

The Potential of Audio Augmented Reality for Journalistic Narratives

Masterthesis in Sound Design
By Demian Wyssmann

Zurich University of the Arts
Mentor: Daniel Hug
August 2026

Demian Wyssmann, Sahlistrasse 46, 3012 Bern
demian.wyssmann@gmail.com
demianwy.ch

Abstract

This thesis examines the potential of audio augmented reality (AAR) for journalistic narratives. As smartphones are widespread and headphones with integrated sensors are becoming increasingly common, a growing number of people have access to devices that could support AAR applications. Drawing on immersive journalism, a visually dominated field that uses virtual and augmented reality to create first-person experiences of journalistic subjects, this thesis asks how journalism and AAR can be connected.

To address this question, the thesis maps related fields of research, including immersive journalism, spatial audio journalism, AAR, immersion, and AAR-specific narrative strategies. Based on this theoretical background, it develops an analytical framework that is applied to three case studies that include the exhibition *Sounds of Silence* and the audio walks *Only Expansion* and *Consent – Walk the Walk*. These examples represent different approaches to AAR and allow a comparison of the technical implementation, narrative design, immersive potential, envisioned roles for AAR applications from the participants' perspective, and journalistic relevance.

The analysis concludes that the potential of AAR for journalistic narratives depends on the relation between the technical system and the subject of the work as the technology must support the intended form of knowledge transfer and narrative engagement. The case studies further show that immersion in this field of application should be understood as a perceptual and psychological state shaped by narrative components. Finally, the thesis argues that AAR requires a high degree of transparency in journalistic contexts. Because spatialized, locative, and interactive audio can make mediated content feel immediate and personally experienced, sources, reconstructions, work processes, and design decisions need to be communicated clearly.

Acknowledgement

At this point I would like to thank the people who supported me and without whom this thesis would not exist.

Dr. phil. Daniel Hug	for your support and advice throughout the entire process this thesis
Cedric Spindler	for your inspiring ideas, trust and constant fascination with the research topic
Prof. Felix Baumann	for listening and opening ears
Manuel Gerber	for guiding me through my time as a student
Jascha Dormann	for sharing your insights and observations
Rupert Jaud	for sharing your experience
My fellow students	for being a community
All teachers	for your knowledge
My friends	for their time
My family	for their love
<3	for everything, everywhere, always

Contents

1	Introduction.....	5
2	Background and related work	7
2.1	Immersive Journalism.....	7
2.1.1	Presence and Immersion	8
2.1.2	Journalistic narratives in Immersive Journalism.....	11
2.1.3	Spatial Audio Journalism	15
2.1.4	Summary	18
2.2	Audio Augmented Reality	19
2.2.1	Characteristics of AAR	20
2.2.2	Technical Components	22
2.2.3	AAR specific Narrative Potential.....	23
2.2.4	Envisioned Roles for AAR Applications.....	26
2.2.5	Summary	29
3	Case studies.....	30
3.1	Method.....	30
3.2	Sounds of Silence.....	31
3.2.1	Sources	32
3.2.2	Technical Implementation.....	33
3.2.3	Narrative Design	34
3.2.4	Immersive Experience.....	35
3.2.5	Envisioned Roles.....	37
3.2.6	Journalistic Assessment.....	38
3.2.7	Summary	38
3.3	Only Expansion.....	40
3.3.1	Sources	41
3.3.2	Technical Implementation	41
3.3.3	Narrative Design	42
3.3.4	Immersive Experience.....	45
3.3.5	Envisioned Roles.....	46
3.3.6	Journalistic Assessment.....	47
3.3.7	Summary	47
3.4	Consent – Walk the Walk.....	49
3.4.1	Sources	50
3.4.2	Technical Implementation.....	50
3.4.3	Narrative Design	51
3.4.4	Immersive Experience.....	52
3.4.5	Envisioned Roles.....	52
3.4.6	Journalistic Assessment.....	53
3.4.7	Summary	54
3.5	Comparison and Discussion.....	56
3.5.1	Technical Implementation.....	56
3.5.2	Narrative design	57
3.5.3	Immersive Experience.....	58
3.5.4	Envisioned Roles.....	59
3.5.5	Journalistic Assessment.....	60
3.6	Summary	62
4	Conclusion	63
5	Bibliography	65
6	Catalogue of Works.....	70
7	List of Figures	71

1 Introduction

This thesis examines the potential of audio augmented reality (AAR) for journalistic narratives. It focuses on the question of what narrative possibilities arise when journalistic storytelling is not conveyed solely through language, images, or linear sound, but is embedded in an acoustic environment that can be experienced spatially. Throughout this thesis the term journalism does not only include news reports but refers factual narrative content broadly, this also includes documentaries, features, interactive environments and in general stories grounded in facts while retaining a narrative form that helps audiences understand events. Applications that can be described as AAR – much like other forms of augmented reality (AR) – combine the physical environment with a virtual one, thereby creating a hybrid space. Unlike visual AR applications, however, the focus here is on the superimposition of real and virtual acoustic environments. This can give rise to acoustic scenes in which it is not possible to clearly distinguish between virtual and physical sound sources.

This connection is particularly relevant for journalistic content because it allows sounds to be anchored in space. The is thus not merely narrated but can be tied to specific locations, movements, or spatial perspectives. The audience then no longer engages with the subject matter from a detached perspective but within the space of the subject matter itself, which simultaneously exists within the audience's own space. This opens narrative possibilities that go beyond traditional linear audio formats: creators can direct attention within a space, make proximity and distance audible, or tie historical, social, and political contexts back to specific locations. However, these possibilities are not neutral. Precisely because journalistic works establish a connection to real events, people, and places, it is necessary to examine how the application in question structures the audience's perception and what role the design of the acoustic environment plays in this process.

This thesis thus ties into the discourse on immersive journalism (IJ). IJ describes journalistic approaches that utilize technologies such as virtual reality (VR), AR or spatial audio and aim to provide the audience with a first-person experience. Compared to established journalistic forms, which are often characterized by a distance between the subject matter, the audience, and the author, IJ shifts this distribution of roles. The audience is not merely addressed as an observer but can find itself within a constructed media situation. This shift opens new forms of communication but also raises questions regarding perspective, staging, authenticity, and responsibility.

Journalistic content is increasingly being consumed via mobile devices, headphones, podcasts, short videos, and platform-based services. At the same time, headphones and mobile devices are being enhanced with features such as noise cancellation, head tracking, motion sensors, and spatial audio playback. This creates the technical conditions that could make location-based and spatially designed audio content accessible to a broader audience. AAR should therefore be understood not only as a technical extension of existing audio formats, but also as a potential journalistic narrative tool whose aesthetic and narrative potential could be more closely defined. Changes in media usage and technical infrastructure make this field of research relevant for those involved in media studies, journalism, human-computer interaction, this thesis explores electroacoustic

composition and sound design. Although the topics under consideration are relevant to both participants and creators and deserve equal attention, this thesis largely examines the field from the creators' perspective. Since no interviews or surveys were conducted with participants as part of this thesis, conclusions about effects can often be drawn only through assumptions based on related fields of research, the literature, and statements by creators.

Although my preference for immersing myself in an alternative world, escaping my physical surroundings and losing track of time in the process, is not part of this thesis, it is likely one of the reasons why this thesis exists. This tendency to digress manifests itself, among other things, in excessive media use, which may well have given rise to the interest and subject matter of this thesis: to understand how media can be used and how they shape our everyday lives, our perceptions, and our access to information. These questions accompany me through my various activities as a sound engineer, composer, and sound designer, fields of work that all take place at the intersection of technology and temporal art forms. In addition to my interest in digressing, these fields of work and my education in Sound Arts at the Bern University of the Arts and Sound Design at the Zurich University of the Arts have shaped this thesis.

Based on this background, I examine how AAR can be used for journalistic storytelling. I will first explore the theoretical foundations of IJ, immersion, and spatial audio journalism. I will then take a closer look at AAR and its associated narrative possibilities, relating them to the roles of AAR application areas outlined by Tran et al. Based on this, I will analyze three case studies to demonstrate how AAR is used in specific journalistic or documentary works and to identify the technical implementations, narrative strategies, immersive effects, and application areas that emerge in the process. The goal is not to describe AAR as a fundamentally superior journalistic form, but rather to highlight its specific possibilities and limitations in the context of journalistic communication.

2 Background and related work

2.1 Immersive Journalism

Since AAR is situated in the field of immersive media, this first chapter will give an overview of its connection to the research and discourse of immersion and IJ. As we will see, IJ is still in an emerging stage which is why “immersion” and terms in relation to it are not used uniformly in literature (Uskali et al., 2021). Instead of using one definition, I will attempt to explain the factors of immersion that are most important for the analysis of factual storytelling in AAR. In this thesis IJ is used to describe a specific journalistic genre and field of work and not journalistic works with *immersive potential* (explained in Section 2.1.1.2) in general.

IJ is a format for factual storytelling that uses immersive systems to create a closer connection between story and participant. An immersive system in this context describes technologies with high sensory information density, such as VR glasses (Agrawal et al., 2020). The goals of IJ further include increasing empathy with the journalistic topic, creating a sense of *presence* for the participants, fostering a deeper understanding than traditional journalism, and promoting transparency by bringing the audience closer to the real event (De La Peña et al., 2010). Discussions on IJ lead back to Nonny de la Peña, also referred to as “The Godmother of Virtual Reality” (Volpe, 2015), who introduced the term IJ in her paper “Immersive Journalism: Immersive Virtual Reality for the First-Person Experience of News” in 2010. According to De la Peña, IJ describes news stories “in which people can gain first-person experiences of the events or situation described” (De La Peña et al., 2010, p. 291). In De La Peña’s definition, IJ represents a specific form of interactive journalism where the participants’ surrounding converts into the surrounding of the narrative. While interactive journalism already allows participants to get an experience of the situation and gain access to a first-person view, deep IJ can facilitate a first-person experience.

Interactive and immersive news stories can take various forms with the most dominant application being 360°-videos (Gabo & Milk, 2015). One example is collection of videos from the pilot project Lookout360° conducted by the European Forest Institute and the Global Editors Network that uses 360°-videos to investigate climate change¹. The prominence of 360°-videos over other forms of immersive media can be related to a production process which can be similar to linear documentaries and the fact that the audience can also consume 360°-videos on smartphone or computer screens if they do not possess VR glasses, which makes them far more accessible than other IJ applications (Uskali et al., 2021). While 360° videos are the most present application, other forms which require a more specific production process and cannot be consumed without VR devices exist and become more widespread with the audience’s access to the appropriate devices.

¹ <https://thelookoutstation.com/projects/lookout360deg-6-month-climate-change-immersive-story-accelerator>

AR or VR environments like the interactive environment *6x9* developed by The Guardian, which transports the participants to a prison cell² or *MLK: Now Is The Time* by TIME Studios that creates a virtual environment around Marthin Luther King Jr.'s "I Have A Dream" speech³. Another example for IJ is the VR documentary "Rebuilding Notre Dame" produced by "TARGO" in partnership with Facebook's Oculus⁴. These examples are applications of journalism that use immersive systems to create works that allow the audience to engage with the subject in multiple ways instead of being limited to steering the orientation of the camera. Serious games, a game genre that focuses on interactive education are another version of *interactive immersive journalism*. As an example, for serious games *Darfur is Dying* designed by TAKE ACTION games in 2006 is a frequently referenced⁵. There are also examples of data physicalization and similar forms of tangible representation of data which can be regarded as IJ by themselves or reinforce important factors of immersion like plausibility or *presence* (Perovich, n.d.). These genres and associated examples already show that the discourse on IJ and the related body of work is visually dominated (Castillo-Ramón et al., 2023; Sanchez-Acedo et al., 2023). If sound is addressed specifically, the discussion is mostly focused on the technical audio format rather than its potential to transport information (Rodero et al., 2015).

2.1.1 Presence and Immersion

To create this perspective immersive journalism harnesses the concept of *presence*, a term that describes the psychological factors that can create the illusion of "being present in an artificial environment", and combines them with immersive technology. To describe immersion only as an objective property of a system would ignore non-technical that can lead to immersion and that immersion is a temporal experience rather than a technical feature. As it is not possible to induce a psychological state but a system or environment can include factors that make immersion more likely, this thesis follows the concept of *immersive potential* and *immersive tendency* and links these factors to the concept of *presence*.

2.1.1.1 Presence

To enable this conversion of the surrounding, the participant must feel "as if they are there." There means the artificial space of the story world. In addition to being there, the participant must believe that the experience is really happening and that the artificial body or body part belongs to their own. These three perceptual illusions can be combined in the term *presence* and were described by Mel Slater as place illusion, plausibility, and body ownership. Place illusion refers to the feeling of being located in the mediated environment. Plausibility describes the impression that the events taking place are really happening, even when the participant knows that the environment is mediated. Body ownership refers to the extent to which a virtual or mediated body, or body part, is experienced as belonging to oneself (Slater, 2009).

² <https://www.theguardian.com/world/ng-interactive/2016/apr/27/6x9-a-virtual-experience-of-solitary-confinement>

³ <https://time.com/6246660/mlk-now-is-the-time-meta-quest/>

⁴ <https://about.fb.com/news/2020/01/rebuilding-notre-dame-vr-documentary/>

⁵ <https://games4sustainability.org/gamepedia/darfur-is-dying/>

While the degree to which *presence* is experienced depends on all described factors, its measurement stays inconsistent (Slater et al., 2022). If participants experience the effect of *presence*, they can start to respond realistically to the artificial environment even if they are aware of its artificiality.

Regarding technology, not only the specific features are important, but the successful display of sensorimotor contingencies, similar to those people rely on in the physical world, is a crucial factor in the human-computer interaction to evoke place illusion (Slater, 2009). *Presence* can support the aim of bringing the audience closer to a journalistic situation and is a driving component to create the first-person experience. However, *presence* should not be understood as identical with immersion. *Presence* describes a specific perceptual relation to a mediated environment, while immersion can also occur through, emotional involvement, strategic attention or engagement. A participant may therefore be deeply immersed in a story without experiencing a strong illusion of spatial presence and vis versa a participant can feel present in a virtual environment without being mentally involved (Jennett et al., 2008)

2.1.1.2 Immersive Potential

When leaving the VR space and trying to broaden the application of immersive parameters the technical factors to enable immersion become less of an importance. On the other hand, immersion as a psychological state needs further clarification to be used as a tool in journalistic narratives. Agrawal et al. argues that immersion is no objective measurable entity but an individual psychological state (Agrawal et al., 2020).

“Immersion is a phenomenon experienced by an individual when they are in a state of deep mental involvement in which their cognitive processes (with or without sensory stimulation) cause a shift in their attentional state such that one may experience disassociation from the awareness of the physical world.” (Agrawal et al., 2020, p. 6)

According to this definition, immersion can happen without external stimulation, e.g. daydreaming, and can also happen while still being situated in the real environment. If immersion is defined as an individual psychological state, it is impossible to create immersion itself; rather, it is only possible to create an environment in which immersion can occur. Immersion itself is a temporal occurrence with factors like a participant’s personal preferences and individual *immersive tendencies*, meaning “an individual’s predisposition to experience immersion.” (Agrawal et al., 2020, p. 6). This individual’s *immersive tendency* cannot be controlled.

In addition to the *immersive tendency* Agrawal et al. described multiple factors that support the possibility to experience immersion and summarized them as parts that characterize the *immersive potential* of a system or content that consists of the following sections (Agrawal et al., 2020).

Perceptual immersion

This type of immersion refers to subjective sense of being surrounded or experiencing multi-sensory stimulation that can create perceptual immersion refers to the feeling of being surrounded by the mediated environment and is an important contributor to the *presence* effect. The effect can be supported by large screens, spatial audio, head-mounted displays, 360-degree visuals, or the blocking out of the physical environment. Agrawal et al. argue that this type should not be confused with immersion as a whole; sensory stimulation can support psychological immersion, but it does not guarantee it. If the story lacks context, agency, or ethical clarity, the experience may remain technically impressive rather than deeply immersive (Uskali et al., 2021)

Narrative immersion

Narrative immersion can be described as the absorption in the narrative or the depiction of the narrative and describes the effect that occurs when participants become mentally absorbed in the story world, its events, characters, or emotional development. Agrawal et al. describe this as one of the main routes into psychological immersion, especially when participants become curious about what happens next, emotionally invested in characters, or absorbed by the fictional or factual world being presented. In IJ, this connects directly to the idea of first-person storytelling.

Strategic immersion

There are two types of immersion that connect to challenge-based absorption. *Strategic immersion* refers to absorption that emerges when participants are absorbed in planning and decision-making.

Tactical immersion

This type of challenge-based immersion describes the absorption through focusing intensely on quick reactions. It can be observed mainly in games when participants become immersed when they must plan, react quickly, solve problems, or balance their skills against the difficulty of the task.

Systematic immersion

This section is a counterpart to *strategic and tactical immersion* as it argues that participants don't have to be challenged to encounter immersion. It refers to the condition that participants accept the rules or logic of a mediated world even when these differ from real-world rules. In IJ, systematic immersion plays an important role for the participants' experience as they may accept reconstructed environments as meaningful journalistic representations but it also raises ethical questions about transparency, accuracy, and whether the audience understands what has been reconstructed (Uskali et al., 2021).

System immersion

This section is not part of the *immersive potential* of a system or environment to elicit immersion but refers to the idea that immersion can be understood as an objective property of a system or technology. From this perspective, a system is more “immersive” when it offers higher sensory fidelity, more channels, better tracking, or more realistic reproduction. In the context of IJ for example the higher information density of 360°- cameras in contrast to 2D-cameras allows the participants to look in all directions from one point instead of being fixed to the perspective which the authors selected. This circumstance can be described as a way to hide the technology used, meaning that it is neither present in the picture nor its effect on the display of the subject can be perceived (Uskali et al., 2021). Agrawal et al. criticize the concept of *system immersion* because it reduces immersion to technical specifications and ignores content, context, perception, and individual differences. Much of the research in the field of IJ has tended to prioritize technology, often at the expense of other elements, particularly the narrative.(Domínguez, 2017).

These effects multiply the factors that influence an immersive experience and broadens the understanding of what an immersive experience can be. To understand the role of immersion regarding factual storytelling in IJ, the next chapter gives an overview over three most common narrative methods in IJ and explores how the concepts of *presence* and *immersive potential* described by Sarvesh Agrawal et al. and Mel Slater differs between these methods.

