally combined it with few Tools from Audio Ease “360 Pan” and naturally some additional plug-ins in Pro tools (for sound editing and sound design).

I was satisfied with their relatively straightforward workflow of FB360, a selection of plugins (including easy encoding plugin which provides multiple muxing options) and good documentation. In the phase of testing, it has offered satisfying results. An important aspect in this decision was also that the director, who wanted to be flexible with exports of the sound and the way he muxes it with video, was himself familiar with the Facebook 360 Workstation.



Proposed workflow for delivering and mixing, https://facebookincubator.github.io/facebook-360-spatial-workstation/Documentation/SpatialWorkstation/SpatialWorkstation.html

For design, mixing and spatialization, I have used Pro tools on Mac and to test the spherical sound and view 360 video I have set up a PC in "slave" mode and played FB360 Video Player on it connected to Oculus Quest with Beyerdynamic DT770 PRO Headphones. The head-tracking information was synchronizing with Pro-tools and in that way, I could control the mix real time.

The majority of the sounds are added and recorded during the sound-editing period. Due to very difficult filming circumstances in the setting (which director also mentions in the interview), we needed to make ADR recordings for almost all the scenes. For the ADR recordings we have used Schoeps CMIT 5 Shotgun Microphone and for each actor a Sennheiser Lavalier Microphone Set EW 112P G4 A with a Voice Technology VT506 Microphone.

General Sound- Concept

To define a concept for sound design and musical design was one of my starting points for approaching the project. During the first conversations with the director, as well as from analysis of his notes in a very early stage, I could create the first drafts of ideas to define which path I would like the sound in the movie to follow. We have discussed the concepts for each of the scenes and could find compromises where my interpretation would not exactly match the director´s vision. The fact that this is a documentary-style movie, in combination with the aesthetic of the pictures and the edit, directed my focus on primarily naturalistic sound concept.

The topic of the caves: through the edit, the notes and through conversations with the director, I noticed that his approach to the sound design in some selected scenes has been seeking the emotional involvement and attempting to portray the "unseen" in the picture. In his notes, Carlos has pointed out specific moments where he could imagine music to play an important role and additionally has placed some temporary music in the reference file.

With all that information I have tried to build a concept for the soundscape to convey an integrated holistic feeling, by following these two ideas:

1. Stay close to the documentaristic nature of the film, characterizing the spaces in the movie and supporting the potential spatial immersion.

2. Interpret and translate the feelings of the protagonists, the “unseen” of the storytelling and support the suspension of disbelief, and the temporary and emotional immersion.

Additionally, I wanted to find a sonic narrative arc and build a story throughout the movie.

Diegesis: Generally speaking, I would call the concept for soundscape of Caves as fluid in respect to diegesis. Staying in the documentary-style world of the movie, but at times, when the correct line-up of space, time, and suiting image for it emerged, the soundscape would be transforming into a mystical, surrealistic world through non-diegetic elements, where the “Caves” could raise their voice.

That “Voice of Caves” I wanted to achieve mainly through musical content. Through conversations with Carlos, I understood that musical design would be kept rather minimalistic in material and expression. Due to the inspiring picture, we could, relatively quickly, define the concept for musical design in Caves. The direction was to fit the general esthetic of the movie in its material: naturalistic and at times touching the unseen. That also informed my decision for the choice of esthetic in acoustic instrumental sounds.

I wanted that the sound-material in the musical design could symbolize the caves and bring out the “human“ aspects of the voice; the wind going through the tunnels, the metal of the carabiners, and the connection to the protagonists. For this purpose, I have chosen following instrumentation:

- piccolo, flute, bass flute
- prepared piano
- percussion

I have recorded particular sounds and effects on each of the instruments, built the gestures and processed the selected material using only minimum number of plug-ins: delays, resonators, reverbs. I wanted to keep the musical design in a haptic, rough aesthetic in order to portray that unknown of the underworld.

The development of that voice of caves has a parallel direction to the timeline of the movie. From a simple “call” (flute gesture: 00:00:53:15), where the protagonists are not yet underground, the caves themselves make attempts of “communication” throughout the movie and slowly build longer more prominent gestures (The Narrow), until the voice begins to “sing” (Lea Alone and Abseil Scene). Those musical moments with gestures approaching a sort of chorus, creates an atmosphere somewhere between the spiritual and the physical world. For the composition, I resorted firstly to improvisation, and afterwards decided upon using a very minimalistic material based on small intervals, quarter tones, dissonances, and atonal melodic motives.

While working on each of the scenes and depending on a setting, I have been using a combination of Set recordings, ADR- recordings, foley-recordings, field recordings and library sounds. All the sounds have been placed in the Facebook360 Spatializer plugin which afterwards could have been routed to various Aux Busses. I have used the workflow suggested by the developer. In the Spatializer plug-in for mono sound-sources, I was able to adjust various settings: position, the width of the sound, the attenuation, enable and point the directionality, enable the room modelling and set the position on vertical as well as horizontal plane. When I was working with ambisonics field recordings, I was able to adjust the orientation of the recording, so it was fitting the position given by the director.

In the session, there was a main 2nd order Ambisonic Aux Track which needed to have an instance of Facebook Control plug-in responsible for converting the mix to binaural headphone monitoring feed and setting the output to the headphone's hardware output. The FBControl plug-in allows to control and automate parameters for room modeling (size and scale) and to simulate room reflections accordingly. On that bus I run few loudness meters (Audioease, FB360), additionally Audio Ease 360 Radar helped me visualize the ambisonic recordings.

3.5. Scene Analysis

I have selected 6 “rooms” which provide both an overview of the applied concept for the sound in the movie, and an introspective look towards specific cases of particular interest in terms of balancing the subjective and objective representation of sound. The selection is not presented in a chronological order, and therefore it does not match the progression of events as presented in the film. Some scenes are of general importance to the narrative of the film and represent crucial moments in the expedition, such as major obstacles or features of the cave-system and difficult passages, while others render an inward picture into the characters, their tasks, and motivations.

3.5.1 Picnic

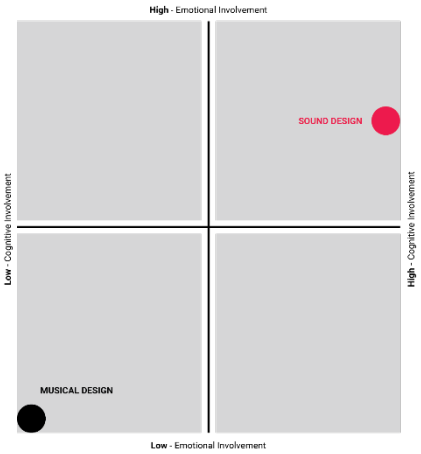
Timecode:

00:04:43:13 – 00:05:32:08

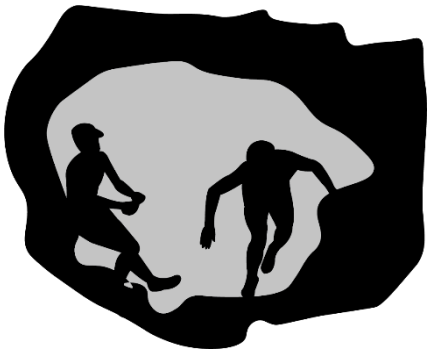
Plot:
After going through a tunnel, the explorers head into what appears to be a room with a waterfall, (suggested by the preceding scene). We land in a very quiet cave where the protagonists take a break.

Nora, Diego, and Lea are having a first resting moment for a snack. While eating, they discuss their next steps. They are planning to "open up" a new passage. They believe, because of a draft, that the next room might be of large volume. Diego points out there might be some work with cleaning the rocks. Lea whishes there will be more of tiny stalactites like the ones they have in the cave they are in now.