2.1.2 Journalistic narratives in Immersive Journalism

To put immersion in context with journalistic narratives this chapter explores the three most common narration methods for 360°-news videos defined by Uskali et al. which are reporter-led narration, source-led narration and invisible/neutral narration (Uskali et al., 2021). Further, these methods’ influence on the goals of IJ, creating a closer connection between story and participants, by increasing empathy with the journalistic topic, creating a sense of *presence*, fostering a deeper understanding than traditional journalism and promoting transparency by bringing the audience closer to the real event. The three categories of narrative methods all focus on spoken narration and ignore other narrative elements that interact with immersion. The most prominent non spoken narrative elements are discussed separately.

2.1.2.1 Reporter-led narration

In reporter-led narration “The reporter is present either as a voiceover or visible in the footage.” (Uskali et al., 2021, p. 15). The reporters presence as a guide creates a constant point of reference that leads through the experience. Through this guidance the focus on the story’s development is enhanced which is an important factor to enable *narrative immersion* (Agrawal et al., 2020). Although this effect is stronger when the narration happens as a voiceover, such a narrator outside of the story may affect the *presence* effect, since the perceived space in strong voice-over is not plausible anymore and diminish the emotional connection and “the feeling of being there” (Cummings & Bailenson, 2016). Another claim of 360° storytelling that is not used in this narrative method is, that immersive journalism enhance the transparency between real event and journalistic

narration. In opposition to traditional media, immersive technologies should minimize the gap in information richness between the actual event and the journalistic story that occurs through mediation by reporting without mediation. In this approach the author is as a middleman gets less important, the richness in information of the real event is increased which should improve realness. Reporter led narration again opens the gap between the report and the real-time event which should be diminished through IJ. When going beyond 360° News videos, the gap between report and real-time event described, could again be diminished, through the direct engagement between characters and participants (Kool, 2016).

Regardless of whether the author serves as the narrator through voice-over or as a visually present character in the story, the author inevitably influences the impression of objectivity. (Aitamurto, 2019). The participants are clearly aware of the authors presence and their power over the creation of the content, which contradicts with the norms of objectivity established in traditional journalism. While in traditional journalism the erasure of the journalist on the audiovisual domain may be an option, this strategy opens the issue of transparency. As Deborah G. Johnson points out this erasure hides the constructed form of the experience (Uskali et al., 2021) While we might be aware of the authors influence in established forms of media, we can get tricked in emerging, unfamiliar technologies. We see this problem exemplified in news games. While games are typically associated with fictional, entertaining content, news games offer us fact-based experiences. If we do not know that the game we are playing is a news game, we may quickly assume that the content is fictional. On the other hand the assumption that the content of a news game depicts unaltered events can overlook the fact that it is a virtual environment constructed by the authors. In order for participants to properly understand the content of a medium, they must be familiar with the form and genre in which it is presented (Uskali et al., 2021). Transparency and objectivity in IJ are fundamental discussions and influence source-led and neutral-/invisible narration.

2.1.2.2 Source-led narration

In source-led narration a character who is present in the story or the story space narrates the story while the reporter is visually and audibly absent. While this narration method can retain the direct guidance through the story as seen in reporter-led narration, source-led narration can provide the possibility to focus on the characters perception of events, describe their emotional and other forms of invisible reaction and can make it easier to feel an emotional connection to the report which are elements related to the effect of *presence* and experiencing *perceptual immersion* (Pillai, n.d.). If the narration happens as a voice-over participants can remain the sense of being surrounded by the experience that follows the sense of body ownership and first-person view as long as it is the characters body the participants inhabit that is narrating and they interpret the voice over as an inner monologue (De La Peña et al., 2010; Slater, 2009).

While this higher interactivity and engagement increases the information density and therefore facilitates a better understanding of the events, it requires the author to construct the experience instead of documenting an event, which again leads us to the beforehand established problems of transparency and objectivity (Aitamurto,

2019). When working with recorded and filmed source material for IJ productions, being absent as a reporter means leaving the room, since one cannot hide behind a camera when shooting in 360°. Being absent as a producer during video-shoot is described as one of the biggest disadvantages by journalists who contributed to “Lookout360°, a 6-month Climate Change Immersive Story Accelerator”.

“In my experience, as a journalist who tries to build a strong connection with my story subject during interviews and filming, I found that leaving the room to hide outside of the 360 shots was one of the biggest drawbacks. I felt that I was losing the eye contact, and especially the connection with my subject.” (Chainon & Chainon, 2018, p. 8)

2.1.2.3 Invisible / neutral narration

This narration method lets the participants see and hear the content with very few interruptions (Uskali et al., 2021). Its focus lays on spatial immersion and the sense of being surrounded. Environments without direct narration force participants to focus only on the environment which can increase the effect of *presence* and enable a different understanding of the subject than traditional journalistic narration. Without explicit narration the participant must discover the environment on their own choose which part of the environment to focus on and interpret it themselves. While participants gain autonomy through this method, it minimizes the possibility of guidance. Participants might concentrate on a part of the experience and miss important elements which means that each participant will end up with an individual take of the story (Aitamurto, 2019).

The invisible narration enables the authors minimize mediation while maintaining high information density for the participants to engage with. This method is unique to immersive and interactive environments and can remove the author further than it is possible in other journalistic formats. However, it is also the method in which ethical questions regarding objectivity, work process and the role of the authors arise most urgently (Aitamurto, 2019; Kool, 2016).

2.1.2.4 Nonverbal narrative techniques

Based on the issues of transparency and objectivity, we have already seen that there are components apart from dialogue and spoken narration that influence the depiction of a subject.

One prominent element apart from spoken narration that influences the *immersive potential* in VR and AR environments is the structure of the artificial space. As these experiences take place in interactive spaces, these spaces have to be designed in ways which are feasible for the participants to read (Jang et al., 2023). How participants behave in virtual, interactive environments can be influenced not only by the narration but also by the design of the virtual space e.g. the way how objects are placed in a virtual room. If a space is hard to read and navigate, it can be difficult for the participants to understand the development of the story or plan and react according to the systems architecture. In 360°-video or 360°-photography, where the participants' position is fixed in space and it is not possible to move through, the participants' agency is limited which makes it easier for the authors to guide through the environment. In VR environments on the other hand, the narrative space offers multiple paths and becomes an active object of interaction, rather than just the background for interaction

(Jang et al., 2023). It therefore also becomes a part of the narration and influences the formation of sense or meaning (Van Der Nat et al., 2023). The deliberate design of artificial spaces to guide the participants' attention does not only influence the participants' experience but can also be used by the authors to develop the experience. One study for example explored, image rendering can be improved by lowering the render quality for parts of the environments which participants tend to overlook (Sitzmann et al., 2017).

All IJ environments require some technical system and each system comes with its specific set of technical specifications. If the technical implementation lacks certain capabilities it can influence the participants' experience. As soon as the latency between a participant's movement and the system's feedback raises, participants are much more likely to experience cyber-sickness and decrease the sense of body-ownership (Caserman et al., 2019). further factors that can influence the experience in a audiovisual environment can be the field of view, frame rate, the realism of the displayed images and the overall tracking accuracy of the system (Sanchez-Vives & Slater, 2005). The effect of sound on the *narrative immersion* is present throughout the experience of an audiovisual environment. Even if the attention of the participants is focused on visual or interactive elements, sound still remains relevant to make it tangible (Domínguez, 2017). The influence of sound on the perception will be discussed throughout this thesis. While the spatial information sound can carry seems to become more relevant in immersive settings to create a plausible environment, the specific technical audio-format is only one influence and can in some cases even be irrelevant for the effect of *plausibility*. In the case of binaural audio in feature journalism, a study by Sara Pérez-Seijo and Paulo-Nuno Vicente showed no relevant difference between monaural audio and binaural audio on participants' perceived immersion (Vicente & Pérez-Seijo, 2022).

Another common narration methods that influences the participants' behavior and influences the possibility to experience immersion is the interaction of the participants with the system and narrative components. While higher degrees of agency can increase *tactical and strategic immersion* the pool of choices challenge for the narrative design as it allows a "choose your own adventure" style of content, which can conflict with the goals of journalism. The multitude of possible paths enables the authors to create a multi-perspective representation of the topic with high information density, but risks a fragmented experience as each participant will end up with an individual version of the subject (Aitamurto, 2019). One possibility to mitigate these risks and offering guidance in VR-environments is the integration of audio beacons which is a core strategy in spatial audio journalism, a subgenre of IJ.

2.1.3 Spatial Audio Journalism

While IJ is visually dominated there is no necessity for a visual layer to pursue the core idea of gaining a first person experience of a journalistic topic (Paíno-Ambrosio & Rodríguez-Fidalgo, n.d.) Audio can already deliver high spatial or *narrative immersion* in mono or stereo and with binaural audio being a technical and financial easily accessible immersive audio tool, *presence* can be achieved with relatively low effort for developer and participants (Wincott et al., 2020). Spatial Audio techniques are widely used in cinema and games and since the technical possibilities advanced and got more accessible its field of application broadened. Spatial audio also referred to as 3D-audio or immersive audio, can be summarized as the recording and reproduction techniques that aim to synthesize a virtual scene which recreates a realistic sense of direction, distance and depth (Harju, 2025). As the evaluation of such scenes is based on the human perception, spatial audio is a tool which aims for perceptual accuracy and not physical accuracy (Hong et al., 2017).

Spatial Audio Journalism practice follows the development of binaural and immersive audio which took place over the last 150 years (Wincott et al., 2021). Starting with the laborious experiments of Alexander Graham Bell on binaural hearing and followed only few years later by Clément Ader who transmitted binaural audio between distant places in 1881 (Ouzounian, 2020; Paul, 2009). Immersive sound has since been used and developed in many disciplines including art, medicine, warfare and communication (Ouzounian, 2020). Its advantage in displaying spatial image is used in cinema since the 1990s and sound artists as well as soundscape composer. Broadcast radio and factual radio on the other hand, the journalistic precursors of spatial audio journalism, used spatial audio formats for a long time only in experimental contexts, even though binaural audio could have been transmitted via stereophonic FM (Wincott et al., 2021). Through online media and broadcasting, spatial audio now gains a wider use and with headphones replacing speakers, binaural audio is accessible to a lot of people. This development led to the creation of radio divisions dedicated to immersive formats like Nuov0son at Radio France and its connected webpage Hyperradio⁶ (Dussert, 2019).

The rise of online journalism not only increased the production of such contents but also the research in immersive audio journalism received more attention. BBC's R&D department conducted research over multiple years and was, together with Radio France, Bayerischer Rundfunk, Radio France International and other radio companies, part of the research project *Orpheus* which built a software system to deliver spatial audio to the audience between 2015 and 2018 (Wincott et al., 2021) The results of the collaboration can still be explored on the dedicated webpage⁷. As the project ended the apps support was terminate and can't be downloaded anymore.

⁶ <https://hyperradio.radiofrance.com/>

⁷ <https://www.orpheus-audio.eu/>

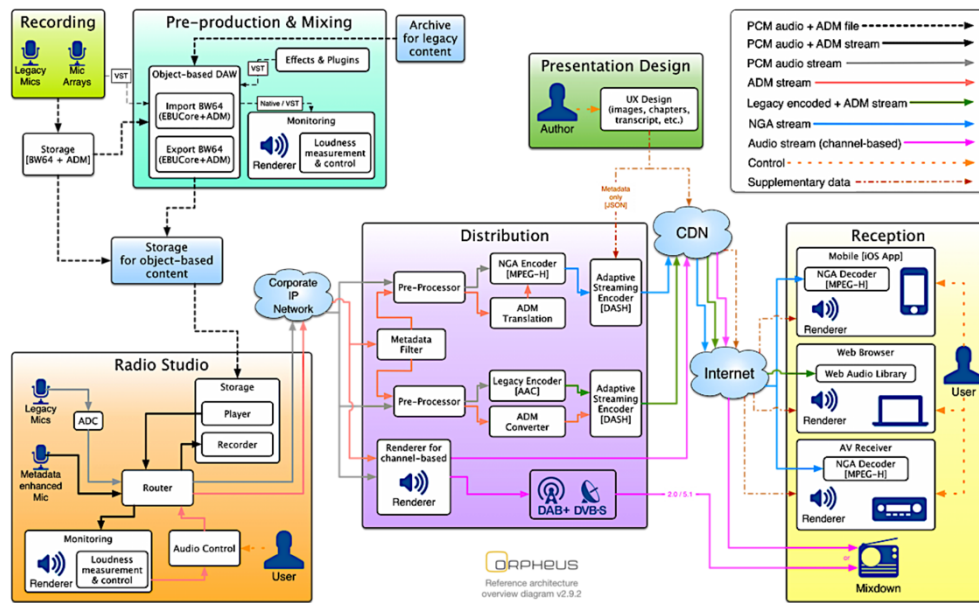


Figure 1: *Orpheus* Reference architecture Infrastructure

The core problem that *Orpheus* tried to solve is the lack of technical standards and consistency in spatial audio and related technology by developing a product with an end to end reference architecture (Figure 1) (Bleisteiner et al., 2018). The impact of this lack of standards gets obvious as soon as one tries to listen to spatial audio content on the webpages hyperradio.radiofrance.com and radiofrance.fr. If there is no instruction on the technical setup that is needed to play productions participants can never be certain if their technical setup corresponds with the selected production if the production. Since many playback devices give little information on the way how different audio-formats are treated until they reach the ears, participants can only assume that it is played correctly. Even channel-based audio formats with clearly labeled, technical names like 5.1, would demand further clarification on channel ordering for the participants to be certain and as soon as we want to play back files which are labeled as immersive, spatial or 3D the format is very unclear. Apart from a lack of technical standards this issue also comes from a non-standardized terminology which makes it hard to tell what exactly is happening in the audio file and how the listener can tune their system or headphones to get the best out of the experience (Wincott et al., 2020)

This problem grows when adding head-tracking. To create the illusion of soundscape that stays fixed when participants move their head the playback system needs to interpret and process the data received from the sensors inside the headphones. How this processing happens stays unclear. Apple's AirPods offer two setting which are referred to as spatial audio. The corresponding part of the manual does not explain how they define spatial audio. The effect of this setting is described as an "immersive, theater-like surround sound environment" (Apple Inc., 2026). By addressing all stages from recording to playback, *Orpheus* aligns a big part of the processing chain and aims for a system which is capable of a consistent participant experience regardless the playback environment.

In spatial audio not only the playback format needs clarification, but the story space design needs to be created in a way which allows the listener to navigate their attention easily toward the stories main components, so the listener can become a participant in shaping the story. *Orpheus* addresses this problem by enabling the participants to change individual layers of the audio content, length of a program and the possibility to display text transcripts (Figure 2).



Figure 2: *Orpheus* iPhone application

While *Orpheus* and spatial audio journalism take advantage of the possibility of spatial audio to create a high sense of *place illusion* and the wide use of headphones, there is yet no consistency in participant experience. In opposition to the step from a screen to spherical stereoscopic displays (e.g. head mounted devices) which marked a huge step in immersion in the visual domain, this is not the case for audio (Wincott et al., 2020). A study conducted by Sara Pérez-Sijo and Paulo-Nuno Vicente showed that interviewees didn't perceive a major change in immersion between monaural audio to binaural audio. Neither in *perceptual immersion* nor in *narrative immersion*. However, the study showed that the strength of binaural audio, the technically most accessible spatial audio format, lays in its possibility to create a scene which feels plausible, to enhance the identification between the listener and the character and the possibility to act as an evocative and central element even without an actual narrative (Vicente & Pérez-Seijo, 2022). These characteristics are important elements of narrative structures in AAR environments as AAR, the audio domain of general AR not only creates an artificial space but is in constant interaction with the physical space. The next chapter leaves IJ to introduce AAR and explores how it can create a plausible environment, emotional engagement, and evocative storytelling without spoken narration and finally how this connects to factual storytelling.

2.1.4 Summary

This section of the thesis provided background information on IJ, introduced the related concept of *presence*, and discussed different definitions of immersion. As the formats and technology used for IJ require specific narrative techniques, the most common narrative methods in 360°-videos were analyzed and discussed in relation to the goals of IJ and the journalistic notions of objectivity and transparency. The exploration of these narratives further showed that narration is not only driven by spoken narration but also by the design of the artificial space and the interaction possibilities of the participants with the technical system. The last part of this section introduced spatial audio journalism and its connection to IJ. By examining different applications of spatial audio journalism, the advantage of this format in creating a plausible environment became clear, and the problem of participants accessing works of spatial audio journalism due to a lack of technical standards was discussed.

2.2 Audio Augmented Reality

On the Reality–Virtuality continuum described by Paul Milgram and Fumio Kishino (Figure 3), AR is located along the line and describes a state in which the physical environment is still present but is layered with an artificial environment, unlike VR, which displays an environment that consists only of artificial objects (Milgram & Kishino, 1994). In visual applications, the most common domain for AR, the virtual layer is mapped onto a real-time video capture of the physical environment. Through the mapping, the virtual layer constantly adjusts to the movement of the camera and creates the impression that the virtual layer stays at the same point in space. The New York Times R&D department created a collection of such applications with a focus on educational content⁸. AR's history starts in the 1950s with the cinematographer Morton Heilig, who wanted to integrate the viewer in the onscreen action. With the introduction of the head mounted display by Ivan Sutherland in 1966 and two years later a first see through head mounted display AR got transported to the participants' head. The research took a further step when Azuma published the first major survey on AR in 1997, in which he offered a widely accepted definition describing AR as the combination of real and virtual elements in which the virtual elements are aligned within the three-dimensional space and capable of real-time interaction (Azuma, 1997). By now we know AR from a variety of systems like parking assistance, AR glasses, AR games, real-time video filters and smart city applications. With devices gaining constantly more computing power and the integration of large language models into AR software the development leads to AR systems with high adaptability and spatial awareness (David, 2025).

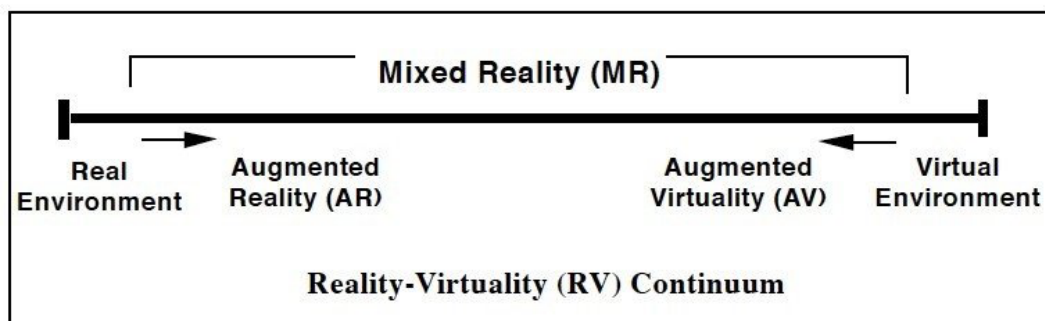


Figure 3: Reality – Virtuality Continuum by Milgram and Kishino

As in IJ, AR is visually dominated (Tran et al., 2026). AR applications that focus on hearing as the primary perceptual sense addressed are referred to as AAR. AAR is a smaller field of research introduced in 1995 by Benjamin Bederson, as he developed a prototype of an automated museum guide capable of position tracking (Arth et al., 2015; Harju, 2021). One advantage for Bederson of AR, in opposition to VR, was the fact that AR is designed to enhance the surrounding environment, rather than isolating the individual and placing them in a virtual environment (Bederson, 1995). While museum guides are still important ways to implement AAR applications, AAR's usage has spread over the last few decades. Some examples are the audio navigational

⁸ <https://rd.nytimes.com/projects/augmented-reality-storytelling/>

systems *Audio GPS*, which used non-speech spatial audio cues to guide participants (Holland et al., 2002), audio games such as the game *BREU: Shadow Hunt* by Team Zeroth, which is designed to be played without visual cues⁹, the immersive theatre experiences by DARKFIELD that use AAR to create a narrative environment for groups¹⁰, and the fictional and factual audio walks accessible through echoes.xyz, a platform for geo-locative audio walks founded by Josh Kopeček¹¹.

2.2.1 Characteristics of AAR

Over time, authors writing about AAR have used the term in different ways to describe closely related concepts. If we use the broad description that AAR includes all situations when something that is somehow “real” and something that is somehow “synthetic” are combined, AAR constantly happens. When we wait at a train station and hear an announcement through the speaker, when the fridge is opened for too long and the alarm goes off, when the phone rings, when an electric car passes, and, in general, every time some sort of speaker emits an audio signal, we usually do not refer to these situations as AAR (Azuma, 1997). To provide a possibility to distinguish between AAR applications and general audio applications, it is important to explore how AAR differs from other forms of AR, what characteristics can be attributed to AAR, and how they differ from other forms of AR. The different approaches show that using one specific definition of AAR could be hindering when thinking of new ways to use sound, but the discussion about possible descriptions and characteristics might help in finding new angles and a general understanding.

2.2.1.1 Taxonomy by Michael Krzyzaniak

That AAR refers to many different applications which are not always clearly definable as AAR was also one observation of Krzyzaniak et al. who examined how the term AAR is used in literature and practice to propose a taxonomy of possible AAR types. This research used a participant-centered perspective and research through design approach with a focus on perception and experience without trying to define audio hardware or algorithms and described six categories characteristic for AAR (Krzyzaniak et al., 2019).

Enchanting silent physical objects with digital sound

Meaning that objects that do not normally produce meaningful sound are extended through auditory feedback. This can, for example, be conductive ink that turns paper into musical instruments, audio maps through which participants can hear how a location sounds, textiles with conductive fiber included to turn them into flexible instruments or loudspeakers, and sports equipment that helps to improve practice methods.

⁹ https://store.steampowered.com/app/2213890/BREU_Shadow_Hunt/

¹⁰ <https://www.darkfield.org/>

¹¹ <https://explore.echoes.xyz/>

Deliberate blending of acoustic and digital sound

A prominent example for this category is karaoke where the voice is merged with the digital playback file by playing the acoustic voice and digital playback through the same loudspeakers. Another example is sound reinforcement in concerts or other live audio setting where digital tools are used to amplify an acoustic signal without overpowering the acoustic source to achieve a mix of both. The category can also be part of site-specific installations where recorded or generated sound mixes with the existing soundscape.

Digital sound-objects placed in real 3d space

This category means that sounds are treated as if they were located at a specific point in the real environment. Examples of this are AR games like *Pokémon Go*, where some of the creatures have a specific sound attached to them, or geolocated sound projects that link sound objects to the participant's location¹².

Overlay of extra audio information onto the real world

This overlay is, for example, used in navigation systems that use audio beacons or language to navigate the participant instead of a screen, or in museum guides that provide audio information about the subject and can be used for cooking assistants or other conversational agents that provide context-relevant information during an activity.

Realtime digital modification of acoustic sounds

This is, for example, known from guitar pedals and augmented musical instruments where acoustic sound is digitally processed. Another application is in-ear translation, like the capabilities of Google Translate to synthesize the recorded speech of the original language in the new language. In opposition to guitar pedals and augmented musical instruments, this example does not include the original acoustic sound but is still part of the concept that analog sound is digitally mediated.

Telepresence / audio merging of remote and local locations

In this category fall telephones or other forms of communication where a distant place is connected to the present one, and sound walks that want to transport the participant to a different location than their current one, or telematic musical performances, where audio connects the participant's present environment with another location.

¹² <https://pokemongo.com/>

2.2.1.2 Description by Matias Harju

Harju used a narrower description of AAR, which is based on the definition of AR by Azuma that describes the three main characteristics of AR systems as *combining real and virtual*, *being interactive in real time*, and *being registered in 3D* (Azuma, 1997). By translating them to the audio domain, Harju concludes that the characteristic essential for an AAR experience is to react in real time to the participant's action (Harju, 2021). This can be achieved by changing the soundscape depending on the participant's location, rotating a soundscape in regard to head movement, or other forms of interaction. While in visual AR the system also has to combine real and computer-generated objects and has to be registered in 3D, Harju argues that AAR can also work when the artificial layer is not dependent on digital processing, for example in pure analog settings, and even if the audio is not registered in 3D, as long as the audio acts in relation to the physical three-dimensional space, for example by playing audio in regard to the participant's location.

2.2.2 Technical Components

As every AAR experience depends on and is influenced by technical components, it is also important to explore their influence on perceptual and narrative components of AAR environments. Therefore, this chapter discusses the technology surrounding AAR in terms of its connection to the characteristics of AAR.

The range of technology that can be used is as broad as the range of AAR applications. At the end of every AAR application, however, there is a device that makes sound perceptible to our brains. These are typically headphones or speakers. Loudspeakers are generally used less often in AAR because they are bound to a shared physical space. Headphone applications, on the other hand, are widely used in AAR applications as they allow creators to create a mobile, private, and participant-specific auditory layer. In recent years, headphones that include features which refer to the characteristics of AR systems described in 2.2.1.2 have become commercially widespread.

Combine real and virtual

Many major brands meanwhile offer headphones with an integrated microphone array to actively cancel out surrounding sounds in addition to the acoustic occlusion that exist through the headphone's ear-cups or earbuds. This active noise canceling is often used by VR oriented systems to avoid distractions caused by the physical environment. The microphone arrays can also work the other way around and bypass the acoustic occlusion by amplifying the sound captured by the microphones into the participants ears. This effect of acoustic transparency makes it possible to blend real and virtual sounds and combine real and virtual environments.

Apart from the most common headphone designs, which are in-ear, on-ear, and over-ear headphones, devices with little acoustic occlusion exist, which means that participants can still hear the physical environment while wearing the headphones without the need to amplify it. This passive acoustic transparency in headphones can, for example, be achieved by using models with open ear cups, directional speakers placed close to the ears, or bone conduction.

Interactive in real-time

Headphones with built-in head tracking are increasingly available on the market. When paired with a smartphone or another device that can access this sensor data and use it for digital processing, the head tracking can be used to create a system that is interactive in real time. In AAR, head tracking is often used in combination with location tracking through GPS, which is included in many commercially available digital devices. The combination of these two methods is the basis for an environment in which the participants can move freely while the audio adapts to it. This combination of tracking is referred to as six degrees of freedom.

Registered in 3D

As headphones are designed to deliver binaural audio, this is the primary audio format that is used to create audio that is registered in 3D. Binaural audio refers to files in a two-channel audio format, for the creation of which filters, so-called head-related transfer functions (HRTF's), were used to simulate human physiology. In an application in which the artificial sound sources stay aligned with their physical counterpart when the participant's position and head orientation change, binaural audio is an important component to create the illusion of a sound coming from the environment while it is actually played on the headphones and can therefore evoke plausible acoustic soundscapes.

These features offer opportunities for AAR without the need of multiple external sensors in commercially available products. With the widespread use of smartphones also the possibility to process the sensor data is commonly available. Through the improvements of large language models recognizing semantic patterns, audio applications are increasingly capable to recognize audio sources and adapt to them. Speech recognition which influences the noise cancelling of headphones is one current example for this development. If these options would be transferable to other audio sources than speech, this would offer the potential of selective acoustic transparency and adaptive noise control (McGill et al., 2020). Through acoustic transparency, head-tracking, location-tracking, binaural rendering, and adaptive processing, contemporary headphone systems already provide many of the technical features needed for AAR. For IJ, this is particularly relevant because headphone-based AAR can connect journalistic narratives to the participant's immediate surroundings without fully replacing the real world. At the time of this study, no such applications are known to exist. As discussed in Section 2.1.2 the narrative layer plays an important role in journalistic applications. In addition to the narrative methods discussed in the context of IJ, AAR applications propose distinct narrative possibilities which need further attention to allow an analysis of its potential for journalistic subjects.

2.2.3 AAR specific Narrative Potential

To better understand how AAR can influence journalistic content, this chapter analyzes the narrative design possibilities that Harju suggests in *Audio Augmented Reality: concepts, technologies and narratives* in relation to the main goals of IJ (Harju, 2025). As discussed in 2.1 these goals include enabling first-person experiences of the events or situations described, restoring empathy, and allowing a deeper understanding through mimetic experience. The analysis focuses on how AAR's specific characteristics, such as spatial sound, real-time

interaction, location-based narration, and the combination of physical and artificial sound layers, can support or complicate these journalistic aims. By doing so, the chapter shifts the discussion from the technical conditions of headphone-based AAR to its narrative potential for immersive journalistic storytelling.

Familiar environment as a narrative space

One possibility that AAR offers is to turn the participants' personal environment into a narrative space. This holds the opportunity to create strong emotional resonance because the story unfolds inside a place that already carries meaning for the participants. In this understanding AAR can make the environment function as more than a neutral backdrop as the surroundings become an active element of narration. Sound can then be designed to blend seamlessly with the environment, or they can be stylized and clearly distinguishable to signal their mediated nature. To use the *familiar environment as a narrative space* can support the-person framing central in IJ.

Personal and collective stories tied to a place

A core narrative possibility of AAR is the ability to attach stories to place through site-specific or geolocated audio structures (e.g., audio walks, museum narratives, oral history formats). Harju describes such practices as a way to embed narratives into the environment so that movement through the space becomes part of narrative progression. This place-based structure supports both personal stories like memories, testimonies and perspectives, and collective narratives like shared history, community identity and contested interpretations of a site. This can create a strong sense of proximity and can increase contextual comprehension by letting participants experience how a place feels sonically and socially.

Hidden narratives

Harju emphasizes AAR's ability to reveal layers of narrative that remain unobservable or unheard without augmentation. This includes hidden stories in a cultural or political sense, like suppressed perspectives and hidden events in a perceptual sense as for example an application that places sounds behind walls, voices in adjacent rooms and traces of earlier use of a space.

Technically and narratively, this category is closely linked to the distinction between acousmatic sounds and attached sounds. Acousmatic sound are sounds without a visible physical counterpart while attached sounds spatially align with an object. Harju treats these as key narrative building blocks because they determine whether the participant experiences a sound as a "ghostly layer" or as an augmentation that belongs to a specific object or location. For journalistic storytelling, *hidden narratives* can be described as a strategy that supports exploration and discovery: participants assemble meaning through movement, listening, and perspective shifts, which can support a more embodied understanding when complex topics are communicated through layered spatial evidence rather than linear exposition.

Collective and personal

AAR can be experienced socially in a shared physical space while still providing personal and individualized narrative layers through headphones or other wearable audio displays. Harju highlights that this shared place with personal feed structure makes it possible to deliver different perspectives, different languages, or different narrative fragments to different participants, even when they are co-present. This can be understood as a narrative design option that enables journalistic focalization strategies as one participant may receive one viewpoint or information layer, while another participant receives an alternative layer, making the multiplicity of perspectives experienceable in a spatial situation.

Augmented theatre

Harju connects AAR to immersive theatre and staged installations, where narration is delivered through spatialized voices, environmental sound design, and timed sequences unfolding inside the participants' real environment or a prepared venue. This category can include both site-specific staging and "reskinned" environments, where the physical space is treated as a controllable narrative container. In both cases, the narrative technique relies on spatial placement, synchronization of events, and directing attention through sound. *Augmented theatre*, when transferred to factual narratives, can be described as a way to stage testimony or situational ambience.

Bringing back extinct sounds and evoking the future

Harju also outlines temporal strategies where AAR overlays the present with soundscapes from the past as for example historical sounds or with speculative future scenarios.

In descriptive relation to the goals of IJ temporal overlays can support first-person proximity by letting participants hear "what was" or "what might be" in the same space they physically occupy.

Emergent narratives

Finally, Harju discusses AAR formats where narrative is not only delivered but can emerge through interaction, exploration, and sometimes collective participation. In such designs, the story is shaped by the participants' movement, attention, and decisions, forming a story world logic where perspective shifts and discovery matter. This opens the possibility for multiple or non-linear narratives in which the participants determine the development.

Taken together, these narrative design possibilities describe AAR as a medium where sound attaches meaning to place through strategies of spatial anchoring, layered revelation, performative staging, temporal overlays, and interaction-driven story formation. These strategies can be used to describe how AAR may support central aims of IJ.

2.2.4 Envisioned Roles for AAR Applications

Despite the growing distribution of AAR-capable technology, the available use cases for it are still limited. To broaden the field of application, it is not only important to summarize the history and existing areas of application, but also to outline future applications and evaluate their significance from the participants' perspective. This chapter therefore provides a collection of possible applications for AAR environments based on the study conducted by Tran et al. as part of the paper "Envisioning Audio Augmented Reality in Everyday Life" (Tran et al., 2026). The authors conducted a collective auto-ethnography study and an online survey to evaluate: "How do people imagine AAR supporting or reshaping their everyday auditory experiences?" (Tran et al., 2026, p. 2) And "What boundaries or concerns do people identify when imagining AAR in everyday life?" (Tran et al., 2026, p. 2). Even though the study is aimed at AAR applications for everyday life, which is a very specific environment and ignores many other settings, these questions describe relevant components for AAR applications for journalistic narratives. In everyday life, participants need technology to work in the background and be implemented into the participants' environment without disruption. When working with journalistic subjects, it can be useful for creators to know if participants can relate to the purpose of the technical setup to adjust to it more easily and focus their attention on the subject. In this regard, the first research question of this study addresses the approach of IJ to make technology invisible to heighten transparency between the participants and the subject (Section 2.1.2).

The study further allows for the evaluation of concepts regarding AAR environments from the participants' perspective. The second research question takes participants' concerns about AAR applications in everyday life situations into account and considers participants' preferred amount of control over the system. While these considerations might differ in settings other than everyday life, they are still included to enable a point of reference for further discussions.

In the collective autoethnography conducted, five researchers examined their experiences with everyday sound and imagined how the experiences could be transformed through AAR. The only survey, which ended with 74 completions, contained three questions on imagined roles of AAR and three questions on control preferences and concerns. The collective autoethnography revealed ten imagined roles for AAR, which were then also used to evaluate the open-ended survey responses and were identified with different frequency. The authors later concluded that the roles can be grouped into four categories that characterize four different general ways in which AAR was envisioned to enhance encounters with sound.

Task– and utility-oriented roles

This section includes the most and third most represented roles of the studies. It refers to AAR as a tool to navigate and provide environmental support. These include applications which are established and well covered by related research. The fact that they are even still prominent even when the participants were prompted to speculate more broadly suggest that they remain important application possibilities for AAR

Roles

- *Reduce* which relates to the potential that AAR could lower or mask unwanted sounds that disrupt focus, rest, or comfort. Participants saw this as useful for managing stress and distraction. This was the most prominent role in both methods used in the study
- *Enhance* describes the possibility that AAR could make important sounds clearer and easier to locate. Participants imagined that AAR could help hearing speech, voices, and subtle cues in noisy settings. This was the third highest response represented across both studies
- *Guide* imagines AAR that could provide prompts and spatial cues for safety and navigation. Participants imagined support in traffic, crossings, and home routines.

Emotional and social roles

This category demonstrates that AAR can also take on tasks that go beyond problem-solving and that AAR's ability to expand affective experiences can allow for a more embodied engagement with subjective topics. One application which approached this category was the research project Orpheus which was discussed in Section 2.1.3.

Roles

- *Personalize* which thinks of AAR applications that could adapt sound experiences to a person's mood, activity, location, or physical state. This role was the second highest ranking response overall.
- *Anticipate* describes AAR applications that could warn participants before sudden or important sounds occur. This could reduce startle while keeping people aware of key events.
- *Remember* describes AAR could use sound to evoke memories and emotional connections like recreating meaningful places or past experiences through audio.
- *Reflect* imagines how AAR could help participants understand how their own sounds are perceived by others and how it could therefore support social awareness in shared spaces.

Perceptual collaborator roles

This category refers to AAR applications that can not only enhance or reduce sounds but translating the sounds context and meaning to the participant and advising them on their action. In this way AAR could help better understanding and influence the way people engage with their surroundings.

Roles

- *Interpret* was envisioned as a role for AAR that could identify unclear or unfamiliar sounds and explain their meaning.
- *Simulate* imagines AAR that could add artificial sound layers to enrich daily life like for example playful, educational, or calming soundscapes.
- *Transform* thinks of how AAR could replace unpleasant sounds with more pleasant alternatives.

Rhythmic and embodied collaborator

The authors add this category without any specific role to add the potential of AAR as a system which “mediates how people live through and make sense of time, sound and social space” (Tran et al., 2026, p. 10). This approach assumes that AAR can operate on different time-scales from short cycles like body movements and sensations to medium-term cycles like activities and routines to long cycles like days and seasonal flow. In addition, this category also addresses how AAR does not only affect the auditory perception but also can affect the participants’ behavior. As an example, the paper mentions that a AAR guiding system like spatial cues can affect navigational pattern. Complex AAR applications that act in this embodied manner and link roles to their corresponding time cycles would then act as a collaborator that is integrated deeply into the everyday lives and could reshape participants’ engagement with sensory events.