Narrative Analysis:
It's a "casual break" during the walk. The atmosphere is relaxed, and the protagonists are having a light chat. They bring their experience and knowledge of speleology in the conversation and speak about their plan to "open up" (exploding) a wall. For the viewer, the prospect of a possible explosion creates suspense and foreshadows a latent risk. For the first time, their plan is revealed, and a sense of time is provided. It is evident there is a strong bond between the explorers and a sense of comfort and camaraderie.



Spatial Analysis:
This room is a low elongated space, fitting 3 to 4 people in its widest section diameter. It is covered with stalactites and stalagmites, and the walls are inclined. There is humidity but there is no water dripping from the walls (according to the director).



Design Approach & Discussion:
Transitioning through a crescendo of sounds of water and percussion (from previous scene), we land in this beautiful, quiet cave. The director's aim was to have a naturalistic setting and to simply watch the speleologists, making the audience feel like one of them, sitting (camera height at eye level), observing, and participating in a friendly talk. I have attempted to follow these notes for the creation of the soundscape and achieve that physical realism.

Because of the clarity of the onset sound, we needed to use the ADR recordings of the dialog, as well as recorded foley sounds of each movement. Afterwards, the sound objects where spatialized accordingly, so the sense of presence is reinforced by the dialog and foleys as they provide a sense of proximity between the characters and the audience. For the ambience, we have decided on using a library sound

of quiet caves (ambisonic recordings FOA) and some subtle elements of flowing wind and light rock movements as mono objects in space. I have added some delay and eq to enhance the reflections, making the room's scale more evident. I have placed the objects in a 2nd order Ambisonic IR Reverb by 360 reverb (Audioease) and adjusted the settings to match both an intuitive perception of the room and the strict director's description of reality (experience based). The dimensions of the room were accordingly set in the Audio360Control plugin which additionally enhanced the depth and in my opinion the 3D perception for this cave. To focus the viewers' attention on the dialog and at the same time give freedom to look around the room, I purposefully did not overload this naturalistic scene with a complex soundscape. In my opinion, the visual content together with growing suspense in the light-mannered exchange of thoughts between the protagonists ("are they really talking about making an explosion underground?!), was more than enough to increase the emotional and temporary involvement of the viewers.

3.5.2 Tunnel

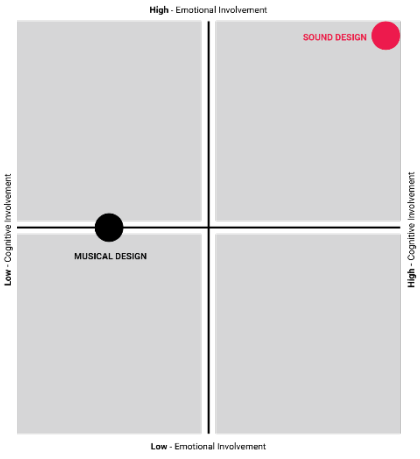
Timecode:

00:03:10:12- 00:03:44:14

Plot:
Following a hard-cut transition, we find ourselves in a new room. It is a and long tunnel filled with water. We see Diego approaching the camera. As he moves through, he passes us close by, Lea follows suit.

We are observing them transition through the tunnel, which is illuminated at its end (probably Nora's light), in the direction where the two are headed. Diego asks Lea if they can go through because it looks that it is quite narrow out there. Lea encourages Diego to put his head down and continue.

Narrative Analysis:
The protagonists are continuing their journey. Here we are in a room with different water sources and deposits. The environment is full of water sounds (there might be a waterfall nearby or some sort of big water flow). Diego and Lea approach. Nora is nowhere to be seen, maybe she is already far in front. We notice that going through this tract of the cave is physically straining, as the two protagonists breathe heavily and must crouch to cross the room. Towards end of the scene, they can stand up again, and Diego expresses his worry that they might not be able to go through the next passage. Lea encourages him to go further. The scene is in its nature transitional but shows the audience new features (water) of the cave, and the possibility that there might not be an open passage.



Spatial Analysis:
This is a very narrow long room. The director wanted to achieve a "noisy, watery environment" in this scene. We feel the proximity (almost surrounding) of a waterfall or a running stream. The pools (or water deposits) we are in, reflect the walls so we can see the rock material.

Design Approach & Discussion:
From the physical "spatial" realism perspective, the tunnel room was a very interesting design challenge, and a source of discussions with the director. This room "arrived" to me without any original sound, and only with reference music, so I had to imagine the space at first. I approached this room as a calm environment, choosing to represent just the stream of water and foley sounds of Diego and Lea. Right after hearing the first idea, Carlos expressed that it didn't sound properly and through some discussions, it became he was trying to recreate an experience, a sensation which was not altogether represented visually, but which he wanted to "explain" this memory through sound. Through naturalistic mix of diegetic and non-diegetic elements, we attempted to increase the level of information. The idea of mixing sounds of a 'fictive' or 'ideal' space was counterintuitive, but with time I quite liked that approach of Carlos which added a level of suggestiveness... does everything have to be explained?

The tunnel was loud because of the refractions, narrowness, and its closeness to a waterfall. It was not simple to achieve the spatial effect and get the balance so that we can still consciously navigate through the scene. I worked with spatialized library mono sounds, placed and automated foley recordings, some field recordings of Carlos, and Slapper Delay, adjusting reflections in the FB360. Altogether, the composition approximated the sound of the tunnel adding realism to the scene, and it suggests a larger body of moving water, as a powerful nearby presence.

Carlos's idea was that Diego and Lea are passing through, towards the "waterfall". In the end of the scene, the protagonists have a short exchange, which in an accurately depicted environment (based on the distance and the increasing levels of the water), should not be audible. To make the dialog in the end understandable, and without compromising the realism of the scene, we have agreed upon balancing down the general loudness, keeping the activity of the water nearby, but minimizing the "waterfall" effect. Through this compromise and fine-tuning I hope we could achieve a balanced combination of immersion and realism.

In the end of the "Tunnel" scene, and as a transition to the next one, there are two elements that intertwine, blending the descriptive but still remote sound of water, with a musical layer. The music consists of subtle rhythmical elements achieved through percussion recordings, which introduce the audience to the soundscape of the next scene. As it may be expected, there are a number of narrower transitional spaces in the cave system. These enclosures are somewhat similar in properties and relate to each other semantically in the narrative style as a "family" of spaces. This scene is closely related to the "The Narrow" (Musical Theater) scene, and it draws reference to its use of music.

These types of scenes where instrumental in my pursuit of a consistent style of soundscape and a coherent concept for the sound arc accompanying the complete story. A useful metaphor for establishing the connection between these scenes was their interpretation as a theatrical choreography of the protagonists, where their movements through space leave sonic traces and hint of an involuntary dance.

During the final mixing and finalization of the project, the director decided upon minimizing further the musical aspect of the tunnel, with the intention of emphasizing the diegetic-non-diegetic storytelling on

naturalistic and supernaturalistic sounds. Given the complexity, the length of the scene, and the amount of visual and auditory information provided, this was probably the right decision, as the development of a more elaborate meta-layer of meaning would be cut short and feel rushed. Perhaps some parallelism between the transitional scenes has been sacrificed, however, the space for other aspects of narrative, some more exploratory (distant water) than those active in the rest of these family of scenes, has been created. Hopefully, the audience can get involved on an emotional and spatial-temporal level.

The musical transition, being a sort of leftover of the "The Narrow" (Musical Theater) scene, could appear quite abrupt, and as a foreign non-diegetic element, potentially even breaking the immersion. Nevertheless, as we proceed in time, the story and the following scenes will strike a fluid balance between subjectivity and objective descriptiveness.