The second research question focused on participants’ boundaries and concerns when imagining AAR in everyday life. The results of the study show that the biggest concerns were that the system might have implicit control over their behavior and the participants might lose personal agency. A further concern was expressed as the possible problem that the system never works quite right and introduces discomfort and possible fatigue. The survey further showed that participants preferred manual control or predefined preferences over full automation. This result aligns with the concern of losing agency and an application which does not integrate seamlessly into the participants’ environment. The collective autoethnography revealed additional tension that AAR applications face in everyday situations. For example, reducing unwanted sounds might support comfort, but could also weaken awareness of other people’s presence. Similarly, softening loud warning sounds might reduce stress, but still needs to preserve their safety function.

The categories that emerged from this study show that the most researched applications in *task- and utility-oriented roles* are still prominent expectations of how AAR environments can be useful in everyday life. The *emotional and social roles* also partly connect to the first category, as it helps participants to navigate emotional and social situations according to their preferences. The results of the second research question open the

discussion on how an AAR application can be adjusted to participants' preferences without interfering with participants' personal agency, while being adaptive enough to fulfill the purpose of the chosen role. The results further indicate that AAR applications need to balance augmentation with environmental and social awareness. For journalistic AAR environments, this provides a useful reference point, because it shows that the technology should support engagement with the subject without disruptive, controlling, or inaccurate implementation. The participants' acceptance of audio augmentation further depends on personal agency and transparency. The results suggest that AAR applications for journalistic subjects could suffer from blurred boundaries between augmentation and physical environment and a system without explicit interaction and control possibilities.

2.2.5 Summary

In conclusion, AAR can be understood as a form of augmentation in which the physical environment remains perceptible but is extended through artificial, spatial, and adaptive layers of sound rather than one specific technology or application. The discussion of different definitions shows that AAR should not be reduced to one specific technology or application but instead describes a range of practices in which real and synthetic sound elements interact with the participant's perception, movement, location, or activity. Regarding the technology surrounding AAR, headphones are especially relevant because they allow an individual, mobile, and spatially controlled auditory layer while the participant remains situated in the physical environment. Through acoustic transparency, head tracking, location tracking, binaural rendering, and adaptive processing, contemporary headphone systems already provide many of the technical conditions needed for AAR. For IJ, this is particularly relevant because headphone-based AAR can connect journalistic narratives to the participant's immediate surroundings without fully replacing the real world. Unlike VR, which transports the participants into a separate artificial environment, the specific characteristics of AAR not only affect the necessary technical features but also lead to a specific narrative design potential.

The last section focused on the envisioned roles of AAR applications identified by Tran et al. that worked out ten roles grouped in the four categories. These roles and the corresponding concerns show that AAR applications need to balance augmentation with participants' agency, privacy and social awareness. This section concludes the background analysis surrounding IJ and AAR which provides the foundational concepts, based on which three case studies are conducted.

3 Case studies

To explore how the concepts discussed in the background section of this thesis, this section uses the findings of the background analysis to examine three examples that represent different sections of journalistic AAR applications. *Consent – Walk the Walk* represents an explicit journalistic application of locative audio, while *Only Expansion* and *Sounds of Silence* are closer to artistic, documentary, and educational AAR practices. These case studies further connect the field of research to the field of practice and allow a link to future research and implementations of the concepts elaborated in this thesis.

3.1 Method

The mentioned works are included because they demonstrate technical, narrative, and experiential strategies relevant for journalistic AAR implementations. Due to the small existing body of work, no single example can guarantee comprehensive coverage of the analyzed categories, and the examples were therefore chosen to ensure this coverage as a sum. To allow a better understanding of the actual audio environment, all examples or a representation of the original work must be accessible at the time of writing. Furthermore, the examples must be comparable to current literature and technical methods, as this field of research is still in its early stages and is evolving rapidly. In the following chapters each work is analyzed individually according a common framework. The framework is based on the background analysis presented in chapter 2. It discusses the specific technical implementation, the narrative design, its potential as an immersive experience, its connection to the envisioned roles discussed in chapter 2.2.4 and evaluates its journalistic components. Each analysis further contains an overview over the used sources. The analyses are followed by a comparison and discussion as well as a summary. The analytical framework used to analyze the examples included five perspective and consisted of the following questions.

Technical Implementation

- What are the components of the technology surrounding the example?
- Does the technical system combine real and virtual, interact in real time and register sound in 3D?
- How does the technical implementation influence the participants' options for action and the interaction between real and virtual environment?

Narrative Design

- Which categories of AAR specific narrative strategies described by Matias Harju can be observed in the example?
- How do the AAR-specific narrative design choices influence the participants' engagement with the subject of the work?

Immersive Experience

- Which components of the example add to its *immersive potential*, and which perceptual and psychological concept of immersion can it be attached to?
- Which conclusions can be drawn from comparing the participants' experience with the example's potential to create immersion?

Envisioned Roles

- How does the example connect to categories of envisioned roles and for AAR by Tran et al.?
- What conclusions about the impact of these categories on the participants' behavior and experience can be drawn from the accessible literature and source material?
- Which components of the example connect to the concerns that the study exposed?

Journalistic Assessment

- How are the concepts of IJ addressed in the example?
- What technical, narrative and compositional components influence the participants' engagement with the topic and to what extent does this differ from an implementation without AAR?

3.2 Sounds of Silence

Sounds of Silence is an exhibition first shown at the Museum für Kommunikation in Bern, Switzerland, between November 2018 and July 2019. The exhibition was created by the audio design and sound scenography studio Idee und Klang together with radio journalist and author Bettina Mittelstrass and the spatial design studio ZMIK. The exhibition explores silence and noise and illuminates different facets of the acoustic, perceptual, and psychological phenomena as well as the historical context. Each station is designed individually and highlights a different facet.

The exhibition's visual layer is very reduced and graphic, leaving all content-relevant knowledge transfer to the audio domain (Figure 4). At the beginning of the exhibition, the participants get a set of headphones with included head-tracking sensor and a smartphone that runs the software to calculate head- and position-tracking and is responsible for the audio playback. The headphone and software system is developed by FRAMED. The exhibition's first two rooms form the introduction to the topic and the explanation of the interactive headphone system. A major part of the exhibition takes place in one big rectangular room only divided by transparent curtains and graphical markings on the walls and floor. The graphical markings and curtains represent the boundaries between the different stations, which are separated content-wise and acoustically. Further, the central room is divided into three parts. On one side, all stations are about silence; on the other side, all stations are about noise. Through the center of the room, all stations, represented as seats, explore the history of silence. From this center part of the exhibition, one can enter three smaller rooms with separate stations. One of them

is the last station before exiting the exhibition, while the other two are two small rooms partly breaking with the visual concept of the rest of the exhibition. This spatial setup was used by the creators to develop a modular exhibition with no explicit order. The participants are guided by a voice through the entire exhibition. The voice works on multiple narrative and communicational levels. It introduces us to the logic of the system, explores the content of the exhibition with us, interacts for us with characters inside the story, guides our expectations, and helps us interpret heard sounds and soundscapes. A detailed video walkthrough of the exhibition can be watched and heard on the webpage from the Museum für Kommunikation ¹³



Figure 4: Graphical layout of the exhibition *Sounds of Silence*

3.2.1 Sources

The analysis of *Sounds of Silence* is based on press releases and summaries of the exhibition as well as a documented video walkthrough, an interactive VR representation of the exhibition accessed on 27 May 2026 at the Museum für Kommunikation and a semi-structured interview with Jascha Dormann, one of the leading sound scenographers working on *Sounds of Silence* as part of Idee und Klang (De Marco et al., 2021; FRAMED, 2015; Kirn, 2018; Museum für Kommunikation, 2018; ZMIK, 2018)

The press releases provide a general overview from the perspectives of Idee und Klang, ZMIK, FRAMED and cdm. Both the video walkthrough and the VR representation are in depth documentations of the exhibition. The VR representation further allows to understand the interactive structure of the exhibition. The semi-structured interview, conducted on 7 May 2026, provides an in dept view from the creator's side about the work-process, narrative design decisions and observations of participants' interaction with the exhibition.

¹³ <https://www.mfk.ch/austauschen/digitaler-besuch/videorundgang-sounds-of-silence>

The work is part of this thesis because it illustrates the narrative and journalistic possibilities of a technical system that is usually limited to supporting visually dominated exhibitions as an audio guide. It combines this tradition of museum guides with the slow journalism genre of radio features and radio plays and creates an environment in which the visual domain is stripped down to symbols and references, leaving the documentary work and knowledge transfer to the audio

3.2.2 Technical Implementation

Sounds of Silence used a headphone-based playback system developed by FRAMED called Usomo to handle tracking and audio playback. Its goal is to create an individualized immersive spatial experience. It can track participants' exact location and head orientation, which allows sound designers to create an acoustic environment in which the participants influence the audio through their movement in real time. In the case of *Sounds of Silence*, the technology used five degrees of freedom, ignoring position tracking in the vertical axis. All audio was encoded to ambisonics to allow rotating the sound scene and, in the end, rendered to a binaural mix playable on headphones (FRAMED, 2018). In addition to the headphones, each participant is given a smartphone, which acts as a playback system and is connected through a cable to the headphones.

Idee und Klang used multiple methods of spatialization, which led to different interactions of the participants with their mixed-reality environment. Narrative parts consisting of mono sources were represented by a ball hanging in space. Participants could move toward and around the ball, locating its artificial acoustic properties exactly at the place where the physical ball is placed. At other stations, participants could turn their heads inside an acoustic scenery like the ice hockey stadium and John Cage's piece 4'33'', getting a detailed sound image able to represent the real location. Most stations consist of a mix of different approaches in the way it can be observed at minute 33:25-38.33 in the video walkthrough. While Sarah Maitland tells us the story of her experience in seclusion on the Isle of Skye the participants listen to the narrated story which is linear and not spatialized, the sonic scenery around the guide voice constantly changes, depending on the participants' location and head rotation inside the station. Due to the limited space and number of separate stations these changes in scenery were close together (Figure 5) and participants could produce shifts in the sonic scenery with only small body or head movements. According to Jascha Dormann's observation of participants' engagement, this circumstance was led to participants moving slowly and exploratorily through the exhibition, constantly observing how the sound environment changed.

The closed headphones used in *Sounds of Silence* limit the acoustic permeability of the mixed-reality environment. While the participants' physical movement and visual surroundings remain connected to the artificial audio layer, the physical soundscape is largely excluded. The mixed-reality relation therefore depends more on spatial correspondence, visual references, and bodily movement than on a direct acoustic blending of real and virtual sound. Jascha Dormann stated that against their worry that the headphones could lead to an isolated experience of the exhibition, people visiting the museum as a group started to communicate through gestures and facial expression.

Sounds of Silence used a technical setup focused on location and head tracking and the participants' possibility to interact through their movement. In the exhibition, the creators focused on creating a diverse and informative walk that demands a seamless integration of the technology. The technology was explained in a separate room to ensure that its logic is understood and can be used by the participants. The system's hardware mitigates a collective experience and diminishes an accurate perception of the physical acoustics. Its narrative layout does not address this issue directly, but through its explorative character, dynamic sonic changes, and visual references, it guides the participants' awareness not only to the audio content but also to their presence and movement in the physical space.

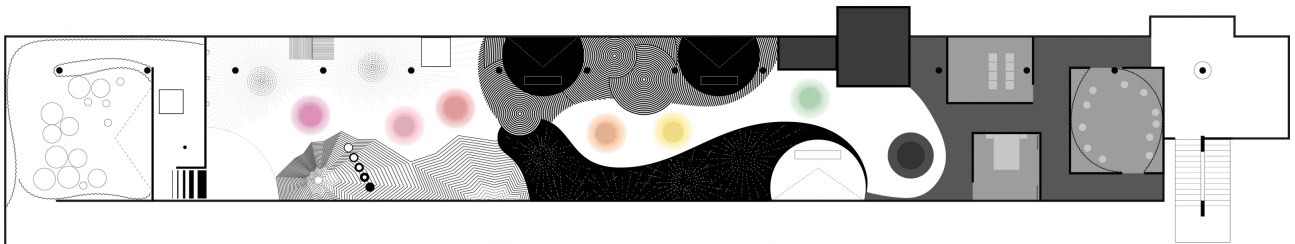


Figure 5: Top view of the exhibition space of *Sounds of Silence*

3.2.3 Narrative Design

The exhibition's use of audio augmented specific narrative strategies depends on the individual station. One element that can be observed throughout the exhibition is the strategy of *hidden narratives*. The station displaying a hotel room is a very explorative example of this. The participants can discover sounds inside an apparently silent environment. The AR environment allows to hide sources to give space to others and vis versa to focus on each noise which could be heard in this environment and interpreted as noise like a tap dripping, neighbors talking, sirens wailing, the air conditionings fan or a clock ticking. Depending on the participants position in space the hotel room gets set in a different context and gets different accessories. The principle of *hidden narratives* is also used in the central part of the exhibition. The central room and therefore a major part of the exhibition is visible from the moment the participant enter the station about the big bang, but they can't make sense of the rooms different parts yet. As there was no printed information that would help the participants anticipate the content, the exhibition needed to be discovered by the participants by physically moving through space.

The narrative strategy of *bringing back extinct sounds and evoking the future* is used in the historical overview that displays past sceneries. To use the narrative potential of collapsing a specific location at multiple points in time into one, the historical scenery could add connection to present time. This would however contradict the narrative design of the exhibition as it focuses mostly on an exhibition space that gets "reskinned" to the needs of the narrative. The visual layers minimalistic style is layered with an independent narrative. This technique allows the creators to be independent from the rooms specific location and context and makes it interchangeable if the dimensions are adapted to the position tracking software. This leads us to the strategy

of using the *familiar environment as a narrative space*. As the work is placed in a museum uniquely created for this exhibition the participants are not in a familiar space, but the soundscapes tight link to the participants movements still triggers a similar effect. The participants start to move slowly observing how familiar movement are part of the narrative space. This link can be connected to Section 2.2.4 and will be addressed in Section 3.2.5. When participants start to strongly interact through their movement, they can create links between parts of the exhibition and shape the narrative to a unique one only developed through this interaction which connects to the strategy *emergent narratives*.

By using a shared space but providing different personal audio depending on the location this example also connects to the strategy of *collective and personal* with the limitation, that the audio stream is the same for all participants when they are present at the same location. These narrative strategies influence the participants' engagement by transforming the exhibition topic into an exploratory process. Silence and noise are not only explained verbally but encountered through movement, discovery, and spatial comparison. Especially the use of *hidden narratives* encourages participants to listen actively and to search for acoustic events that would normally remain unnoticed. This supports an engagement with silence and noise as perceptual phenomena rather than as abstract concepts. The narrative design therefore connects the thesis topic to the participants' bodily and auditory experience.

The staged and scenes partly use the strategy *augmented theatre*. The spatial placement and telepresence of sources as well as the intimate interaction of the guide voice with participants depend on the interactive environment of AAR. these scenes are not used to *augment theatre* but act independently from a physical scene they only address a part of the potential of this strategy. The guide voice is the driving element of the narration but not unique to AAR environments I will therefore discuss its role as part of the journalistic assessment of this examples analysis.

3.2.4 Immersive Experience

Sounds of Silence constantly feeds multilayered and narrated audio sceneries to the participants. The audio's development is coherent with visual cues which help the participants to navigate through the room and anticipate switches between stations. The participants hence are likely to encounter *perceptual immersion*. The first station of the exhibition introduces the participants to silence. There is almost nothing to hear. Sometimes we hear a dog barking. It's so calm that we can hear the snow falling. The guide asks us to encounter silence for a minute. To not lose the participants awareness due to the emptiness in the audio and the narrative stagnation the audio is supported by a large-scale video of a hazy, snowy, field with individual trees (Figure 6). According to Dormann the high amount of visual information should compensate the auditive emptiness. The creators build this slow room to create a buffer between the space outside and inside the exhibition, giving participants time to switch between both worlds. The last station picks up on this idea of reduced information but without the visual compensation. The participants enter a dimly lit room with seats and have an encounter with Zen Buddhist Niklaus Brantschen who guides the participants to experience silence before talking about

his background and training in silence. Compared to the first station the participant is not asked to experience silence as acoustic phenomena for a specific time but as an emotional and psychological state for as long as they please.



Figure 6: First station of *Sounds of Silence* introducing participants to silence

As a continuous element and familiar narrative tool the guiding voice has an impact on the narration of entire exhibition. It influences our expectations of what will come next and helps us interpret sound scenes and sounds we just heard. It is therefore a driving element enhancing *narrative immersion* by connecting separate sound design elements to a story or placing them in the same context, like the station at minute 08:12-08:46 of the video walkthrough, that introduces us to multiple trumpet sounds all the same measured level but different perceived loudness or the station at minute 18:05-19:00 of the video walkthrough where the participants hear different sounds of a specific measured sound pressure level going from 10 to 95 dB SPL.

While the potential to experience *perceptual immersion* and *narrative immersion* are strong in *Sounds of Silence*, it gives the participants only little access points to encounter *strategic immersion* and *tactical immersion*. The participants don't know what will happen and can't strategize their next moves. An exception is the station where the participants enter a hotel room and can discover sounds and deliberately mix them giving them the freedom to plan their next steps. *Tactical immersion* can't be experience in the mean of being "completely attentive to the task on hand due to a stream of demands for quick reactions" (Agrawal et al., 2020, p. 4). On the other hand, a strong pull to *systematic immersion* can be observed. Unlike *tactical immersion*, *systematic immersion* focuses on the participants' capability to accept and act inside a set of rules or laws applying to a mediated object not corresponding to a similar situation in reality (Arsenault, 2005). The

comparison between *immersive potential* and documented participant behavior suggests that *systematic immersion* was more clearly observable than *narrative immersion*. Dormann's observation that participants moved slowly and explored the sound changes indicates that they accepted the rule system of the exhibition and adapted their behavior to it. However, because no systematic participant interviews are available, conclusions about *immersion* or *presence* remain limited. The strongest available evidence for immersion is therefore behavioral rather than self-reported.

3.2.5 Envisioned Roles

In this exhibition the addressed roles change depending on the station while the *task- and utility-oriented roles* *guide* and *enhance* are the most present ones. When participants lean towards screen embedded in the wall to enter the acoustic scene of the interior of a car, a cave and a library at minute 10:40-11:20 of the video walkthrough the sound gets reduced to single sources. Throughout the exhibition the participants are guided by localizable point-source audio, the guide voice, visual markings and the physical layout of the exhibition. The interactive system also addresses the participants as an embodied collaborator as it creates a tight link between the participants movement and the corresponding change in the audio scenery and can shift their awareness to the body movements, changing the perception of their actions in space. Participants do not only receive information, but learn the logic of the exhibition by moving, turning and approaching sound sources. The observation that participants moved slowly and explored how the sound environment changed supports this reading.

The exhibition also partly connects to the role *interpret*. The guide voice explains what participants hear, contextualizes acoustic phenomena and helps them make sense of sounds that might otherwise stay abstract. In this regard, the AAR environment does not act as an autonomous perceptual collaborator in the sense of automatically identifying sounds, but the combination of spatial audio, narration and exhibition design takes over an interpretive function. A further connection can be made to the role *simulate*. Many stations create artificial or reconstructed auditory situations that allow participants to experience environments which are not physically present in the museum space. These simulations do not primarily aim to replace the physical exhibition room, but to make acoustic phenomena experienceable through spatial sound. In this way, *Sounds of Silence* uses AAR to create soundscapes that can be explored through movement and listening.

At the same time, *Sounds of Silence* only has a limited connection to the roles *personalize*, *reflect* and *transform*. The exhibition does not adapt strongly to individual preferences, and it does not allow participants to change the mix or define their own auditory layer. On the other hand, this limitation ensures a similar experience for each participant. This example therefore does not mainly happen through personalization, but through guided exploration, embodied movement and interpretation. In addition to the envisioned roles for AAR applications, the clearly distinguishable artificial audio from the exhibition space and the markings that represent the boundaries of each station addresses the concerns mentioned in Section 2.2.4 as they support the participants in putting changes in the audio scenery into perspective.

3.2.6 Journalistic Assessment

As discussed in this exhibitions analysis on its technical implementation the headphone and tracking system allows the participants to move freely through the space exploring the exhibitions content. The interaction of the participants with the content and close link to the head and body movements facilitates an embodied experience, different to a linear, fixed media work.

The interactive environment and the sound design tailored specifically to each station enables the participants to access a first-person view of the depicted scenery. As soon as the guide voice enters a scene the first-person view but is strongly mediated. Instead of allowing participants to interpret all acoustic scenes independently, the voice frames perception, directs attention, and often takes over interpretive authority. This reduces ambiguity and supports knowledge transfer, but it also limits the openness commonly associated with immersive journalistic first-person experience. At the station on John Cage's piece 4'33'', the hotel room and the last room of the exhibitions the guide voice moves into the background and participants navigate more independently. While the guide voice limits the participants independence Dormann states, that the prominent use of it ensures that all participants can follow the content. The last station of the exhibition is also the only station where the participants are directly addressed by somebody else than the guide voice.

A lot of the architecture of *Sounds of Silence* priorities the direct and predictable knowledge transfer over more interactive methods with smaller predictability or less mediated stations with higher interpretation demands. These decisions make *Sounds of Silence* mediated and leave the power over the narration to the authors while retaining points that allow a first-person experience. The example's architecture reflects the museum and educational context in which *Sounds of Silence* took place.

3.2.7 Summary

The exhibition's modular structure and open architecture in the central room do not allow for broad narrative arcs. The stations are self-contained, and the participants can only assume little about what they will experience when entering a new station. The limitation on the narrative side creates room for exploration and discovery due to its unforeseeable structure. To ensure a coherent experience and knowledge transfer, the participants are guided through the process by a constant guide voice. The exclusion of the physical acoustics limits the real-artificial mix to peer-to-peer interaction and the synchronicity of visual/graphical and acoustic changes. The exhibition developed a close link between movements and artificial sound scenery using detailed position and head-tracking and a structured introduction of the participants to the system's mechanics. The technical implementation enables the creation of plausible artificial soundscapes, laying the basis for a potential to experience *perceptual immersion*. *Sounds of Silence* uses multiple AAR specific narrative strategies described with the most prominent being *hidden narratives* and the possibility of the participants discovering them.

The observations of the participants' movements and the analysis of envisioned roles conclude that the connection of subject-specific sound design and the tracking and playback system not only guides the

participants but can act as a *rhythmic and embodied collaborator* that can lead to a more mimetic understanding and embodied experience of the topic with a partial first-person perspective. As the work guides interpretation, it does not provide a lot of space for ambiguity and interpretation by the participants. This decision enables people of various age groups and participants without experience with interactive audio environments to navigate the exhibition and ensures the desired transfer of knowledge. Through the museum-based setting, the mediated character of the example is framed clearly and locates its journalistic relevance in terms of its demonstration of how AAR can support embodied knowledge transfer.

3.3 Only Expansion

The next case study highlights an adaptive audio walk created by Duncan Speakman, which was first shown at IDFA DocLab in 2019 (Speakman, 2019). *Only Expansion* addresses climate change and reflects on life on a planet in crisis by changing the acoustic city soundscape and allowing the participants to experience sonically how their life is affected by climate change. The creator Duncan Speakman is an artist and researcher based at the Pervasive Media Studio in Bristol, working with mobile audio and locative media. As part of his PhD at UWE Bristol on the temporal structures of AAR, *Only Expansion* served as a case study next to *Dark By Then*, also created by Duncan Speakman. At the start of the walk, the participants receive a pair of isolating over-ear headphones equipped with microphones on the outside of the headphones. The headphones and microphones are connected to the playback unit. In addition, the participants were handed out a booklet containing chapter names and short prompts for each chapter (Figure 7).



Figure 7: Participant of *Only Expansion* with modified headphones and booklet

At the beginning of the experience, an usher or, in some cases, a recorded voice explains to the participants that the only spoken words they will hear are the chapter names, at which point they should turn the page of the booklet to the next chapter and read the corresponding prompt. The sonic layer of the audio walk consists of field recordings, music, and the microphone input of the binaural microphone arrangement placed on the outside of the headphones. During the walk, the interplay of the sonic layers changes, and the microphone input is temporarily cut out or processed. The audio walk is divided into an outward and a return journey, making the start and finish point the same place. The first five prompts create the outward journey and ask the participants to find a certain place or move in a certain way. Examples for this are prompts like “find the

quietest place” or “begin walking, let yourself be guided by the flow of things”. The sixth prompt starts the return journey, asking the participants to retrace their way back to the starting point. The final prompt asks the participants to stop in sight of their starting location, at which point the soundtrack will eventually stop and end the audio walk.

3.3.1 Sources

The analysis of *Only Expansion* is based on a fixed audio adaption created by Speakman, an interview between the Speakman and journalist Kent Duncan Speakman’s thesis “Composing Entanglement: Temporal Structures of AAR” (kentbye, 2019; Speakman, 2021, 2025). Statements and descriptions made about *Only Expansion* by Duncan Speakman and publishers of the work were used to support the analysis and are labeled individually. The interview and the thesis provide detailed information on the Speakman’s background theoretical framework and implementation process. The thesis further includes participant statements that connect to the research questions of this case-study. The fixed audio adaption is not a representation of the actual audio walk but works a insight into the examples acoustic aesthetics.

This work is included in this thesis because it demonstrates an application of AAR that emphasizes the blend between physical and artificial acoustics and addresses a factual topic without relying on a conventional spoken narration.

3.3.2 Technical Implementation

The audio content is replayed through isolating headphones. The headphone’s ear cups are each equipped with a microphone. the microphones input signal can be fed directly to the participants and allowing the participants to hear the acoustic environment in binaural format and creating active acoustic transparency or conversely also facilitates the possibility for active and passive noise cancelling. This setup can support place illusion because the mediated microphone signal remains linked to the participant’s actual surroundings. However, the effect depends on the plausibility of the acoustic transparency, the balance between live and recorded sound, and the participant’s willingness to accept the mediated soundscape as connected to the visible environment. Acoustic transparency also has the effect of causing participants to perceive their surroundings differently, even without the author using artificial sound sources. Minor changes in timbre caused by the recording and playback media, deviations from the participants’ usual listening volume, and the physical presence of the headphones may be enough to bring about this change.

The headphones and microphones are connected to a rectangular unit, containing a raspberry pi computer responsible for the audio playback and processing of the microphone signal input, an audio interface and microphone preamps (Figure 8). The audio playback and signal processing is based on a fixed timeline, and the participant’s agency is limited to the choice of location and way of movement and is intersected through the actions taking place around the participant’s location. Since the real acoustics is present during the audio walk and gets processed to interplay with the artificial audio content the mixed reality domain meaning the

overlap between artificial and real can be regarded as high. The mediation of the real acoustics can create an unfamiliar impression of a familiar and visually present acoustic environment making the participants aware of their surrounding addressing one of the core aims of AR (Furht, 2011).

In addition to the headphones and processing unit, the participants receive a booklet. The booklet contains titles referring to chapters of the audio-walk and a prompt for each chapter. The booklet acts as a simple guiding system as explained in the introduction. This separation of the guiding system from the audio domain allows Duncan Speakman to commit the audio to abstract narrational elements while the participants still have a simple, familiar guide. The field recordings played back include recording conducted with the microphone system which the participants wear which lets the direct audio input and recorded audio blend more easily and can enhance *place illusion*.



Figure 8: Booklet, playback unit and headphones
used in *Only Expansion*

3.3.3 Narrative Design

Through adding recorded sound to the familiar acoustic environment Duncan Speakman creates a sound scene foreign to the participant which is still placed in the location the participant is physically present. Even if the participants are not familiar with the exact location, they are still familiar with its acoustic properties as long as they are used to similar places. Speakman uses this narrative strategy referred to as using the *familiar environment as a narrative space* through on multiple levels. He mediates of the direct microphone input and blurs the lines between physical and mediated sources making the participants question their perception and making them acutely conscious of their own presence. By placing the environmental field recordings in a

setting where they are connected to the human presence of the individual and surrounding infrastructure, Speakman avoids that participants might envision idealized landscapes untouched by human presence. As Speakman wants the participants “to experience sonically how your own life might change in the future” (Speakman, 2019) and not how it affects a distant place, the experience aims to focus on the acoustics surrounding the each participant and envision how it might change and the multiple layers using the principle of *familiar environment as narrative space* can be seen as an approach to this goal.

In *Only Expansion* the participants choose their own root which will include acoustic events only happening in this specific moment and is played back through the headphones creating a personalized audio stream only participants of the audio walk can listen to. This creates a plausible and personal experience in which Speakman discusses a collective matter, the effect of climate change on our life, and can be connected to the narrative strategy *personal and collective stories tied to a place*. The real-time processing partly represents the strategy of *emergent narratives*. It creates a unique experience connected to the participants’ environment but does not adapt or react to environmental factors, as Harju explicitly mentions the generative possibilities of this strategy. As the work envisions a future state, *bringing back extinct sounds and evoking the future* is another AAR narrative strategy used in *Only Expansion*.

Through interviews Speakman observed how the perception of the participants regard their own presence and the activities taking place around them change. The acoustic overlap make the participants conscious of their own presence and heightens their awareness of their surroundings and was reported by participants that are cited in Speakman’s thesis.

“I started becoming aware of stuff around being around, I kept hearing my breath”

(Speakman, 2021, p. 164)

“I come around the corner and up the steps and I realise its picking up my breath”

(Speakman, 2021, p. 169)

“... I was being guided and the sound had a lot to do with that as well, the, yes, the composition, the sound composition really kind of like, like created a sort of bubble of- bubble’s not the right word though, because I felt much more aware of my surroundings than I usually do”

(Speakman, 2021, p. 104)

The tools used to let the participants experience how they are and could be affected by climate change uncover *hidden narratives* in our familiar environment and can reshape participants' connection to their surroundings.

"First of all I thought this is silly, because I know this area, but it changed it completely. It made me feel as if I wasn't here, I was somewhere else" (Speakman, 2021, p. 145)

"I never noticed half these buildings actually exist. You think I'll have a pause and actually look around" (Speakman, 2021, p. 146)

"I spotted this alley that I'd never been through before; I suddenly found myself in an environment where everything was new" (Speakman, 2021, p. 146)

This audio walk uses AAR's possibilities to temporarily change experiential time and uses measurable times as a compositional tool to affect the participant's spatial-temporal perception of oneself and the world. The narrative strategies which can be observed can't capture the full narration of this work. To achieve entanglement between participant and the environment. Speakman also uses music as a narrative element. This is not unique to AAR but it's contribution to the narrative structure makes it relevant for this analysis.

Speakman discovered that the introduction of music influences the walking pace of participants and affects when participants choose to stop at a specific location. He describes this effect with Kristine Jørgensen's concept of the trans-diegetic use of sound in video games (Jørgensen, 2007). While music scores are a non-diegetic or extra-diegetic elements in movies and linear media, meaning that they can't be heard by the characters and elements inside the story world but are perceived outside the story world and don't affect the progression or plot, music can influence story world in interactive settings. When a gamer or participant is or represents a character inside the story world and can hear the non-diegetic sounds like sudden changes in the music and sound-scene through voice-overs (prompts, hints and explanations), these sounds influence the behavior of the character. Non-diegetic sounds, in this case music, start to influence the diegetic domain. Sounds with this effect are described by Jørgensen as trans-diegetic sounds and are used by Speakman in *Only Expansion* to influence the experiential time and shaping the audio walk's temporal structure.

3.3.4 Immersive Experience

The three main audio layers used in *Only Expansion* (environmental field-recordings, microphone input and music) dynamically change during the walk and affect the overall density of information and the subjective sense of being surrounded. The mixed reality relation can be put in three broad states:

Direct audio only

The audio heard by the participants only consist of the signal received by the microphones and the processing applied to it

Direct and recorded audio

The microphone signal and pre-recorded field recordings or music play can be heard at the same time

Recorded audio only

The microphone signal is cut out completely or processed in a way that the participants can't connect to their environment anymore

Each of the states implies a different connection between the participants and their environment and therefor addresses the participants' possibility to submerge in the environment in different ways. Since interviewees state big shifts in their perception through recorded audio as well as through the real-time audio, any of the described combinations can be considered to evoke an envelopment in an interactive environment as immersion is described by Witmer and Singer (Witmer & Singer, 1998).

While *perceptual immersion* can be experienced in any of the situations, each state forms a different relation between participants and environment. The more dominant the recorded audio becomes, the more isolated the participants become from their environment and therefor a part of the story world. The isolation of participants from their surrounding to eliminate possible distraction, which is considered a significant part to create perceptual immersion in VR, now takes on a different role (McMAHAN, 2003). While the layering of artificial and real audio creates a high awareness of oneself and the environment, isolation draws attention to the participant's inner world, as the effect of music is described by one participant. In the state of heightened awareness on the other hand the real environment becomes part of the interactive system by which they are enveloped. This forms an access to experience *perceptual immersion* that allows for a more embodied experience due to the haptic presence of the story world.

As this thesis uses Agrawal et al.'s approach to describe immersion as a temporal psychological state depending on the *immersive potential* of a system or content and the *immersive tendencies* of the individual, the effect of these shifting connections between the participants and their environment would have to be researched further through specified interviews to be fully captured.

As described earlier the audio content, especially the pre-recorded music and the real-time processing of the microphone signal had an impact on the movement behavior of the participants. One interviewee states that they felt the fear of missing the next station by moving too slow and started to hurry. This reaction of participants to the felt urgency inflicted through audio connects the narrative domain to a strategic challenge and allows participants to take action and supposedly increases the participant's possibility to interact with the system even though the audio walk is based on a fixed timeline.

The influence of music is part of the abstract narrative of the experience and can facilitate a weaker narrative pull and *narrative immersion* than more common audio narrative tools like a guide voice would and runs the risk of losing the participants' attention. The prompts in the booklet counterweight this risk through clear tasks, leaving the participants to small quest and making *narrative immersion* more accessible while leaving space in the audio domain for sound design and music and facilitates access points that allow the participants to take part and explore the experience – components crucial to experience *systematic immersion*. *Only Expansion* therefore does not create immersion through one dominant mechanism, but through shifting relations between environment, body, sound, and time. Its *immersive potential* lies less in stable illusion and more in the continuous disturbance and reconfiguration of the participant's relation to the surrounding world.

3.3.5 Envisioned Roles

The walk's ephemeral interplay between physical and artificial environment *enhances, reduces and guides the participants* overall movement, attention and acoustic perception and is supported by the booklet as a primary guide. It is not individual events that are emphasized or silenced but the full scenery is changed at multiple points in the experience and focused on a holistic change of perception through AAR. Its strongest connection to Tran et al.'s framework therefore lies in its ability to address *emotional and social roles* and work as a *rhythmic and embodied collaborator* rather than in task-oriented assistance.

The roles of *remember* and *reflect* can be observed in the open quests that evoke memories of locations and translate not perceptible time domains to a perceivable timeframe. In addition, the role *reflect* can be observed in the way participants how they became aware of their own presence, breathing, movement and relation to the surrounding city through the mediated environmental sound through microphones and headphones. The audio walk cannot be characterized as a perceptual collaborator since it leaves the interpretation of the audio scenery completely to the participants.

The participants' statements included in Speakman's thesis suggest that the audio walk can create changes in the perception of time, self and environment. These effects are reflected in the categories *simulate, reflect* and *transform* and are important drivers for a *rhythmic and embodied collaborator*. Participants stated that the audio walk affected their movement pace. As the audio walk includes geological time and anticipates a climatic future the audio walk also addresses timespans that are usually not perceivable by humans. The audio walk attempts to transform them into a perceivable form and combine them with the physical acoustic to form the

mix between individual and environment which blurs this boundary and augments our perception and can create entanglement. The connection to personalize is limited but present. The system does not adapt to individual preferences, but each participant's route, pace, acoustic surroundings and momentary attention influence the experience. The limited control of the participants over the system in combination with a narrative design that deliberately blurs the boundaries between artificial and physical environment raises the question if participants felt a loss in personal agency confusion and how safety concerns were as the audio walk might temporarily blend out approaching vehicles or other physical audio cues that are important for the participants.

3.3.6 Journalistic Assessment

Through the audio walk participants were able to gain first person experience of the perceptual condition of how it could feel like to live in a world transformed by climate change rather than a representation of climate. Participants' statements make clear that they felt connected to the topic during the audio walk. The use of abstraction in *Only Expansion* adds transparency regarding its mediated nature in a system capable of blurring the border between physical and artificial sound sources and connecting the different acoustical spaces. The acoustic overlap is used by Speakman to create a new environment which includes sound characteristics of the immediate surrounding as well as those of distant or abstract spaces while keeping the participants anchored in the visually present location. The multiple audio layers affect the participants' perception of time which adds a net of connections between forms of time to the spatial link between places. This complex interplay uses the advantages of AAR to create a mimetic understanding of the effects of climate change, a phenomenon that occurs on a timescale that cannot be experienced directly.