3.5.3 Lea in Water

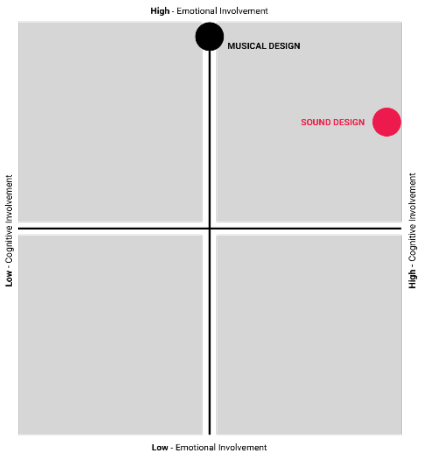
Timecode:

00:03:44:15 – 00:04:43:13

Plot:
This room is another tunnel, this time, almost filled with water, and the camera is only on the surface of the water. Our field of view is very narrow, as the ceiling is right above us. We see one of the protagonists approaching us, it's Lea. Jump cut- and now we are right next to her face, carrying us with her. We see that the cave is gradually changing its shape and it is opening.

Nora, who is waiting in the open space, calls Lea. Lea struggles through the tunnel and the water, behind us is Diego. Making his entrance, breathing heavily. Nora stands up and encourages the group to go further. The presence of water raises in dynamic, are we approaching a new room?

Narrative Analysis:
After the Tunnel surrounded by water we are being placed in a quiet environment, claustrophobic, narrow and desolated. Now water is the dominant perceptible element, we are immersed in it. From the light at one end of the narrow tunnel, there is light, which helps us to see movement coming from the opposite side. We are suddenly next to Lea, we see her eyes wide open, we hear her struggled breathing as she advances, submerged and leaning her head to stay above the water. She looks forwards and is focused. As she progresses, we hear water, we are getting closer to a new place. There is a sense of relief as Lea meets Nora who has already reached a more open space where she can stand. The timing of the movements (Diego joins the scene shortly afterwards), highlight the difficult passage through the narrow space. Nora encourages the group to go further. We follow into a smaller corridor; the big water source is getting closer. Are we about to see the waterfall?



Spatial Analysis: This is a room composed of a water-filled tunnel with a very narrow passage, leading to a small pool, then climbing up to a more open space through a sort of gate, where the protagonists can stand. There is a restricted perception of the room, as the camera is just above water level, the ceiling hovering close. It seems, the room leads to more open spaces through a passage.

Design Approach & Discussion:
In the previous scene there is a lot of movement from different elements: the water, two protagonists (Diego and Lea crouching though the tunnel), and dialog. The beginning of this scene, in contrast, is dominated by water related sounds which are slower and more still (light flows and occasional drops of water around, falling from the ceiling into the pool). The opening of this scene shows a new space, and it provides the audience some time to readjust to this new environment before further action happens. This build-up in tension is intensified by the nearness of the water and the realization of movement and signs of struggle approaching from the distance. Up until this point, the audience has been able to follow the group closely, being "guided" by their flashlights, their calm voices, and movements. Now the viewer is left alone in this claustrophobic setting. Even though the sound we hear carries the naturalistic components of the scene (light drops and water far away) it also contains this raising tone and some unworldly whistling in the space whose origin is not readily identifiable. The placement of those musical elements "head-locked", in stereo and not spatialized, had the intention to contrast the naturalistic sounds and to intensify the suspense, creating disorientation and relativization of time.

For that scene, an idea of combining the naturalistic elements with some extradiegetic sounds was discussed to create an ambiguous soundscape and accentuate for the viewer a sense of claustrophobia, solitude, and uncertainty.

The musical design, which begins in the previous scene with a percussive gesture, and transitions into light processed flute-whistles with increasing constant tone, functions as support for the scene's atmosphere and psychological mood. The jump-cut, which brings us close to a yet "unknown" perspective of Lea, and her emotions at the time, is letting us experience the situation 'first-hand'. We switched from a static camera to a hand-held camera carried by Lea, with which we hear her breathing and bring back a naturalistic, and documentary style into play. This cut into documentary style is partly reflected in the sound design, performing an ambiguous role, attempting on one side to bring us back into the physical world, and on the other to make an emotional connection to the character. The dichotomy creates balance through opposition and reinforces the narrative immersion by providing both context and a new plane of content.

We have started the scene with a supernaturalistic approach to the sound, working on the perceptual realism and transferring into a more naturalistic one, building the physical realism of the space. The transition between the rooms was achieved through combination of Impulse Response Reverbs crossfading, delays, and automated room transition in the FB360 Control Plugin. The water movements (foley) and recordings ADR were placed in a space and had adjusted directionality. The approaching waterfall increases in dynamics, while the spatialized sounds of water overtake the scene.

3.5.4 The Narrow

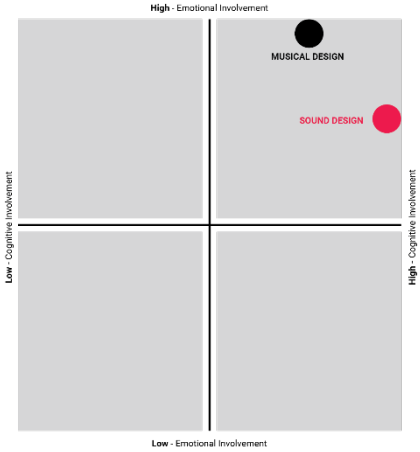
Timecode:

00:05:32:16 – 00:06:13:13

Plot:
After observing the protagonists' break for picnic, we land in a very narrow tunnel together with the group. We watch their movement, as they crawl through that small space, all in one direction.

One after the other, first we see Nora, then Diego follows her, and as last comes Lea. We are very close, and thus can hear them struggle, their eyes focused on the one in front. They don´t exchange one word in this silent procession, as they all seem follow their 'steps' carefully.

Narrative Analysis:
After the Picnic break, where protagonists discussed further steps, the audience is being brought into a scene of pure action and movement. We have a chance to see a new room. As the protagonists traverse the long room, it seems as they are mimicking each other's movements. At this near distance, we can hear their clothes dragging through the rocks, the metallic jingling of their harnesses and carabiners, and the strenuous breathing, all placing the audience in a close encounter with their current reality. The protagonists are focused and aware. They seem certain of their actions even though they explore new territories (as mentioned in picnic scene). They know the feeling of being in these narrow spaces and convey experience in the process.



Spatial Analysis:
This room is very narrow and cylindric. The dimensions force the crew to crawl horizontally on their bellies. From the perspective of the camera and because the protagonists show signs of fatigue, it seems as though this room is quite long. The atmosphere is dry, and we see no water on the rocks either. The director mentioned that this room was very silent.

Design Approach & Discussion:
The director wanted to bring the viewer into this real situation which often happens to speleologists, going through a very narrow space, and being challenged by the architecture of the underground. The linearity of the space and their movement was a reference I wanted to follow and support through a line of musical gestures that would emulate directionality.

To showcase the features of the space, and achieve the physical realism on this room, I worked with foley recordings which where precisely spatialized for each of the characters and mapped closely to their movements. Their recorded reactions and sounds (grunts, breaths, gasps, etc.) as well as very light ambience, have contributed to the diegetic representation on this scene.

A further layer of the naturalistic soundscape came from an input from the director, who wanted to achieve a very dry environment, almost completely silent. The lack of any reflections given the narrow shape and dimensions of the room, gave this scene a kinetic character focused on highlighting the physical effort, providing a sense of speed (or slowness).

The soundscape is completed with a musical design which is surrealistic. The aim was to support the narrative side of the story and add to the growing presence and increasing "personality" of the caves themselves. This is one of the first moments where the entire cave system starts "communicating" with the group of speleologists, who are venturing deeper and deeper inside of them. We wanted to accentuate this intimacy and connection with sound. The caves themselves, begin to create short "musical" gestures, manifesting music as the voice of a fourth protagonist of the film: space.