By making the participant's immediate environment the site of the story, *Only Expansion* implicates the topic rather than observing it and puts the participants' experience and perception into the centers of the story. The audio walk can be understood as an alternative to documentary formats that primarily situate the environmental crisis in distant or unknown locations, which are intended to be made understandable through empathy and the corresponding immersive media. This logic has been criticized for legitimizing the experiences of others only when those experiences are tangible and can be absorbed by the audience. In doing so, it ignores the distance between the individual and their fellow human beings and overlooks the fact that we can never fully grasp the inner worlds of others or represent them through technology (Bollmer, 2017).

3.3.7 Summary

Only Expansion represents an application of journalistic AAR that highlights the potential of AAR to discuss non-fictional topics as part of a mimetic experience while leaving the interpretation of it to the participants. By using a combination of music, field-recordings and real-time microphone signal Duncan Speakman points out the effect of hearing on our perception and presents an alternative to the dominant representational strategy of immersive environmental journalism. The analysis of the narrative design points out how the combination of using the *familiar environment as a narrative space* and *hidden narratives* can be adapted to an abstract

storyline and how the impact of non AAR specific narrative tools can change when used in an AAR environment. The technical implementation is designed specifically for this work and consists of the modified headphones, playback device and booklet yet the booklet is the only part which requires direct interaction. Despite the fixed time structure and therefor supposedly linear progression of the composition the perception of time is manipulated in multiple ways and *Only Expansion* offers the potential for *strategic challenge*, *narrative immersion* and *systematic immersion*.

The mixed-reality acoustic environment keeps shifting the participants' sense of *presence* and connection to their environment and creates a gateway to experience the subjective sense of being surrounded. This is also observed in terms of the examples connection to the envisioned roles shows the examples possibility to act as a *rhythmic and embodied collaborator*. If the link to the real environment through the real-time microphone signal is present and plausible, the virtual and physical space can align and create place illusion without the need of an artificial representations of the physical environment. The booklet on the other hand takes over the major part in guiding the participants and builds the basis for *narrative immersion*.

Even though one of the main goals of IJ is to increase the participants' empathy regarding the documented topic, a lot of western news and media in 360° videos miss this goal due to their narrative design. By creating stories set in a supposedly foreign world that is discovered through documentation, authors divide between audience and a documented space that is left after the video ends. Works like *We Wait* (Aardman Animations & BBC Connected Studios, 2016) position the participant as an observer of a geographically and culturally distant crisis, a dynamic that critics of IJ have connected to broader questions of postcolonial representation in documentary media (Bollmer, 2017; Wincott, 2024). *Only Expansion* proposes a different logic: rather than transporting the participant to a distant crisis, it implicates the participant's own environment in the story. The participants are rather presented with possibilities of a changing acoustic environment and a collection of perspectives to create an embodied understanding of a complex process.

3.4 Consent – Walk the Walk

An examination of *Consent – Walk the Walk* adds an application to the analysis closely linked to radio feature journalism that is labeled as landscape journalism, by the authors. The geo-locative audio walk, created by Chris Brookes and Emily Deming and published in 2018 in St. Johns, Canada, discusses a trial from 2017 in which an in uniform and on-duty police officer was charged with sexually assaulting an intoxicated woman he drove home the city's nightclub precinct on December 21, 2014. The jury trial which lasted two weeks ended in an acquittal and led to days of protest in the Newfoundland city. After the case was reopened by the Canadian supreme court in 2019 but was later canceled due to an error in how the judge chose which jurors were dismissed (CBC, 2020), Constable Carl Douglas Snelgrove was found guilty in a third trial in 2021 and sentenced to four years in prison (Mullin, 2021; Smellie, 2021).

The audio walk must be understood in relation to the legal situation at the time of its publication. It was created in response to the controversial 2017 acquittal and the public debate that followed. Later legal developments, including the ordered retrial and the 2021 conviction, change the retrospective context of the case, but they do not change the original journalistic situation in which the audio walk was produced.

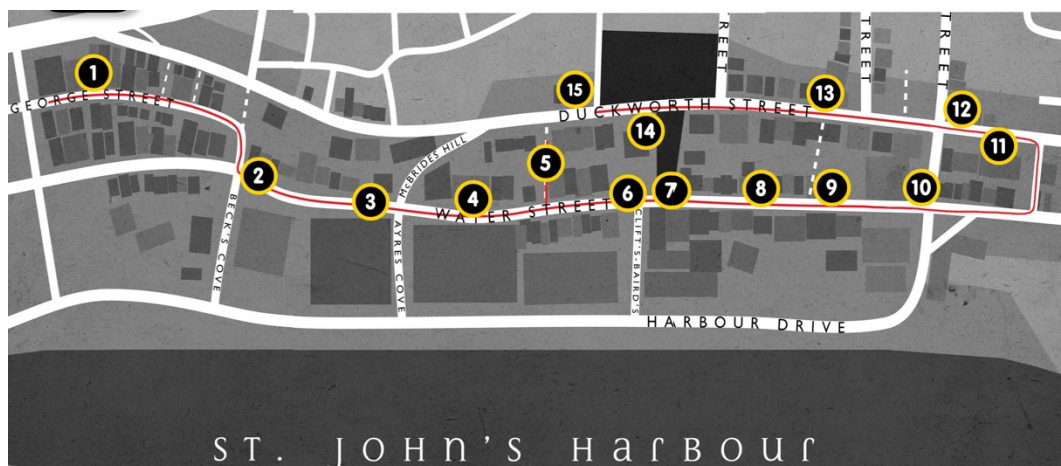


Figure 9: Map of *Consent – Walk the Walk* with numbered stations

The participants follow the proceeding of the 2017 trial as an observer by walking a route of approximately 1,2 kilometers through the city that involves fifteen stations that represent locations mentioned in the trial (Figure 9). Each station is one part of the full story which unfolds chronologically regarding the night of the assault. The audio walk was implemented in a dedicated smartphone app that explained the process, guided the participants through the city and automatically played the audio when the participants reached the corresponding location. The actual content was exclusively audio based while the app provides further background information about the trial and the making of the audio walk. The authors attach locations to corresponding pieces of the story and by doing so bring thematically connected but time and location independent moments together. The participants go through the trial by listening to a re-enactment of the

transcribed testimonies made in court and recreated audio scenes depicting the events that happened on the night of the alleged assault. For this thesis the audio walk is particularly interesting because it is a direct use of locative audio for journalism and shows the possibilities and difficulties of AAR in the context of a news report.

3.4.1 Sources

The following analysis is mainly based on a fixed media web-version of the actual audio walk accessed on the battery radio webpage and the article “Consent: Objectivity and the Aesthetics of Reenactment in Locative Audio Journalism about a Sexual Assault Trial” by Jeanti St Clair written for the journal *RadioDoc Review* in 2019 (Brookes & Deming, 2018; St Clair, 2019). Further literature which was used to support the analysis is cited in the text.

The web-version of the audio walk made it possible to get to know the audio walk’s subject and supported the article by St Clair, as the article frequently refers to stations of the audio walk. The work is included in this thesis because it represents an application of AAR that is close to traditional news journalism and audio reportage and does not rely on specified hardware.

3.4.2 Technical Implementation

Since the audio walk is played in an app on the private smartphone and listened to with private headphones the technical system is not uniform and requires the participants to possess a compatible phone. This implementation requires the project to be adapted to the requirements of smartphones, and the exact setup can’t be controlled by the creators. The application’s interface shows a photograph of the Supreme Court of Newfoundland and Labrador building in St. John’s as well as three icons at the bottom of the screen. One to enter the map, one to access background information on the case and reading instructions about the app and one to access links for sexual assault services, information on consent and media reports on the Snelgrove case. The app accesses the phones GPS location and plays the audio automatically as soon as the participants reach the corresponding location and is therefore interactive in real-time. The location-based playback allows the participants to access a layer of history not represented in other forms of remembering like architecture and statues. When approaching the fourteenth location the participants can choose to which of the closing arguments they want to listen to and if they to listen to any or both. The audio content is not registered in 3D to resemble the real location but played in a fixed stereo format.

According to Harju such implementations can still be considered AAR since the participants still interact with the physical world even when they parts don’t match acoustically. As the audio content doesn’t adapt to the participants’ movement or environment ones the audio file starts playing, the descriptive sound design becomes a more important factor for participants to align the story with their visual surrounding. The technical implementation structures the participant’s action primarily through movement between locations rather than through continuous interaction with the audio. Unlike *Sounds of Silence* or *Only Expansion*, the participant

does not influence the composition, spatialization, or processing of the sound once a section has started. Agency is therefore concentrated in navigation, pacing, and the moment of triggering audio at specific sites. The interaction between real and virtual environment is based on geographical correspondence rather than acoustic registration or real-time sonic adaptation.

3.4.3 Narrative Design

Consent – *Walk the Walk*'s general narrative structure follows a linear unfolding of the story with a clear buildup ending in the closing arguments and the judgement of the court. The narrative and dramaturgical decisions of the authors following the linear storytelling do not only connect to the narrative design analysis but will also be addressed in the analysis on the immersive experience and the journalistic assessment.

When focusing on AAR specific narrative strategies, one important aspect of the geo-locative audio walk is, how it is embedded in the city's geography and connects specific locations to personal and *collective stories tied to a place*. The audio walk anchors the personal story of the victim in the specific or representative location. It further follows and discusses important events of St. John's recent past and collective memory. In doing so the audio walk enables the development of place literacy and makes the trial's testimonies accessible for the public. The geo locative approach used in this audio walk can also be connected to the narrative strategy of using the *familiar environment as a narrative space* and the principle of *hidden narratives*.

Due to the restricted permission to use the original audio recordings of the trial for the audio walk, the creators had to transcribe the testimonies. These transcriptions were then read by actors, recorded by the authors and arranged to be used as a representation of the testimonies at court. In this process the transcripts were reordered to build a narrative that follows the chronological timeline of the events of the night of the assault which does not represent the chronological order of the trial. This timeline is also reflected in the ordering of the locations the participants pass and by the authors to strengthen *narrative immersion*, as explained in this work's analysis of the immersive experience. The re-enacted situations of the courtroom are interrupted by sceneries of descriptive sound design which are used to depict events of the night of the assault. When participants listen to the testimonies, they take part in a discussion about the events rather than being present at the events itself. Therefore, the night of the crime is shown to audio walk participants through the perspective of the witnesses. The personal testimonies allow participants to get to witnesses while they can still be distant to the locations in which the events happened. The descriptive sound design can close this gap through letting the participants experience one location at multiple points in time and can connect the time of the crime with the time of the sound walk and building a bridge between the participants physical environment and the story space. As this interlinking of past, present or future is used for past events and not past sounds, it is not *bringing back extinct sounds and evoking the future* but also acts as a historical document that makes the past accessible as it is this strategy's potential.

3.4.4 Immersive Experience

The story's linear buildup creates a clear narrative structure and opens the possibility for *narrative immersion*. The transcripts from the courtroom as well as the route of the audio walk were arranged in chronological order to the night of the crime. Unlike the trials chronology which expands over two weeks, the night of the crime only expands over several hours and can be represented easier in the audio walk. In the reordering the authors also facilitated a juxtaposition of the different testimonies that refer to the same event which highlights the contradictions in the statements. The potential for *narrative immersion* gets strengthened by the audio-walk's topic, which is related to the news genre of true crime, which is considered to trigger strong emotional reactions in the receivers and profit from sensationalism even when they are aimed at objectivity (Wiltenburg, 2004). While this audio walk's potential for *narrative immersion* has multiple points of connection, the technical implementation prevent control over the potential to experience *perceptual immersion* since it depends a lot on the participants' personal setup. In a case where the participant uses transparent headphones and completes the walk at night, the descriptive sound design can overlay the participant's physical environment and can create a more location embedded experience, while another participant with a different setup, or bad internet connectivity listening to the walk during the day might not experience any congruence between the story world and their physical environment.

The audio walk also shows little potential for *strategic immersion* and *tactical immersion* since the participants' only moment of agency, next to the decision at which time they enter the next location and trigger the next section of the story, is at location 14 when they can choose which closing argument they want to listen to.

The opposite of the linear audio walk could be unnumbered locations where participants can choose their own route. As St. Clair states, a participants free choice of route may end in loosing track of the plot and can confuse the participants, as observed by Josephine Reid in their analysis a non-linear locative audio (Reid et al., 2005; St Clair, 2018). Since the analysis does not include participant interviews or systematic observations, conclusions about actual immersion remain limited. The work's *immersive potential* can be assessed through its design but not fully compared with documented participant experience. Based on the available material, *Consent – Walk the Walk* appears to offer its strongest *immersive potential* through *narrative immersion* and place-based contextualization, while its technical setup offers weaker access to experience *perceptual*, *systematic*, *strategic* or *tactical immersion*.

3.4.5 Envisioned Roles

Consent – Walk the Walk has the least direct connection to Tran et al.'s envisioned roles if these roles are understood as technically adaptive sound processing. The most direct connection lies in the role *guide* that is part of the *task- and utility oriented roles*. The app guides participants through the city by connecting specific locations to parts of the story. As the only real-time input of the system is the participants' location, it does not *enhance*, *reduce* or *transform* the participants' physical acoustic environment. Nevertheless, the audio walk still integrates participants by connecting movement and location to the corresponding parts of the subject. For

this reason, the roles should be applied here in a broader sense that includes how the audio layer changes the participants' understanding of the physical environment.

Following this broader definition, the guidance mentioned above is then not only navigational, but narrative. Participants move through public space in the order defined by the authors and encounter the subject in relation to the places mentioned in the case. The system therefore structures both physical movement and narrative attention. The audio walk further creates a connection between physical surroundings and local collective history which connects to the role *remember*. It can further initiate a reflection on the personal, social and legal definition of consent which connects to the role *reflect*. As the audio layer gives meaning to ordinary places in relation to the participants' location a further connection to the role *interpret* can be made. These complex connections show the potential of AAR and geo-locative media for journalistic applications that can be achieved while using established narrative forms and could be an approach of imagining journalism applications in ubiquitous AAR environments.

3.4.6 Journalistic Assessment

Consent – Walk the Walk shows many parallels to feature journalism, a journalistic genre that can serve as a hyperlocal form of feature journalism (St Clair, 2018). This also means that the audio walk's claims on objectivity and depiction of reality can be compared to the ones generally used in journalism. Using a re-enacted version of the courtroom, rearranging the testimonies, adding music and descriptive sound effects are all mediated layers which can distort the report and might contradict the SPJ code of ethics principle to "Seek Truth and Report it". Being aware of this contradiction, the authors use the introduction to the audio walk to make their work process and the re-enactment transparent and therefore allowing the participants' to contextualize the content. By using popular technical tools and minimal potential for *presence* by experiencing *perceptual immersion*, the authors can assume that the participants are capable to distinguish between the physical and mediated dimensions of the audio walk and doesn't demand for further explanation about the mediated nature of the audio walk.

As St. Clair analysis *Consent – Walk the Walk* in regard of its journalistic validity, she differs between feature journalism and news journalism and places the strength of factual locative audio in feature journalism rather than news journalism. As in other forms of feature journalism, this audio walk crafts engagement by discussing "personal experiences that are related to contemporary events of perceived public interest" (Steensen, 2011, p. 51). Radio documentaries are also characterized using scenes, ambient sound, narration, music, and interviews, all of which are carefully structured through dramaturgical scripting and editing.

These same storytelling elements and production techniques are also employed in the creation of locative audio documentaries. Like written and audible feature journalism, locative audio documentaries require a high production effort which doesn't align with the frequency in which news reports have to be published and thus is also described as slow journalism by St.Clair. The dramatization and multilayered use of sound enables the

participants to take on a different perspective follows the description that a “News reportage tells us, whereas feature journalism and documentaries show us” (St Clair, 2019, p. 29).

The narrative strategies used by Brookes and Deming moves the work away from news report into a fictional literary journalism but also aids the contextualization of the topic and events by facilitating an understanding and broader discussion on the definition of consent in Canadian law. Unlike some forms of IJ, *Consent – Walk the Walk* does not place participants inside the direct perspective of one protagonist. Instead, it creates a situated listening position between witness, observer, and member of the local public. This limits first-person embodiment but supports journalistic distance and reflection. As observed in the analysis of the narrative design the authors enable place literation and anchor points in time through the corresponding locations by attaching audio to these locations. This layering of relevant sights with an artificial counterpart creates space for a historic layer not through more long-term medias like statues, platters or other forms of long-term remembrance.

3.4.7 Summary

To achieve a clear focus on the narrative, this work uses a common technical setup based on a smartphone software with no customized hardware and, through its technical implementation, becomes a mixed-reality experience that closely resembles linear radio journalism. Participants are generally familiar with the technology which overshadows its limitations in terms of *immersive potential*, compensates for the resulting shortfall in perceptual transparency and allows the participants to focus on their surroundings and the content of the audio walk without having to learn the system’s rules first. The choice of technology to access the audio walk further ensured that the journalistic content using an evolving media is still available for the majority of the local population (Canadian Telecommunications Association, 2023)

The geo-locative interaction, core to the audio walk, is used to place the personal testimonies in the public space and connect the collective history of the community to the participant’s memories of locations across the city. This place-based knowledge transfer consists of three situations (the night of the crime, the trial and the sound walk) that follow the same chronological order linearly and refer to the same locations. Through the arrangement of the statements, the authors highlight their contradictions and illustrate the complexity of the topic which can, in combination with other narrative design decisions, create immersion through *narrative immersion*. The lack of overlap between artificial and real environment and interaction possibilities mostly prevents the participants from experiencing other forms of immersion. In terms of the envisioned roles this example only connect to the role *guide* as the technical system’s context awareness is limited. If the roles are applied in a broader sense the effect enhancing the physical environment with an additional historical layer shows that this example addresses *emotional and social roles* and *perceptual collaborator roles*.

The strength of geo-locative applications for journalistic topics is highlighted, when analyzing it from the journalistic content’s point of view. It shows that the use of AAR and concepts related to IJ for journalistic

implementations needs a transparent communication about the work's mediation and work process and needs to name sources as well as their use in the work. The literal adaption of journalism for this evolving media raises the most urgent questions about objectivity and ethics of journalism. As *Consent – Walk the Walk* used characteristics related to feature journalism to produce an environment that explores the boundaries between objectivity and subjectivity based on an emotionally charged topic, this audio walk served as a prime example to observe the need for further research.

3.5 Comparison and Discussion

The three examples analyzed represent different approaches to how the characteristics of AAR can be used for journalistic content and were able to be contextualized through the analysis within the framework of AAR and IJ. To capture the breadth of the field of application, examples were selected that differ not only in their technical implementation but also in their subject matter. Each example relates differently to the literature underlying the analysis and makes it possible both to test the concepts in practice and to understand how they interact. The discussion of the case studies follows the same structure as each analysis and is summarized at the end of the chapter

3.5.1 Technical Implementation

When comparing the technical implementation of the three examples, they can be placed on a spectrum between precise spatial tracking, adaptive acoustic transparency and accessible geo-locative playback. *Sounds of Silence* uses the most precise tracking system of the three examples. The position- and head-tracking system allows the audio to react directly to the participants' location and head orientation and creates a close link between physical movement and sound scenery which allows participants to explore the audio layer through movement connected to their own body. At the same time, the closed headphones strongly reduce the physical acoustic environment and therefore limit the direct acoustic overlap between real and virtual sound. The mixed-reality relation in *Sounds of Silence* is therefore mostly created through spatial correspondence, visual references and bodily movement.

Only Expansion uses a different strategy. Its technical system has no control over the exact location of the participants. Instead, the microphones mounted on the outside of the headphones allow the physical acoustic environment to become part of the audio composition and influence the participants' perception and sense of presence. The technical system used in *Only Expansion* creates a strong overlap between physical and artificial acoustic components than in the other examples and achieves to blur the line between physical and artificial environment.

Consent - Walk the Walk uses the simplest technical system. It does not adapt the audio content to head movement, walking speed or the acoustic environment after the playback has started and the interaction between physical and mediated environment is based on geographical correspondence rather than on acoustic registration or continuous real-time adaptation. Compared to the other two examples, this creates less *immersive potential* on a technical level. At the same time, this simplicity makes the work more accessible and closer to established media practices like radio feature journalism and podcast listening.

The comparison shows that the technical system has a strong influence on the participants' agency and determines which senses are engaged but does not determine the level of journalistic information. *Sounds of Silence* has a highly developed tracking and playback system, but its journalistic relevance lies mainly in its ability to create embodied knowledge transfer in an educational context. The technical setup used in *Only Expansion*, can blend the physical and artificial environment but the interaction with the system remains mostly indirect. *Consent - Walk the Walk* uses the least advanced technical system but is the most direct journalistic example because it places trial material, testimonies and public debate in the locations connected to the case. The technical implementation of AAR should therefore not be evaluated only by its technical complexity, but by how well it supports the topic, the participants' role and the intended form of knowledge transfer.

3.5.2 Narrative design

The analysis of the examples use of AAR specific narrative strategies showed, that the importance of the individual strategies varies a lot. In the exhibition *Sounds of Silence*, the most important strategy is the use of *hidden narratives*. The participants enter a visible exhibition space, but the meaning of the different stations does not exist before it gets uncovered by listening and moving through the room. The same narrative strategy is used differently in every example. While *Only Expansion*'s use of *hidden narratives* relies on blurring the line between physical and artificial environment to reveal a narrative which incorporates the present one, *Consent - Walk the Walk* adds an additional historical layer that acts in parallel to the existing one.

In *Only Expansion* and *Consent - Walk the Walk*, the strategy of using the *familiar environment as narrative space* is much more central. The participants do not enter a specially designed exhibition space in *Only Expansion* but move through an existing urban environment that is altered through the processing of the microphone signal, added recordings and music. In *Consent - Walk the Walk* the strategy allows a strong connection between a location and its meaning in regard to the story. The narrative strategy is used in a more journalistic way to create place literacy. The narrative strategy to use the *familiar environment as narrative space* is not present in *Sounds of Silence* as it is developed for a space that specifically designed for this exhibition.

Another central strategy in *Consent - Walk the Walk* is the ability to explore *personal and collective stories tied to a place*. The subjective and personal testimonies are placed in a public space and carries the documentation of a subject that is part of the recent collective past. A part of this strategy is also used in *Only Expansion* through addressing the collective matter of climate change and connect it to the participants' environment.

In addition to these most present narrative strategies the examples all used multiple other strategies that support the central narration. The narrative Strategy of *collective and personal* can be observed in all examples. The use of headphones and individual playback engines for each participant means that each participant hears a different audio stream while sharing a common space with other participants. The advantage of headphone-based applications to address only one participant at a time is crucial for all examples. It shapes the participants

real-time interaction with the subject in *Sounds of Silence* and *Only Expansion* and makes the layering of historical information in *Consent - Walk the Walk*.

All examples contain parts that connect to the strategy of *bringing back extinct sounds and evoking the future*. Interweaving the physical and virtual environment to envision an environment impacted by climate change is the most elaborate use of it. *Sounds of Silence* and *Consent - Walk the Walk* also combine different timelines, but do not embed them in the participants' physical environment.

The abstract narration used in *Only Expansion* mostly prevents the use of *augmented theatre* as it was applied in the other examples. The staging of scenes and spatial placement of their sources, as observed in *Sounds of Silence* and *Consent - Walk the Walk* connects partly to this strategy, as it contains theatrical elements which act independently and is not used to augment physical scenes.

None of the example's narratives fully emerges from environmental factors and the strategy *emergent narratives* therefore is not a central part of any of them. Since the narrative development can differ between participants in *Sounds of Silence* and *Only Expansion* and the real-time processing of the participants' environment in *Only Expansion* creates unique audio walks and represents the narrative strategy of *emergent narratives*. The potential of AAR systems to react and adapt to environmental factors is not applied in any of the examples and the impact of generative elements on the narration of journalistic AAR can't be analyzed.

The comparison of the narrative design shows that AAR can create engagement with a topic in multiple ways and varying degrees. In *Sounds of Silence*, engagement emerges through discovery and exploratory listening. In *Only Expansion*, engagement emerges through a changed perception of self, environment and time. In *Consent - Walk the Walk*, engagement emerges through the connection of personal testimony, public space and collective memory. The AAR specific narrative strategies therefore do not only create novelty or interactivity but change the relation between the participant and the subject of the work. The analysis showed that the criteria of AAR specific narrative strategies are not always met completely. This observation together with the finding, that every example also uses general narrative strategies shows how complex the narrative design of practical works can be and how closely related general and AAR specific narrative strategies are. Used in combination it can create an environment in which the subject is encountered through movement, location and bodily presence instead of simply listening to it.

3.5.3 Immersive Experience

When comparing the immersive experience of the three examples, it becomes visible that immersion cannot be described as one single effect. It also became clear, that it is possible to recognize components which can lead to immersion by analyzing the technical setup, narrative and interactive structure but the evaluation of their immersive effect would rely on self-reported statements. However, because the analysis is not based on systematic participant interviews, the strongest evidence for immersion remains behavioral. By exposing components that can lead to immersion, the analysis contextualized a central selling point of immersive media and made it possible to analyze the example in relation to IJ.

It became clear, that *Sounds of Silence* offers a strong potential for to experience *perceptual immersion* and *systematic immersion* that is based on the technical setup and made accessible through the narrative design, which introduces the exhibitions logic. The documented participant behavior suggests that they accepted the system's rules and adapted their movement to the audio environment. The heightened awareness of participants about auditory changes and their own presence that resulted from that was also stated by participants of *Only Expansion*. Duncan Speakman's audio walk creates *immersive potential* through shifting relations rather than through one stable illusion. The participants move between moments in which the real acoustic environment is present, moments in which it is mixed with recorded material and moments in which it is reduced or removed. The *immersive potential* lies in this changing relation between body, environment, sound and time. This example's *immersive potential* not only created by being surrounded by sound, but by the disturbance of ordinary listening and the constant reconfiguration of the relation between self and environment.

Consent - Walk the Walk on the other hand crafts its strongest *immersive potential* through enabling *narrative immersion*. The linear structure of the route, the chronological arrangement of the testimonies and the emotional weight of the topic support engagement with the story. At the same time, the technical setup offers only limited access to *perceptual immersion* or *systematic immersion*. This weaker potential for perceptual immersion can also be seen as a journalistic advantage, because it helps to maintain distance between the participant, the re-enacted material and the physical city.

This comparison shows that journalistic AAR should not necessarily aim for maximum *immersive potential*. In some contexts, immersion can support embodied knowledge transfer, as in *Sounds of Silence*. In other contexts, unstable or reflective *immersive potential* can help participants question their own perception, as in *Only Expansion*. In a direct journalistic context like *Consent - Walk the Walk*, too much perceptual immersion could risk blurring the boundary between evidence, reenactment and interpretation. AAR therefore has to balance involvement and distance depending on the topic. The goal is not always to make the participants feel as if they are fully inside the story, but to create a meaningful relation between the participants, the mediated content and the physical surrounding.

3.5.4 Envisioned Roles

The comparison of the three case studies in terms of their connection to the envisioned roles for AAR application shows that this perspective of analysis can be useful if they are not understood too narrowly as a technical function. Tran et al.'s study focuses on AAR applications for everyday life while the analyzed examples are bounded works and therefore do not have to address the same requirements.

Sounds of Silence mainly addresses the roles *guide*, *enhance*, *interpret* and *simulate*. The guide voice helps participants navigate through the space and interprets the exhibition while the technical implementation provides the foundation to acoustically reskin the exhibition space. This reskinned space in combination with the technical implementation creates the strongest connection to Tran et al.'s framework in the form of a

rhythmic and embodied collaborator as the participants' adapt their movements to the spatial audio system in to explore the exhibition.

Only Expansion connects most strongly to the emotional, reflective and embodied roles. It does not guide participants toward clearly defined information points, but changes their relation to their environment, movement and sense of time. The roles *reflect*, *simulate*, *transform* and *remember* are therefore more relevant than the *task- and utility-oriented roles*. The work is also the clearest example of the *rhythmic and embodied collaborator*, because it affects walking pace, bodily awareness and the perception of long temporal processes such as climate change.

In *Consent - Walk the Walk*, the connection to the envisioned roles is less direct. The system does not react to the participants' acoustic environment and does not operate as an adaptive perceptual collaborator. If a broader definition of the study is applied the connection between *Consent - Walk the Walk* and the envisioned roles become clearer. The attachment of meaning and collective memory to public locations address roles *remember* and the narrative design highlights the ambiguity of the subject and can create an access point for the role *reflect* as participants are encouraged to reflect on the different definitions of consent.

In terms of the second research question of the study, the analyses can only make assumptions, as none of the literature discusses whether the participants would have needed higher personal agency and control or whether an unclear distinction between artificial and physical audio was concerning. In *Sounds of Silence*, the graphical support and the clear distinction between physical and artificial space could help participants to anticipate when and how the audio scene changes and support a conscious interaction with the technical system. In *Only Expansion*, the mixing of physical and artificial sound is a central aesthetic and narrative element. It might, however, be that this effect at the same time mitigates the participants' awareness of their surroundings in a moment when it would be necessary. The limited potential for perceptual immersion and the possibility to explicitly control the audio walk in the smartphone app might create a distance between subject and participants in the example *Consent - Walk the Walk*. This might be helpful for the participants to distinguish between factual and reconstructed elements of the audio while still being able to follow the development of the story.

3.5.5 Journalistic Assessment

The journalistic assessment of the three examples reveals the main tension of AAR as a medium for non-fiction subjects of work. AAR can create embodied and situated forms of understanding, but it also strengthens questions of mediation, objectivity, authorship and ethical distance.

Sounds of Silence shows how AAR can support embodied knowledge transfer. Its relevance for journalism does not lie in news reporting or objective depiction of reality in a narrow sense, but in the possibility to make acoustic and perceptual phenomena experienceable. *Only Expansion* shows how AAR can address a complex documentary topic without relying on the promise of full empathy or direct access to another person's

experience. The participants are not transported into a distant crisis or placed into the position of a represented other. Instead, the work changes their relation to their own surroundings and implicates their immediate environment in the topic of climate change. This makes the work relevant for a critique of IJ that understands empathy as a problematic goal when it claims to make the experience of others fully accessible. *Only Expansion* rather creates a mimetic and reflective understanding of climate change by making the participants' own perception unstable.

Consent - Walk the Walk is the most explicit journalistic example and therefore also the example in which questions of objectivity and transparency become most urgent. The work uses reenacted testimonies, rearranged transcripts, music and descriptive sound design. These elements strengthen narrative engagement, but they also open questions about mediation and journalistic responsibility. The transparency about the use of transcripts and reenactment is therefore central. The example shows that AAR applications for journalism need to communicate clearly which parts are documented, which parts are reconstructed and how sources are used. Across the three examples, AAR appears strongest when it does not claim to remove mediation. Its potential lies rather in making mediation spatial, audible and embodied. In each example, the audio layer does not simply add information to a place. It reorganizes the relation between participant, space, body, memory and narration. The comparison also shows that the first-person position in AAR has to be understood carefully. First-person experience can help to create embodied understanding, but it can also reduce critical distance if it is presented as direct access to reality.

In *Sounds of Silence*, the participant gets a partial first-person experience of acoustic scenes, but this experience is strongly guided. In *Only Expansion*, the first-person perspective is used to question the participant's own perception rather than to simulate another person's inner world. In *Consent - Walk the Walk*, the participants is not placed inside the perspective of the victim, the accused or the witnesses, but they take on a situated listening position between observer, witness and member of the local public. These differences show, that journalistic AAR should not simply aim to place participants "inside" the story. Depending on the topic and authors intention, it can be more appropriate to position them as participants, observers, explorers, witnesses or reflective participants.

3.6 Summary

The discussion of the technical implementation placed the examples on a spectrum between precise spatial tracking, adaptive acoustic transparency and accessible geo-locative playback and concluded that the technical implementation affects participants' agency and determines the addressed senses. However, the technical complexity of the technical system does not determine the level of journalistic information, as can be seen in *Consent - Walk the Walk*, the example that uses the least advanced technical system while being the most direct journalistic example. The technical implementation should therefore also be evaluated in regard to its ability to support the topic and intended form of knowledge transfer.

In terms of narrative design, the comparison revealed how the most prominent AAR specific narrative strategies used in the examples connect and that all examples partly address most of the AAR specific narrative strategies in varying importance. The findings that all examples use general narrative strategies in addition to the AAR specific narrative strategies and that all examples partly address the majority of the AAR specific narrative strategies showed that AAR can create engagement in various ways and degrees and that the narrative design possibilities available in AAR can change the relation of the participants and the subject of work.

As in the discussion of the technical implementation, the comparison of the examples in regard to their *immersive potential* showed that its use must be evaluated in terms of how well it supports the journalistic context. The aim for maximum *immersive potential* would for example be unsuitable in *Consent - Walk the Walk*, where the subject depends on the participants' possibility to distinguish between facts and interpretation, but is an important factor to enable a more embodied knowledge transfer in *Sounds of Silence*.

The comparison of the analyses relating to the envisioned roles of AAR observed that the degree in which the example can be connected to the study varies a lot and the second research question of the study cannot fully be answered as no interviews about the participants' concerns and control preferences are available. To explore the effect of a broader framework the analysis of *Consent Walk - Walk* also examined how it can be connected to the study apart from the audio domain. This analysis concluded that Tran et al.'s framework cannot be fully adapted to journalistic works, as the study's focus on everyday life situations differs from the analysis's examples.

The comparison of the journalistic assessments unveiled how different the knowledge transfer can be and how complex the system is which influences it. All examples, independent of their specific way of addressing a subject, raise questions of objectivity and transparency. Using AAR for journalistic topics might therefore not be appropriate when the goal is maximum objectivity but offers the potential to reorganize the relation between the subject, the participants and the author. Additionally, the comparison showed that the use of a first-person experience depends on the specific subject and content, and that the specific perspective from which participants engage with the subject must be considered in the design process.

4 Conclusion

To investigate the potential of AAR for journalistic narratives, this thesis explored the related fields of research, IJ and AAR, with a focus on their distinct narrative characteristics and narrative methods. The findings of this research formed the foundation for a framework along which three case studies were conducted to understand how the examined field of application for AAR is applied in practical examples.

The background analysis on IJ showed how the key components *presence* and *immersive potential* impact the most common narrative methods used in 360° video formats and bridged IJ to AAR by introducing spatial audio journalism. Based on the narrative methods discussed in this chapter, it became clear that a journalistic use of applications that are considered to create immersive environments has to consider the influence of *presence* and *immersive potential* not only as perceptual components and psychological states but also as narrative tools that impact the subject of work. The analysis of AAR continued these findings by exploring the characteristics of AAR and how technical components of AAR applications reflect these characteristics. It showed that the potential of the technical components to reflect the characteristics of AAR is related to specific spatial and narrative design strategies. To conclude the background analysis, the chapter investigated possible roles for AAR applications and, in doing so, tied into the observed issue in the chapter on spatial audio journalism that, even though the technical environment for AAR applications is widely available, the actual application possibilities are limited.

The case study in this thesis analyzed the exhibition *Sounds of Silence* and the audio walks *Only Expansion* and *Consent - Walk the Walk* from five perspectives to evaluate how the findings of the background analysis are represented in actual works. The case study showed that each of the examples possesses a specific focus and that they differ in their relationship to journalism, ranging from educational and documentary formats to explicit journalistic practice. *Sounds of Silence* creates an environment in which the participants can explore the exhibition and the different facets of the subject. *Only Expansion* blends the physical and artificial environment to change the participants' sense of *presence* and the relationship to their environment. *Consent - Walk the Walk* enriches the urban space with collective history to create a spatially embedded journalistic discourse.

The significant differences between the examples showed that the area research describes a field of practices in which artificial audio, physical space, movement, and narration are combined in different ways rather than treated as separate concepts. The analyses further lead to three central conclusions. First, the journalistic potential of AAR depends on the relation between the technical system and the work's subject. The technology must support the form of knowledge transfer and narrative design, which does not necessarily come from maximum immersion or interaction. Second, immersion in AAR should not be reduced to *presence*, empathy, or sensory intensity. The examples show how exploration, disturbance, reflection, absorption in the narrative,

and embodied listening act as starting points to build *immersive potential*. Third, AAR requires a high degree of transparency in journalistic contexts. Because spatialized, locative, and responsive audio can make mediated content feel immediate and personally experienced, the work process, sources, reconstructions, and authorial decisions need to be communicated clearly.