The movement of the group, the quiet and sure rhythm of that scene and the stillness and focus perceptible in this room, have inspired me to think of the process that created these rooms and the forces that over the years have carved the rock diligently. This odd pair of actions, a sort of collage, in the contrast of movement and stillness and the mundane passing of humans vs the eternal process of nature, create a musical theater that sets the first stage for a concert of the "voice of the caves".

That musical theater accompanies the group in their rhythmical movement through the tunnel and tells a story parallel to the simple action of their journey. Through that combination of the documentary sound design and the strong contrast with musical gestures, the soundscape makes a statement of involving the viewers by engaging multiple senses and appealing to different types of cognitive inputs. Ideally, the scene will awake curiosity and spark emotion while maintaining the awareness of their presence in space.

The musical scene begins already transitioning from the picnic scene, as a symbol of the "motivation" of the characters which has been revealed. A piano is introduced, evolving elements into a rhythmical motive of the percussion which "transfers" a musical flow (as referential to tunnel) into flutes in a subtle continuum. Those musical gestures come as bulks (like the packets of pressure of air passing through that tunnel) and they have a clear linear development, varying the intensity of the material. The end of the scene (00:06:01) foreshadows the next scene transitioning into musical "wind".

In order to accentuate the dualism and ambiguous feeling in this scene, the sound design is being spatialized and the musical design is head locked. The Caves are raising their voice, but this voice is telling a different story than the physical narrative of the room. The combined effect is supporting the narrative immersion and the emotional involvement. There is one more design element which is symbolizing and bridging two worlds: wind.

The wind, which also plays a role in the next scene, is already lightly to be perceived here, but to create that ambiguity, we decided on giving the responsibility of the physical sound of wind to an instrument. The eolian sound effects (flute), carry the reference to wind from this scene into the next.

The choice to translate the wind in this scene through music was meant to create a progression developing across multiple scenes and depict the developing "voice" of the Caves which will grow into the "Song of Caves" (not part of the analysis) at a further time in the story (00:06:38). The scene is a rough blend between surrealistic and documentaristic soundscape. Through that dualism the suspension of disbelief is building up.

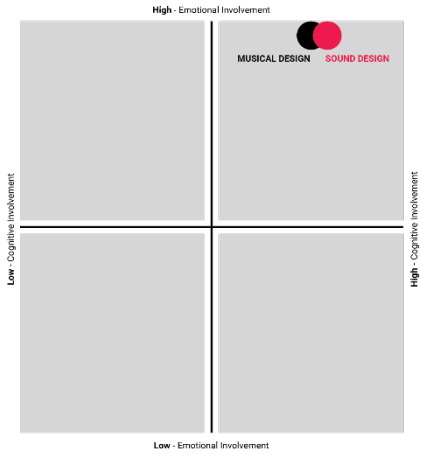
3.5.5 Bringing the Water, “Thousands of Years”

Timecode:

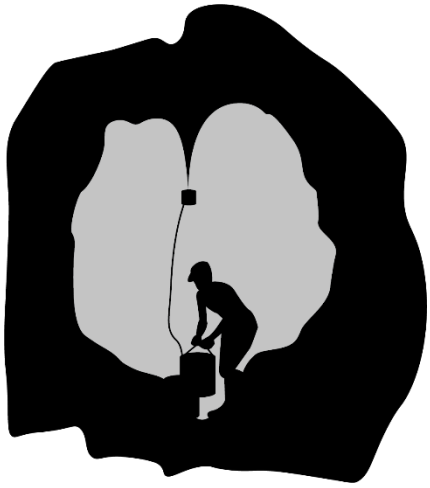
00:09:22:17- 00:10:00:15

Plot:
Lea is being sent to bring water by herself. We watch her leaving the previous scene and the group. Diego says: “don’t get lost” as she leaves the other room.

In “Bringing the water” we see Lea sitting on the bottom of a room. There is a small construction set up for the purpose of collecting water. It is a simple hanging funnel that collects drops channeled by a prominent stalactite. Lea manages the set-up on the bottom of the cave, she changes the hose from the funnel, setting an empty container and taking the full one back to the group.



Narrative Analysis:
Lea is “lightly” warned by Diego in a previous scene on not to get lost, to what she replied that she will manage it. We don’t know how far Lea had to go to reach that place. By now we understand that some of the dangers and difficulties associated with their journey are not being fully disclosed. The protagonists, being experienced practitioners and aware of the challenges associated to their activity, are able to joke about them and can address them with a certain lightness.



We watch Lea doing what seems to be a routine activity (for us an adventure) as her swift and efficient movements indicate she knows how the system works. The camera is placed about 3.5 meters away from her and gives the audience an ample view of the magnificent room. Thanks to the simple action taking place, the audience has time to discover the surroundings, to image how the protagonists survive in the inhospitable environment and to ponder on how this structure has been created.

Spatial Analysis:
This room has a complex architecture. Through two sources of light, we can perceive the bottom or “floor” of the cave where Lea is seated. The light also shines on the ceiling, revealing beautiful shapes and geometries, probably created by the water. The cave doesn’t have a closed form, it is hard to estimate its horizontal extent. It is evidently humid since the speleologist's depend on this humidity to collect their water. The ruggedness and complexity of the walls and ceiling, however, does not reduce sound reflections.

Design Approach & Discussion:
In the notes from the Protocol, Carlos refers to this scene as "Thousand years", probably in an attempt to describe the formation process that can be seen solidified in the complex geometries of hanging stalactites. My goal was to convey some of that history and inspiration with the sound as well. I planned to achieve this by strengthening the viewer's sense of presence in the room, and to incorporate elements suggesting the passage of time, or bringing the audience's attention to time-measures- rhythms.

Carlos wanted a room in which one could hear water dripping lightly, but at the same time he wanted to portray a strong surrealistic story in the background, the patient sculpting of the walls by single drops and threads of water: Movements, transforming and affecting one another, shaping themselves slowly.

I interpreted the intention of Carlos as a homage to the beauty of the caves and as an opportunity for the viewer to just observe and learn. For "Thousands of Years", I have been working with single attacks of the hi-hat, processed through resonators, delays, playing with the reversal of the sound in combination with winds (picc, fl, bfl).

I interpreted the passage of time and the slow morphing of the stones, into motives which are constantly evolving, at times repeating themselves with small differences and staying in constant movement of various intensity. In this room, melody does not really have a direction of expression, it is just there. A mono-foley of the dripping water, sent into impulse response Reverb, accentuates the rhythmical drops in the actual space.

The idea of translating the room's history into music was to me interesting and I definitely got inspired by it. I doubt that viewers will fully grasp the extent of the metaphor and our interpretation of this scene, since it is very isolated, and the length of the scene might allow the viewer enough time to it. Additionally, the idea itself is very abstract.

I think though, because of that simple analogy of the water drips to the music (hi-hat), blended with the naturalistic sound effects and the provided time to experience the scene (we watch a simple activity for more than 30 seconds), the viewer is still allowed to emotionally involve on a different, more personal level with that space. The distance between the camera and the protagonist leaves enough room for the audience to drift the attention and gaze away from the simple action, to wonder and add a reflective "surreal" pause between the two highly naturalistic scenes surrounding this sequence.

This scene as well, treats the diegetic sounds as objects and they are specialized in the room. To bring the history closer to a realistic room, I have spatialized the single hi-hat notes in the space, so that the musical design comes closer and becomes more present on its own. Up until now, I have seldomly spatialized musical elements, but it seemed the right approach to the intended effect. However, due to the duration and subtlety of the effect I don't think that the decision had a significant impact on the immersion of the audience.

At first, I have identified this scene as the surreal, mystic scene, as the musical design which is non-diegetic, is placed in the foreground. This scene lays very much on the border between two worlds where Caves sing on their own, but the protagonist is very involved in her world.

3.5.6 Abseil

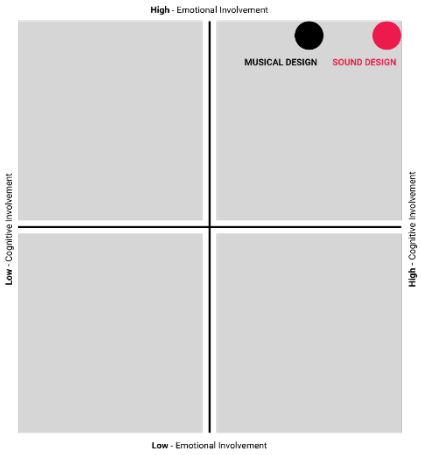
Timecode:

00:15:12:17 – 00:16:17:14

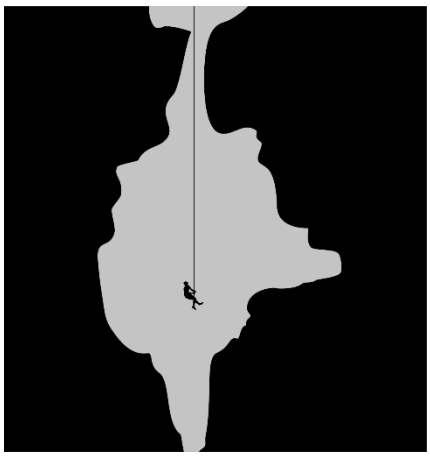
Plot:
The protagonists are close to the end of their journey. We know from the past scene, that they almost ran out of food and will proceed to the surface again. This scene is set in a gigantic, vaulted room.

We see a person struggling up on a rope, we see the water falling through the walls. There is light coming from above and from the bottom. Towards the end of the scene, we hear Diego encouraging Lea to continue climbing.

Narrative Analysis:
Our protagonists have taken us on an adventure trough the innards of the caves and, as announced in the previous scene, they will now head up to the exit. Until now, we have not yet seen what Lea has referred to as "new territory". Is this the Cave they have been willing to explore? We watch her climbing up patiently and struggling through the long rope. The setting looks very dangerous; however, the speleologist reflects a calm and persistent attitude. Diego's presence in the scene (and with it the sense of team-effort) is reminded by him calling on her and motivating her. They watch each other's back while leaving this monstrous underground.



Spatial Analysis:
This gigantic gallery opens to our view. It's the biggest cave-room in the Swiss alps. It is 180m high and covered with water dripping on the walls. We can see the endless darkness in the bottom.



Design Approach & Discussion:
As mentioned before, in the scene "The Narrow", the Caves start to create their own musical structures, and we start isolating their character and hearing their voice. In a similar scene to the "Abseil" the one right after the Tunnel: "Lea alone" (00:06:38:00-00:08:01:00), the Caves start singing their Chorus for the first time. In the scene "Lea alone" we are being placed very close to Lea, and we hear her movement, breath, light fear (?) and we infer respect in the eyes. She is watching the cave while going down, is she hearing something? A parallel soundscape approach for the materials has been used in the current space referencing "Lea alone", the narrative, though, has evolved.

“Abseil” was another homage to the caves from the director. He wanted to show the viewers, this biggest cave in the Swiss Alps, which many would otherwise never get to see. With that, he aimed to portray the magnificence, calmness and as well the danger of the adventure the group has taken.

I wanted to support that approach in the soundscape by enhancing the interesting physicality of the room in sound. However, the visual content and the narrative was bringing our attention to this gigantic space, so I wanted to emphasize the emotional involvement. This scene could be interpreted as a potential climax of the film in terms of visuals and narrative, and this was an aspect I also wanted to reflect in the soundscape. I chose to work with field recordings of the director (FOA) as well as library ambiences and mono objects specialized around the room to convey a “respect” to the physicality of the room and provide a better sense of its scale. I have added some filter delay to the objects, a fitting IR Reverb and adjusted the Facebook Control Plugin for depth.

The protagonist is (acoustically) unnaturally close to us in this scene (we hear her breathing and struggling, although she is well away from the camera). In the wet and reverberating environment of the cave, we might be able to hear her presence, but this was an unnaturalistic choice to support the narrative and show the physical struggle of the protagonist climbing up.

Musical sound design was a chorus of caves (material introduced in “Lea alone scene”) including previously introduced single detuned gong motives which were supported by a processed bass flute and flute chords. The “call” of caves, previously heard in the field (00:00:53:15), comes back here for the last time but instead of hearing a quartertone-glissando chord, we now hear single glissando tones, as objects placed around the room. This symbolizes a farewell from the caves to our protagonists.

During the final mixing, the director decided upon eliminating a few elements which were enlarging the importance of the musical design in the scene. Additionally, we have decided on giving a new balance between the physical room soundscape and the music in this scene. The musical design acquired a bigger presence in previous versions, and upon finalizing the project, the number of elements, musical gestures and dynamics has been adjusted down. At first, I thought that this might break the suspense. I wanted to support the emotional and visual culmination at that point.

In retrospective, I think, that because of the visual content in the scene and the narrative of the story, the suspension of disbelief was not broken or decreased. The viewers could stay closer to the protagonist who was struggling in her dangerous climb. Potentially this decision even positively affected the emotional immersion in this scene.

4. Discussion & Conclusions

> *"Exploration is in our nature. We began as **wanderers, and we are wanderers still."***
> *-- Carl Sagan, Cosmos*

Through an analytical introspection into the soundscapes of "Caves", I have addressed in this work the relationship between sets of core concepts, such as realism and design, authenticity and spuriousness, and presence and narrative, all underlying the experience of the 360 VR documentary film. The questions posed by the investigative line, have targeted the elusive balance of aspects, ranging from technology and psychology to even personal affinity and pre-existing domain knowledge, which influence the fascinating phenomenon of immersiveness in media. The selected use-case, servers to exemplify and explore the design decisions, the enabling technology and the challenges and learnings resulting from the project.

A theoretical foundation based on the works of researchers in cross-discipline problem spaces like game-design, film, and interactive media, as well as in narrower more specialized domains like VR Technology or Psychology, provided me with tools to probe at the "fabric" of immersiveness and to look closer at its enabling principles and its malleable underpinnings in storytelling and narrative, tracing back to the most elemental inner-workings of our perceptual apparatus. The interplay of space, time and narrative played a crucial role in the interpretation of the relationship between a story and the (Emotional & Cognitive) involvement of the audience. The contribution of immersive technology in the multisensorial engagement and through this, the enabling of a "positively tuned state" of perception (or Suspension of Disbelief) was likewise key to understanding the scope of effect of spatial and narrative sonic design.

With the help of the designed framework based on spatial and narrative descriptions and the cognitive and emotional impact, I dissected the elements and motivations behind the sound and musical design. The framework allowed for a consistent inquiry into the aforementioned interplay between space, narrative, and sound.

In applying the same inquiring instrument and logic, a certain differentiation emerged based on the classification borrowed from Görne (2017), grouping the analyzed soundscapes into the following categories:

- **Naturalistic**: "Picnic" & "Tunnel"
- **Supernatural/Attention-Guiding**: "Lea in water" & "The Narrow"
- ***Semi* - Surreal, Mystic**: "Bringing the Water" & "Abseil"

The analysis also revealed that most of the scenes had a very fluid profile of the mix between the designed characteristics of naturalistic and surrealistic sound design and musical design, where the competing or complementing intensity of these two sonic layers mapped closely to the general desired "weight" of the narrative. Additionally, a clear differentiation was not always strictly feasible, since the design often involved the interconnectedness and merging of the layers, or a functional inversion in their typical role (music as informational, descriptive tool and sound-design as an evocative and referential medium), thus some soundscapes display a duality and "in between" approach to narrative.

Among the main findings of the analysis, is that a strong correlation of immersiveness and narrative delivers a more intense and engaging experience, activating both the emotional and cognitive involvement. It was counter-intuitive to realize the interchangeability of sonic functions and the high "tolerance" an audience can have for mixing extreme diegetic and non-diegetic elements so long as they are coherent.

This, for me, highlights the importance of a holistic approach of an all-encompassing soundscape vs a component-based design of sound elements and music composition. The role of VR Technology and spatialized audio was seen to provide a baseline for the Suspension of Disbelief which allows to incorporate other narrative fluid elements.

Limitations & Further Research

Even though one might consider the lack of full "freedom" (DOF) in 360 VR Film, as a hindrance to immersiveness, the very "negligible" conscious action required (head-tracking) is actually liberating the audience from distractions. The speed and the imperceptibility or transparency of this interface (strong coupling with our body) intuitively reinforces focus and supports the linear narrative sought by the director. (In our use case the use of light in combination with 3d sound was paramount to achieve this).

Because VR Technology is often still considered a new medium, representing some technical hurdles, and requiring some experimentation and innovative approach to narrative, directors of linear media often shy away from it as a veritable storytelling platform. However, as seen in this project, the very same aspects that make it challenging, are opening possibilities for interesting and creative results. What remains an obstacle for the time being, or at least a differentiating factor in experience, is that the infrastructure required for "screening" is more elaborate and isolating in comparison to regular film. But again, as far as I can say, the technology also makes a huge contribution to the immersiveness of the audience. It will remain a task of further research to validate the possibility of achieving similar results (also balancing a wide-ranging soundscape utilizing surrealistic and naturalistic elements) without the support of Immersive Technology.

I believe, that exploring and challenging certain preconceptions about VR technology, can hopefully help uncover further possibilities for storytelling while diving into new narrative worlds.

Bibliography

Arango, B. (2021, November 26). *What's the difference between 360° video and VR video?* Retrieved from Filmora: https://filmora.wondershare.com/virtual-reality/difference-between-360-video-vr.html

Arron, M. (2016, May 31). *Immersive storytelling is everywhere and there's no going back.* Retrieved from The Guardian: https://www.theguardian.com/media-network/2016/may/31/immersive-storytelling-360-vr-technology-marketing

Barnard, D. (2019, 05 05). *Virtual Speech.* Retrieved from Degrees of Freedom (DoF): 3-DoF vs 6-DoF for VR Headset Selection: https://virtualspeech.com/blog/degrees-of-freedom-vr

Barton, L. (n.d.). *Immersive Technology*. Retrieved from advrtas: https://advrtas.com/immersive-technology/

Bolter, J. D., & Grusin, R. (1999). *An Introduction to Immersive Technologies.* MIT Press. Retrieved from Vista Equity Partners: https://www.vistaequitypartners.com/insights/an-introduction-to-immersive-technologies/

Colm O Fearghail, C. O. (2018). Director's Cut - Analysis of VR Film Cuts for Interactive Storytelling. *International Conference on 3D Immersion (IC3D).* Brussels: IEEE Signal Processing Society.

Cuddy-Keane, M. (2005). Modernist Soundscapes and the Intelligent Ear: An Approach to Narrative through Auditory Perception. . In P. J. Edited by James Phelan, *A Companion to Narrative Theory* (pp. 382-398). Blackwell Publishing Ltd.

Cummings, J. J., & Bailenson, J. N. (2016, 05 15). How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology*, pp. 272-309. doi:10.1080/15213269.2015.1015740

Encyclopædia Britannica, 11th Edition. (1911). Cambridge: University Press .

Encyclopaedia, T. E. (2020, May 29). *Encyclopedia Britannica.* Retrieved from realism: https://www.britannica.com/art/realism-art

Fu, X. (2021). Acoustic Narrative. In *Chinese Narratologies* (pp. 199-221). Singapore: Springer.

Gardonio, S. (2017, October 29). *What's the Difference Between VR, AR, MR, and 360?* Retrieved from Medium: https://medium.com/iotforall/whats-the-difference-between-vr-ar-mr-and-360-139fcf434585

Georgiou, M. (2020, November 10). *Is Immersive Storytelling the new marketing frontier.* Retrieved from Forbes: https://www.forbes.com/sites/forbescommunicationscouncil/2020/11/10/is-immersive-storytelling-the-new-marketing-frontier/?sh=3c21ad03698c

Goldman, J., & Falcone, J. (2016, March 9). *Virtual reality doesn't mean what you think it means.* Retrieved from CNET: https://www.cnet.com/tech/mobile/virtual-reality-terminology-vr-vs-ar-vs-360-video/

Görne, T. (2017). The Emotional Impact of Sound:A Short Theory of Film Sound Design. *EPiC Series in Technology*, *1*, pp. 17-30. Hamburg.

Hameed, A., & Perkis, A. (2018). Spatial Storytelling: Finding Interdisciplinary Immersion. Lecture Notes in Computer Science. In R. Rouse, H. Koenitz, & M. Haahr (Ed.), *Interactive Storytelling. 11318*, pp. 323-332. Springer, Cham. doi:https://doi.org/10.1007/978-3-030-04028-4_35

Häuselmann, P. (2016). Das Karstsystem Siebenhengste-Hohgant-Schrattenfluh: Versuch einer Synthese. *STALACTITE*, 28-41.

Ji, Q., & Raney, A. A. (2016, 12 1). Examining Suspension of Disbelief, Perceived Realism, and Involvement in the Enjoyment of Documentary-Style Fictional Films. *Projections: The Journal for Movies and Mind, 10*(2), 125-142. doi:https://doi.org/10.3167/proj.2016.100207

Lombard, M., & Ditton, T. (1997, September 1). At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication*, p. unpaginated. doi:10.1111/j.1083-6101.1997.tb00072.x/

Murray, J. H. (2017). *Hamlet on the holodeck: The future.* MIT press.

Nilsson, N. C., Nordahl, R., & Serafin, S. (2016, November 2). IMMERSION REVISITED: A REVIEW OF EXISTING DEFINITIONS OF IMMERSION AND THEIR RELATION TO DIFFERENT THEORIES OF PRESENCE. *Human Technology, 12*, pp. 108-134. doi:http://dx.doi.org/10.17011/ht/urn.201611174652

Passi, A. (2019, December 25). *Elements of Good Story Telling.* Retrieved from UX Planet: https://uxplanet.org/storytelling-elements-31b2a6a7e373

Realistic, Responsive Audio Is The Next Frontier In VR Innovation. (2018, January 3). Retrieved from cbinsights: https://www.cbinsights.com/research/vr-audio-tech/

Rodrigo Oliva, J. B. (2019). Video and storytelling in a digital world: . *Comunicação e Sociedade, vol. 32*, 459-476.

Ryan, M.-L. (2001). *Narrative as Virtual Reality; Immersion and Interactivity in Literatureand Electronic Media.* HE JOHNS HOPKINS UNIVERSITY PRESS.

Schubert, T. W. (2006). The sense of presence in virtual environments. *Media Psychology (Zeitschrift für Medienpsychologie)*, 69–71.

Slater, M. (2003, 01 01). A Note on Presence Terminology. *Presence Connect, 3*, p. unpaginated. Retrieved from https://www.researchgate.net/publication/242608507_A_Note_on_Presence_Terminology

Steinicke, F. (2016). *Being Really Virtual : Immersive Natives and the Future of Virtual Reality.* Web.: Springer International Publishing.

Väljamae, A., & Tajadura, A. (2007). Perceptual optimization of audio-visual media: moved by sound .

Wikipedia, c. (2021, October 31). *Siebenhengste-Hohgant-Höhle.* Retrieved from Wikipedia, The Free Encyclopedia.: https://en.wikipedia.org/w/index.php?title=Siebenhengste-Hohgant-H%C3%B6hle&oldid=1052847779

Wikipedia, c. (2022, January 9). *Karst.* Retrieved from Wikipedia, The Free Encyclopedia: https://en.wikipedia.org/w/index.php?title=Karst&oldid=1064669612

Witmer, B. G., & Singer, M. J. (1998, June). Measuring Presence in VirtualEnvironments: A PresenceQuestionnaire. *Presence, 7*(3), 225-240.

Statement of independent work

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it by myself without the use of any sources beyond those cited, and all passages and ideas taken from other sources are cited accordingly.

Authored by:

Last Name:	**First name**:
Pakiela	Patrycja

With my signature I confirm that

- This is my original work
- I have documented all methods, data, and processes truthfully
- I have mentioned all persons who were significant facilitators of the work

Place, date	**Signature**
Zürich, 12.01.2022	Patrycja Pakiela

Appendix

1. Interview with Director Carlos Isabel Garcia

Q-1:
The following quote opens the film: "Mankind has landed on the moon and flies around in space. But under our feet there are about a million kilometers of cave systems – only one per cent of which has been explored.". What was the main motivation to tell this story; is the story's focus on the activity of "caving" and the place what you wanted to share, or where you also interested in exploring the human psyche in this context? If both, how would you describe the balance?
A-1:
So, it's about both, it's about the place and it's also about the people. It's about the place, because I think in 360 it's also important, it's not only we usually think it through 360 we compare it or think it as a film and it's a little bit more. It's film with a space and a space is quite important. So, it was about the caves and battles about caving. So, caves I like them because no one has been there, most of the people haven't been there are like hidden places and places that are about the human psyche. I like the way caves deal with fears, Uhm, and I like the way these kind of regular people like the speleologists because they are regular, they are not like the usual heroes that we see movies in the mountain and like that the way they deal with fears.

Q-2:
The film is in its format and style a documentary, however, the choice of medium (technology) opens the possibility of challenging linearity and reconsidering the experience of the audience. Could you comment on the choice and impact of the technology in the storytelling and in your exploration of interactivity and immersion? Did you want the viewers to feel the mediated storyline of Caves, or did you want them to feel present in the setting, being "there"? In short, what was the balance you strived for: story vs. presence, immersion vs. storytelling?
A-2:
I wanted to make a film about caving since few years. I was taking to Thais (the editor) about 2d film but since I have experience with 360, I thought this would be a perfect medium. Through that you can tell a story through presence and immersion, tell story about space and you can get close to people. About storytelling, it's quite simple because you are taking a "tour" inside the caves. Quite often I was losing orientation in 360, 3 degrees of freedom films and I wound it always problematic when there are too many 180-degree jumps, so I wanted to challenge that. And also, when you are in caves orientation is often a problem and people lose it, I wanted to bring the viewers and make them feel present with a group of speleologists, being part of that experience. That is why I chose this technology, and a cave, "space" was also a protagonist for me.
Q-3:
From the references I have gotten from you in the beginning of our collaboration, I interpreted the direction you wanted for the film's music as playing the role of a "character and the voice of the caves". I took this reference and attempted to apply it in the sound-design and music composition. Do you think we achieved this effect, and was it how you imagined it from the beginning (sound playing a quasi-character role in the storytelling in a sort of musical theater)? In this sense, one could say a strong interplay between image and sound was pre-designed, would you agree?

A-3: Yes, I agree. I was looking as you said so and as we talked many times, I thought the music as a character, as presentation of the caves of what we don't see. And yeah, it was more or less the way I imagined. So yes, you did that. It in the work it shows very good.

Q-4:

In the press release you write the following: "Thanks to 360° Experience it is possible to convey to a wide audience in both physical and emotional terms how it feels to crawl through narrow passages, to hang in dizzying shafts, to observe with amazement from very close quarters the fascinating halls hidden underground." An aspect highlighted here, which I would like for you to elaborate on, addresses the question of immersive-ness, placemaking or spatial experience you seek to convey in the film. How do you see the role of sound in achieving these conceptual goals (for example in enhancing the perception of physical realism)?

A-4: A I think it's quite important sound because most of this base perception comes from our ears, not from our eyes, this is something most of the people got confused but the space perception that the room perception comes from our eyes come from, from our eyes. So, I think that sound has as well really important role in this space perception. The perception of, of space and physical realities. Yeah. It's quite important. Actually, much more important, that would we see because also in the film is what we see is quite less. There are always just a few spots of light, and the rates come from the information come from our ears.

Q-5:

"In a cave, trapped in the darkness, the audience can only see places where there is light. This makes it possible to create precise narration in 360°spaces.". This is an inspiring strategy for storytelling within the selected medium, and we also had discussions about spatializing the musical parts and creating a "soundscape". What do you think is the resulting impact or contribution of the spatial sound for physical and temporal orientation and "navigation" within the space described in the story? What about its narrative counterpart (in conveying emotions and suggesting moods, etc.)?

A-5: Sound has got a huge impact and contribution because a through the sound you see the different places, so each park got like a kind of theme like where there are these when she's getting or were there these are all stalactites and stalagmites and the way you're telling them so with through the sound, it's then different. Each place could like each have its own sounds. So, it has contribution, so it also creates rhythm and so the music also changes in each room. So, it helps to make a differentiation in between rooms. And I think it helps the whole story organize the whole story.

Q-6:

There was a conflicting function for sound, where at times you were looking for a fidelity of the experience (for example compared to recordings) versus an intuitive and perhaps more suggestive (but incorrect) rendering of sound (for example, the scene of the "Ankunft Biwak", where discussed the "dryness" of the room and the low level of reflections in the cave). How would you describe in summary the role of sound in the film: more as a tool to create physical realism, or rather a narrative element knitted in the story, or some fluid in-between? Based on some viewer feedback, sometimes the "realistic" approach seems counterintuitive (people expect echo). Would you change your interpretation for these scenes (where perhaps the realism was more important for you) based on this feedback?

A-6: That's a funny one I think it's kind of some fluid in between it's also to create valleys sometimes it's like ah, no disappointing a with that, based on some viewer feedback. Sometimes the realistic approach seems counterintuitive - people expect a go will you change your interpretation for these scenes where perhaps the readings was more important for you based on this feedback? And sometimes so I think it's like flow is in between a getting physical realism so being kind of scientific but it's not a scientific feeling but sometimes hmm like as an example the sound in 2001 A Space Odyssey there's no sound without a doubt showing in the space. I think it's much more interesting like this and creates kind of tension and in the case of 2001 Space Odyssey. I don't care much about feedback from strangers. There's just sometimes a few places caves where you hear some echo and some reflection from your own sound but there's not much echo and so I think it's kind of fluid in between so it helps to narrate the story and it helps also to, to achieve the physical reality has to be in between and I don't really care if the people I know some people like there's no way no echo God no, there's no echo when I gave this idea. I would leave it this way. I like it.

Q-7:
In the film, you use a combination of shot styles that bring us in different proximities to the space and the characters. Could you describe your concept for utilizing wide shots versus close-ups to the protagonists (Lea in the water) and perspective (3rd person vs 1st person) and its connection to sound (realism vs suggestive elements)?

A-7: Because of the geometry of the camera in 360 a lot of people say that it doesn't work in 360 and decide on not to choose close –up perspectives, because it looks. But I don't think that is the case that is why we have decided on some close-ups and to play with that. Camera is often a character; first person perspective and it is with the groups of other explorers as a character. Sometimes is getting closer to then, like in caves while the expedition. And because those close-up shots mostly happen to Lea, and that's why we decided in the synopsis that the Lea will be the main Protagonist, because we get closer to her. And we were looking for that scenes where we could play around with the perspective of the camera and make it for in VR. But is always about the whole group of friends, we follow them without prioritizing. About sound, one of the scenes where the perspective and "deformed" face of the Character could potentially not work is "Leas in Water" where she holds the camera, but because of the music which creates that surreal and mystical level it transfers viewers into this other subjective dimension, this is what I really like.

Q-8:
There are some challenges and limitations presented in using VR Technology. Some of them manifest themselves throughout the production, post-production and even screening. Would you care to describe some of these issues which were of difficulty or interest to you, perhaps some limitations served also as a source of inspiration? How would you describe your overall experience with VR technology?
A-7: VR technology I think it's quite interesting because it's a kind of new technology even though it exists in the since the 50s 60s. A while looking into cinema history, the cinema to develop another structure and new grammatic took more than 30 or 40 years in film to accept jump cuts and, and the way we have it in film nowadays, so VR offers, it's like a new medium, so you can rethink it and it´s about trial and error. I like these things from VR, because it's new, and you can make many tests and try new things. And sometimes it works sometimes it doesn't work and challenges and limitation. Of course, in the Caves, humidity, dirt darkness, dust darkest. And also, while editing you need much stronger hardware even though I feel with an amateur camera GoPro, computer had to be strong. And one of the main problems he while showing VR is infrastructure. Because most of the places they don't know how, how it works and what does it mean. And sometimes it's much more like say side an event, they will show case some VR but it's not the main point, people think it's model, like an attraction that something sad, but I think it will change make sure that a few festivals there are not so many but there are a few that know what those things means. And yeah, it's always a pain -the infrastructure, as you see like you go to a festival and they've got only VR headsets, but not headphones. And, and those things and also another thing is that people are not used to consume VR so that's also funny, but sometimes they are overwhelmed by the experience, and they cannot watch more than two minutes or three minutes and they get lost. Not with my film, I have to say that it worked quite well many people said also for newcomers to the VR.

2. Sound-Protocol

00:00:20:17 (Laufen Wiese)	00:01:10:06 (Doline)	00:01:53:19 (Wurzeln)	00:02:23:18 (WasserLaufen)	00:03:18:20 (EngeWasser)
200725_001 Ambisonics ∞ Video	005_LE0000 *[?] Mono ∞ Video Lea Lavalier	200725_004 Ambisonics ∞ Video	SR006R *[?] Stereo ∞ Video	200724_002 Ambisonics Wasser
004_LE0000 Mono ∞ Video Lea Lavalier	200725_002 *[?] Ambisonics ∞ Video		SR006F *[?] Stereo ∞ Video	200708-T006 Mono Wasser
000_140729 Mono ∞ Video Diego Lavalier	001_140729 Mono ∞ Video Diego Lavalier		006_LE0000 *[?] Mono ∞ Video Lea Lavalier	004_LE0000 *[?] Lea Lavalier ****
200720-T001 Ambisonics Zicaden	000_140801 Mono ∞ Video Nora Lavalier		005_DI0000 *[?] Mono ∞ Video Diego Lavalier	003_DI0000 *[?] Diego Lavalier ****
	[?]		003_NO0000 *[?] Mono ∞ Video Nora Lavalier	002_NO0000 003_DI0000 Wasser
♫	♫	[?] Wasser flies bis in die nächste Bild! (Wir fliesen mit den Wasser)	[?] FOLEY Wasser Bewegung PVC (DI; NO; LEA)	[?] FOLEY Wasser Bewegung PVC
			[?]	

∞ Video = Synch mit Video
*[?]= Schnitt im spur
[?]= Offvoice
♫=Music
[?]=SoundDesign
TO DO
SPAZIALIESIEREN
FOLEY

00:03:53:03 (Siphon Lea)	00:04:51:24 (Picnic)	00:05:41:04 (Enge kriechen)	00:06:21:24 (Enge warten)	00:06:47:01(A bseilen Sacht)

002_LE0000 Mono ∞ Video Lea Lavalier	200725_010 Ambisonics ∞ Video		005_140801 *? Mono ∞ Video Nora Lavalier	007_LE0000 Mono ∞ Video Lea Lavalier
001_140731 Mono ∞ Video Nora Lavalier	012_LE0000 Mono ∞ Video Lea Lavalier	006_LE0000 *? Mono ∞ Video Lea Lavalier	010_LE0000 *? Mono ∞ Video Lea Lavalier	200725_006 Ambisonics Gleiche Ort
001_140728 Mono ∞ Video Diego Lavalier	007_140729 Mono ∞ Video Nora Lavalier	008_DI0000 *? Mono ∞ Video Diego Lavalier	006_140729 *? Mono ∞ Video Diego Lavalier	
? FOLEY Wasser Bewegung PVC	006_140801 Mono ∞ Video Diego Lavalier			
♫		♫ Wind!		♫
		? FOLEY Bewegung PVC (DI; NO; LEA) Carabiner	? FOLEY Bewegung PVC (DI; NO; LEA) Carabiner	? ????

∞ Video = Synch mit Video
*?= Schnitt im spur
?= Offvoice
♫=Music
?=SoundDesign
TO DO
SPAZIALIESIEREN
FOLEY

00:07:22:03 (Sacht)	00:07:57:12(A nkunftBiwak)	00:09:37:22 (Wasserhollen)	00:10:14:14 (Essen)	00:11:10:04 (Schlafzimmer)

005_DI0000 *? Mono Diego Lavalier	SR006F Stereo ∞ Video	SR007R Stereo ∞ Video	006_LE0000 Mono ∞ Video Lea Lavalier	014_LE0000 *? Mono ∞ Video Lea Lavalier
? Wasser Tropfen/ Schlafsäcke	006_NO0000 Mono ∞ Video Nora Lavalier	006_NO0000 *? Mono ∞ Video Nora Lavalier		
?	007_LE0000 Mono ∞ Video Lea Lavalier	007_LE0000 *? Mono ∞ Video Lea Lavalier		
	005_DI0000 Mono ∞ Video Diego Lavalier	007_DI0000 *? Mono ∞ Video Diego Lavalier		
		? Lea N / D Bewegung TicTac Sprengung	? ????	?
	♫	♫		

∞ Video = Synch mit Video
*?= Schnitt im spur
?= Offvoice
♫=Music
?=SoundDesign
TO DO
SPAZIALIESIEREN
FOLEY

00:14:59:21 (Picnic vor S)	00:15:39:17 (Steigen Saal T)	00:16:39:02 (Steigen L/D)	00:17:30:23 (Aussen)	
008_DI0000 *? Mono ∞ Video Diego Lavalier	1. 4 Mono Canalen Stereo ∞ Video		1. 4 Mono Canalen Stereo ∞ Video	
011_NO0000 *? Mono ∞ Video Nora Lavalier		002_LE0000 *? Mono ∞ Video Diego Lavalier	009_DI0000 *? Mono ∞ Video Diego Lavalier	

010_LE0000 *? Mono ∞ Video Lea Lavalier	009_LE0000 Mono (∞) Video Lea Lavalier	004_DI0000 *? Mono ∞ Video Lea Lavalier	012_NO0000 *? Mono ∞ Video Nora Lavalier	
			011_LE0000 *? Mono ∞ Video Lea Lavalier	
? Lea N / D Bewegung	? Ambi Lea	? ????	? Uhu AMBI Lea N / D Bewegung???	
		♫	♫	

∞ Video = Synch mit Video
*?= Schnitt im spur
?= Offvoice
♫=Music
?=SoundDesign
TO DO
SPAZIALIESIEREN
FOLEY