Due to the limited number of case studies and a lack of experience reports from participants, the findings of the case studies can only be referenced by the available literature and author statements. The research on the potential of AAR for journalistic subjects faces further difficulties, as many AAR applications are site-specific and only accessible for a short period of time and will always partly have to rely on documentation to follow their development.

Following the limitations mentioned, future research that focuses on the question how AAR and journalism interact, could conduct interviews with participants to empirically ground the findings of this thesis. In addition to a more concrete picture of the field of research, experience reports focusing on journalistic narratives could further be used to support authors in navigating through this complex field and inspire software development aimed at creating systems that capture the complex narrative and perceptual connections that AAR offers and journalistic subjects deserve.

5 Bibliography

- Aardman Animations, & BBC Connected Studios. (2016). *We Wait*. Aardman Animations & BBC Connected Studios. <https://www.bbc.co.uk/taster/pilots/we-wait>
- Agrawal, S., Simon, A., Bech, S., Bæntsen, K., & Forchhammer, S. (2020). Defining Immersion: Literature Review and Implications for Research on Immersive Audiovisual Experiences. *Journal of the Audio Engineering Society*, 68(6), 404–417.
- Aitamurto, T. (2019). Normative paradoxes in 360° journalism: Contested accuracy and objectivity. *New Media & Society*, 21(1), 3–19. <https://doi.org/10.1177/1461444818785153>
- Apple Inc. (2026). *Spatial audio and head tracking*. Apple Support. <https://support.apple.com/en-us/guide/airpods/dev00eb7e0a3/web>
- Arsenault, D. (2005). *Dark Waters: Spotlight on Immersion*. GAMEON-NA International Conference, Montréal, Québec, Canada.
- Arth, C., Grasset, R., Gruber, L., Langlotz, T., Mulloni, A., & Wagner, D. (2015). *The History of Mobile Augmented Reality: Developments in Mobile AR over the last almost 50 years* (arXiv:1505.01319). arXiv. <https://doi.org/10.48550/arXiv.1505.01319>
- Azuma, R. T. (1997). *A Survey of Augmented Reality*. 6(4), 355–385. <https://doi.org/10.1162/PRES.1997.6.4.355>
- Bederson, B. B. (1995). Audio augmented reality: A prototype automated tour guide. *Conference Companion on Human Factors in Computing Systems - CHI '95*, 210–211. <https://doi.org/10.1145/223355.223526>
- Bleistainer, W., Bogaards, N., Meier, M., Färber, N., & Baume, C. (2018, May 30). *Object-based broadcasting – for European leadership in next generation audio experiences—D2.5: Final Pilot Progress Report*. Orpheus.
- Bollmer, G. (2017). Empathy machines. *Media International Australia*, 165(1), 63–76. <https://doi.org/10.1177/1329878X17726794>
- Brookes, C., & Deming, E. (2018). *Consent: Walk the Walk*. <http://batteryradio.weebly.com/landscape-journalism.html>
- Canadian Telecommunications Association. (2023, January 19). *Canadians among global leaders in internet usage and smartphone ownership, Pew Research Center study shows*. <https://canadatelecoms.ca/news/canadians-among-global-leaders-in-internet-usage-and-smartphone-ownership-pew-research-center-study-shows/>
- Caserman, P., Martinussen, M., & Göbel, S. (2019). Effects of End-to-end Latency on User Experience and Performance in Immersive Virtual Reality Applications. In E. Van Der Spek, S. Göbel, E. Y.-L. Do, E. Clua, & J. Baalsrud Hauge (Eds.), *Entertainment Computing and Serious Games* (Vol. 11863, pp. 57–69). Springer International Publishing. https://doi.org/10.1007/978-3-030-34644-7_5

- Castillo-Ramón, R. E., Camacho-Ninayahuar, A. G., Turriate-Guzman, A. M., & Bravo-Guevara, D. R. (2023). Immersive Journalism: A Systematic Review of the Literature in Scopus over the Last Five Years. *2023 IEEE 6th International Conference on Computer and Communication Engineering Technology (CCET)*, 48–52. <https://doi.org/10.1109/CCET59170.2023.10335133>
- CBC. (2020, September 25). Mistrial declared in RNC officer Doug Snelgrove sex assault case. *CBC News*. <https://www.cbc.ca/news/canada/newfoundland-labrador/doug-snelgrove-mistrial-1.5738599>
- Chainon, J.-Y., & Chainon, C. (2018). *For Editors & Newsroom Managers*.
- Cummings, J. J., & Bailenson, J. N. (2016). How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology*, 19(2), 272–309. <https://doi.org/10.1080/15213269.2015.1015740>
- David. (2025, October 8). The Future of Augmented Reality: Where AR is Headed in 2026. *AUGmentecture Blog*. <https://www.augmentecture.com/blog/the-future-of-augmented-reality-where-ar-is-headed-in-2026-and-beyond/>
- De La Peña, N., Weil, P., Llobera, J., Spanlang, B., Friedman, D., Sanchez-Vives, M. V., & Slater, M. (2010). Immersive Journalism: Immersive Virtual Reality for the First-Person Experience of News. *Presence: Teleoperators and Virtual Environments*, 19(4), 291–301. https://doi.org/10.1162/PRES_a_00005
- De Marco, R., idee und klang, & av Edition GmbH. (2021). *Sound Scenography/ Klangszzenografie The Art of Designing Sound for Spaces*.
- Domínguez, E. (2017). Going Beyond the Classic News Narrative Convention: The Background to and Challenges of Immersion in Journalism. *Frontiers in Digital Humanities*, 4, 10. <https://doi.org/10.3389/fdigh.2017.00010>
- Dussert, M. (2019, September 4). Le son binaural, la révolution sonore qui va faire chavirer vos oreilles. *L'ADN Tendances & Mutations*. <https://www.ladn.eu/mondes-creatifs/son-3d-binaural-revolution-sonore/>
- FRAMED. (2015). *Usomo*. <https://usomo.de/>
- FRAMED. (2018). Usomo Audiosystem in Sounds of Silence. *usomo*. <https://usomo.de/projekt/2018-museum-fuer-kommunikation-bern/>
- Furht, B. (2011). *Handbook of augmented reality*. Springer.
- Gabo, A., & Milk, C. (2015). Clouds Over Sidra. *MIT - Docubase*. <https://docubase.mit.edu/project/clouds-over-sidra/>
- Harju, M. (2021). *Exploring narrative possibilities of audio augmented reality with six degrees of freedom*.
- Harju, M. (2025). *Audio Augmented Reality: Concepts, Technologies and Narratives* (1st ed.). Focal Press. <https://doi.org/10.4324/9781003627289>
- Holland, S., Morse, D. R., & Gedenryd, H. (2002). AudioGPS: Spatial Audio Navigation with a Minimal Attention Interface. *Personal and Ubiquitous Computing*, 6(4), 253–259. <https://doi.org/10.1007/s007790200025>

- Hong, J., He, J., Lam, B., Gupta, R., & Gan, W.-S. (2017). Spatial Audio for Soundscape Design: Recording and Reproduction. *Applied Sciences*, 7(6), 627. <https://doi.org/10.3390/app7060627>
- Jang, S.-Y., Park, J., Engberg, M., MacIntyre, B., & Bolter, J. D. (2023). RealityMedia: Immersive technology and narrative space. *Frontiers in Virtual Reality*, 4, 1155700. <https://doi.org/10.3389/frvir.2023.1155700>
- Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, 66(9), 641–661. <https://doi.org/10.1016/j.ijhcs.2008.04.004>
- Jørgensen, K. (2007). *'What are Those Grunts and Growls Over There?'* Computer Game Audio and Player Action [PhD Dissertation]. Copenhagen University.
- kentbye. (2019, December 12). #847 DocLab: Augmented Audio Tour with Duncan Speakman's "Only Expansion." *Voices of VR Podcast*. <http://voicesofvr.com/847-doclab-augmented-audio-tour-with-duncan-speakmans-only-expansion/>
- Kim, P. (2018, November 28). An exploration of silence, in a new exhibition in Switzerland. *CDM Create Digital Music*. <https://cdm.link/sounds-silence-exhibition-switzerland/>
- Kool, H. (2016). *The Ethics of Immersive Journalism: A rhetorical analysis of news storytelling with virtual reality technology*. 9(3).
- Krzyzaniak, M., Frohlich, D., & Jackson, P. J. B. (2019). Six types of audio that DEFY reality!: A taxonomy of audio augmented reality with examples. *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound*, 160–167. <https://doi.org/10.1145/3356590.3356615>
- McGill, M., Brewster, S., McGookin, D., & Wilson, G. (2020). Acoustic Transparency and the Changing Soundscape of Auditory Mixed Reality. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16. <https://doi.org/10.1145/3313831.3376702>
- McMAHAN, A. (2003). *Immersion, Engagement, and Presence*.
- Milgram, P., & Kishino, F. (1994). *A Taxonomy of Mixed Reality Visual Displays*. E77-D(12), 1321–1329.
- Mullin, M. (2021, May 15). 'Justice has been served,' Crown says after Snelgrove found guilty of sexual assault. *CBC News*. <https://www.cbc.ca/news/canada/newfoundland-labrador/snelgrove-trial-may-15-1.6028320>
- Museum für Kommunikation. (2018). *Sounds of Silence Videorundgang*. Website des Museums für Kommunikation Bern. <https://www.mfk.ch/austauschen/digitaler-besuch/videorundgang-sounds-of-silence>
- Ouzounian, G. (2020). *Stereophonica: Sound and space in science, technology, and the arts*. The MIT Press.
- Paíno-Ambrosio, A., & Rodríguez-Fidalgo, M.-I. (n.d.). *A proposal for the classification of immersive journalism genres based on the use of virtual reality and 360° video*.
- Paul, S. (2009). Binaural Recording Technology: A Historical Review and Possible Future Developments. *Acta Acustica United with Acustica*, 95(5), 767–788. <https://doi.org/10.3813/AAA.918208>
- Perovich, L. J. (n.d.). *Combining art, science, technology, and community to move towards collective action on environmental challenges*.

- Pillai, S. E. (n.d.). *The Texture of Interiority: Voiceover and Visuals*.
- Reid, J., Cater, K., Fleuriot, C., & Hull, R. (2005). *Experience Design Guidelines for Creating Situated Mediascapes*.
- Rodero, E., Mas, L., & Larrea, O. (2015). *Use of sound and speech in the integrated journalism age. Shaping the future of News Media* [BOOK].
https://www.researchgate.net/publication/280568295_Use_of_sound_and_speech_in_the_integrated_journalism_age_Shaping_the_future_of_News_Media
- Sanchez-Acedo, A., Carbonell-Alcocer, A., Gertrudix, M., & Rubio-Tamayo, J. L. (2023). Metaverse and Extended Realities in Immersive Journalism: A Systematic Literature Review. *Multimodal Technologies and Interaction*, 7(10), 96. <https://doi.org/10.3390/mti7100096>
- Sanchez-Vives, M. V., & Slater, M. (2005). *From Presence Towards Consciousness*.
- Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535), 3549–3557. <https://doi.org/10.1098/rstb.2009.0138>
- Slater, M., Banakou, D., Beacco, A., Gallego, J., Macia-Varela, F., & Oliva, R. (2022). A Separate Reality: An Update on Place Illusion and Plausibility in Virtual Reality. *Frontiers in Virtual Reality*, 3, 914392. <https://doi.org/10.3389/frvir.2022.914392>
- Smellie, S. (2021, December 11). *Newfoundland police officer sentenced to four years for sexually assaulting woman while on duty*. <https://www.theglobeandmail.com/canada/article-newfoundland-police-officer-sentenced-to-four-years-for-sexually/>
- Speakman, D. (2019). *Only Expansion*. <https://duncanspeakman.net/projects/oe/>
- Speakman, D. (2021). *TEMPORAL STRUCTURES OF AUDIO AUGMENTED REALITY*.
- Speakman, D. (2025, June 17). *Archive Expansion—Counterfactual Mixtape*.
<https://soundcloud.com/duncanspeakman/sets/the-counterfactual-mixtape>
- St Clair, J. (2018). *Stories that walk with you—Opportunities in locative audio for feature journalism*.
- St Clair, J. (2019). Consent: Objectivity and the Aesthetics of Re-enactment in Locative Audio Journalism about a Sexual Assault Trial. *RadioDoc Review*, 5(1). <https://doi.org/10.14453/rdr.74>
- Steensen, S. (2011). The Featurization of Journalism. *Nordicom Review*, 32(2), 49–61.
<https://doi.org/10.1515/nor-2017-0112>
- Tran, T. T. M., Yoo, S., Weidlich, O., Cao, Y., Yu, X., Cheng, X., Ye, Y., Gulbransen-Diaz, N., & Parker, C. (2026). *Envisioning Audio Augmented Reality in Everyday Life*.
<https://doi.org/10.1145/3772318.3790620>
- Uskali, T., Jones, S., Sirkkunen, E., & Gynnild, A. (2021). *Immersive Journalism as Storytelling: Ethics, Production, and Design* (1st ed.). Routledge. <https://doi.org/10.4324/9780429437748>
- Van Der Nat, R., Müller, E., & Bakker, P. (2023). Navigating Interactive Story Spaces. The Architecture of Interactive Narratives in Online Journalism. *Digital Journalism*, 11(6), 1104–1129.
<https://doi.org/10.1080/21670811.2021.1960178>

- Vicente, P.-N., & Pérez-Seijo, S. (2022). Spatial audio and immersive journalism: Production, narrative design, and sense of presence. *El Profesional de La Información*, e310519.
<https://doi.org/10.3145/epi.2022.sep.19>
- Volpe, J. (2015, January 24). *The Godmother of Virtual Reality: Nonny de la Peña*. Engadget.
<https://www.engadget.com/2015-01-24-the-godmother-of-virtual-reality-nonny-de-la-pena.html>
- Wiltburg, J. (2004). *True Crime: The Origins of Modern Sensationalism*.
- Wincott, A. (2024). Un Bon Voyage Sonore: Avoiding Listener Discomfort in Immersive Audio Documentary. *Journal of Radio & Audio Media*, 31(2), 357–373.
<https://doi.org/10.1080/19376529.2023.2225062>
- Wincott, A., Martin, J., & Richards, I. (2020). *State of the art and future needs of spatial audio journalism*.
- Wincott, A., Martin, J., & Richards, I. (2021). Telling stories in soundspace: Placement, embodiment and authority in immersive audio journalism. *Radio Journal: International Studies in Broadcast & Audio Media*, 19(2), 253–270. https://doi.org/10.1386/rjao_00048_1
- Witmer, B. G., & Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3), 225–240.
<https://doi.org/10.1162/105474698565686>
- ZMIK. (2018). *Sounds of Silence—Zwischen Stille und Lärm*. <https://www.zmik.ch/de/projekte/sounds-of-silence>

6 Catalogue of Works

Brookes, Chris, and Emily Deming (2018) *Consent – Walk the Walk*

Geo-locative audio walk

St. John's, Canada: Battery Radio

<http://batteryradio.weebly.com/landscape-journalism.html>

Accessed 2026-06-03

Museum für Kommunikation (2018) *Sounds of Silence*

Exhibition with spatial audio guide

Bern: Museum für Kommunikation

Created with Idee und Klang, Bettina Mittelstrass, ZMIK, and FRAMED

<https://www.mfk.ch/austauschen/digitaler-besuch/videorundgang-sounds-of-silence>

Accessed 2026-06-08

Speakman, Duncan (2019) *Only Expansion*

Audio augmented reality audio walk

First presented at IDFA DocLab

<https://duncanspeakman.net/projects/oe/>

Accessed 2026-06-17

7 List of Figures

Figure 1	<i>Orpheus</i> Reference architecture Infrastructure Bleistainer, W., Bogaards, N., Meier, M., Färber, N., & Baume, C. (2018) <i>Object-based broadcasting – for European leadership in next generation audio experiences—D2.5: Final Pilot Progress Report</i> . p. 10	16
Figure 2	<i>Orpheus</i> iPhone application https://orpheus-audio.eu/free-orpheus-app-for-object-audio-based-experience-published/index.html accessed 2026-08-12	17
Figure 3	Reality – Virtuality Continuum by Milgram and Kishino https://www.researchgate.net/figure/Milgram-and-Kishinos-Mixed-Reality-on-the-Reality-Virtuality-Continuum-Milgram-and_fig1_321405854 accessed 2026-08-13	19
Figure 4	Graphical layout of the exhibition <i>Sounds of Silence</i> https://www.zmik.ch/de/projekte/sounds-of-silence accessed 2026-08-14	32
Figure 5	Top view of the exhibition space of <i>Sounds of Silence</i> https://www.zmik.ch/de/projekte/sounds-of-silence accessed 2026-08-14	34
Figure 6	First station of <i>Sounds of Silence</i> introducing participants to silence https://www.zmik.ch/de/projekte/sounds-of-silence accessed 2026-08-14	36
Figure 7	Participant of <i>Only Expansion</i> with modified headphones and booklet https://duncanspeakman.net/projects/oe/ accessed 2026-08-10	40
Figure 8	Booklet, playback unit and headphones used in <i>Only Expansion</i> https://duncanspeakman.net/projects/oe/ accessed 2026-08-10	42
Figure 9	Map of <i>Consent – Walk the Walk</i> with numbered stations https://batteryradio.weebly.com/landscape-journalism.html accessed 2026-08-13	49

Declaration of Autonomy Master-Thesis

I hereby confirm that I have written this Master's thesis entitled "The Potential of Audio Augmented Reality for Journalistic Narratives" independently and without outside help.

Any content taken from other sources such as texts, images, audio, graphics, software, etc., whether verbatim or analogous, is correctly cited with full attribution of authorship and source. In addition, all passages created with the help of AI-supported programs are clearly marked and provided with the exact name of the program used and the prompt applied. It is declared how AI tools were used for the translation of my own text, idea generation, brainstorming or similar.

Furthermore, I confirm that the thesis has not yet been published and has not been submitted in an identical or similar form as an examination or final project at another university, educational institution or in another degree program.

I acknowledge that a violation of these requirements may have legal and disciplinary consequences in accordance with § 26 Regulatory Framework for Bachelor's and Master's Degree Programmes at Zurich University of the Arts in conjunction with §§ 8 ff. Ordinance of the Universities of Applied Sciences Act.

With my signature I confirm the accuracy of this information:

First name: Demian

Last name: Wyssmann

Swiss matriculation number: 16-264-566

Date: 14.08.2026

Signature:

