Multimedia Tools and Applications (2025) 84:16105–16143
https://doi.org/10.1007/s11042-024-19288-4

The Role of Audio in Immersive Storytelling: a Systematic Review in Cultural Heritage

Alessandro Giuseppe Privitera[1] · **Federico Fontana**[1] · **Michele Geronazzo**[2]

Received: 1 December 2023 / Revised: 19 April 2024 / Accepted: 22 April 2024 /
Published online: 26 June 2024
© The Author(s) 2024

Abstract
Nowadays, Virtual and Augmented Reality technologies play a supportive role in many research fields. In cultural heritage, various examples are available, including storytelling and narratives, where they can provide immersive and enhanced experiences to visitors and tourists, especially for entertainment and educational purposes. This review aims to investigate the opportunities that soundscape design and advanced sonic interactions in virtual and augmented environments can bring to cultural heritage sites and museums in terms of presence, emotional content, and cultural dissemination. Nineteen-two papers have been identified through the PRISMA methodology, and a promising positive effect of sonic interaction on user experience in a virtual environment can be observed in various studies, notwithstanding a general lack of specific contributions on the use of sound rendering and audio spatialisation for improving such experiences. Moreover, this work identifies the main involved research areas and discusses the state-of-the-art best practices and case studies where sonic interactions may assume a central role. The final part suggests possible future directions and applications for more engaging and immersive storytelling in the cultural heritage domain.

Keywords Sonic interaction design · Virtual/augmented reality · Immersive audio · Storytelling · Cultural heritage

1 Introduction

Storytelling and narrative have traditionally been key for cultural heritage preservation throughout human history, embracing diverse contexts ranging from entertainment to culture and knowledge [1]. With the progressive evolution of communication means and multiplica-

✉ Alessandro Giuseppe Privitera
privitera.alessandrogiuseppe@spes.uniud.it

Federico Fontana
federico.fontana@uniud.it

Michele Geronazzo
michele.geronazzo@unipd.it

[1] Department of Mathematics, Computer Science and Physics, University of Udine, Udine, Italy

[2] Department of Engineering and Management, University of Padova, Padova, Italy

tion of the supports, stories started to be passed down not only verbally but also with graffiti, paintings, poetry, theatrical plays, musical pieces, and then books, photos, and films until the revolutionary carriers put available by the electronic and finally digital era.

In recent years, the use of technology within digital humanities has become essential for dissemination, education, and research in the cultural heritage field. In particular, augmented and virtual reality (AR and VR, which together form a common ground of mixed reality) are relevant enabling technologies and are nowadays widely used by both academic and corporate research as well as development in multiple application domains ranging from neuroscience to sports training, medicine, education, and entertainment [2]. Cultural heritage can also benefit from mixed reality experiences in a variety of areas, ranging from the reconstruction of cultural sites [3] to the creation of accessible knowledge made possible by technology-mediated direct experience [4, 5]. A *soundscape*, the acoustic real or virtual environment as perceived or experienced and as understood by a person, in context, gives a valuable contribution to tourism and cultural heritage enhancing cultural proposals [6]. In this context, immersive storytelling refers to the use of mixed reality techniques in storytelling, with the goal of eliciting a deeper sense of immersion and presence. Mixed reality offers innovative ways for storytelling, providing immersive experiences that can captivate and emotionally engage audience and facilitate learning [7]. Moreover, it introduces new affordances that allow more ecological interactive experiences, as suggested by various studies (e.g., [8, 9]). Heightened engagement, emotional connection, and enhanced interaction contribute to increased interest in the subject, thereby enhancing knowledge fruition and dissemination [7]. Audio is a key element in allowing users to become more immersed in the virtual experience and can also be an important channel for information about the environment [10, 11].

Focusing on cultural heritage, disparate studies, and reviews are available on mixed reality storytelling [12, 13], audio [6] and technology for audio AR [14]. However, very few comprehensive contributions that deal with these aspects altogether have been found. Some recent studies put a general focus on audio mixed reality and storytelling [15], but no systematic review is available about the specific topic of cultural heritage, especially including audio as a key aspect.

The insights presented here aim to define a state of the art in the available interactive audio virtual environments for cultural heritage storytelling, and to illustrate the most useful techniques for enabling audio mixed reality experiences by reviewing the current literature in the field. Additionally, we discuss key elements that create an immersive experience in cultural heritage fruition, as well as how interactivity and personalized, emotional content can be used for educational purposes and knowledge dissemination. Finally, we investigate how audio mixed reality can help storytelling in cultural heritage and the limitations of the current research in discovering and proposing new forms of advanced interactions. From here, we will outline some promising directions. In particular, along with a systematic review of available works, we propose the following research questions:

- Q1. What enabling technologies and platforms are available for immersive storytelling in audio mixed reality for cultural heritage? These topics will be discussed in Section 3
- Q2. What are the available works? What insights does the literature research provide about storytelling in audio mixed reality for cultural heritage? The theme will be analysed in Sections 4 and 5.
- Q3. Which aspects of User eXperience (UX) and interaction are being studied with regard to storytelling in audio mixed reality for cultural heritage research? The reader will find the topics in Section 5.

– Q4. Which implications can be drawn from the literature regarding the design of future research? These matters will be discussed in Sec. 6.

In order to identify a common characterization and to provide some answers to the questions at hand, the paper is organised as follows: Section 2 introduces the most important concepts and surveys state of the art. Section 3 illustrates tools and technological platforms available from the design and implementation of storytelling in audio mixed reality for cultural heritage. Section 4 describes the PRISMA review method and criteria used for the selection of works. Sections 5 and 6 present analysis results and a discussion based on previously described criteria. Finally, Section 7 discusses the results and proposes some research insights and application scenarios.

2 Background

In the context of cultural heritage storytelling, audio rendering, and mixed reality have been introduced and investigated since the seventies. In one of the early theoretical examples, the idea of enhancing the experience with multisensory elements was introduced by Youngblood [16]. The author coined the term *expanded cinema* to describe a form of art that goes beyond traditional cinematic media. This includes special effects, computer art, and multimedia elements. One of the first prototypes for audio augmentation in VR [17] was an automated audio guide, permitting the inclusion in the environment of synthetic audio based on the visitors' position to avoid isolation and enhance social interaction during museum visits. In another seminal study [18], the authors introduced a classification of four categories of museum visitors based on physical navigation, artwork enjoyment, and information browsing for creating an interactive audio guide prototype in AR. The model records the user's physical movement inside the museum to dynamically classify visitors with a non-intrusive approach. Finally, in a pioneering work [19] AréViJava was presented as one of the first platforms for virtual tourism in which an avatar guides tours along virtual places. In particular, the authors describe the design process for the reconstruction of the Brest harbour site (France) as it was in 1810, enabling a virtual guided tour. During the tour, different viewpoints were suggested by the virtual guide, together with additional audio and video documents accessible through a website.

In most such works, audio is an integral part of a broader multi-sensory perspective (audio/visual, audio/haptic, audio in service of system feedback) or just one element part of a complex systemic experience, in which its specific contribution is difficult to evaluate, or no audio-specific evaluation is considered. In such a context, we believe that a systematic study highlighting the contribution of audio and mixed reality in cultural heritage storytelling would be of help to improve the sonic interaction design of cultural experiences and improve research in the field.

2.1 Sonic interactions in virtual environments: The immersion-coherence-entanglement model

This review exploits the theoretical and philosophical lens offered by the Sonic Interactions in Virtual Environments (SIVE) new field of study, which refers to the human-computer interplay through auditory feedback: "the study and exploitation of sound being one of the principal channels conveying information, meaning, aesthetic and emotional qualities in

immersive and interactive contexts" [20]. SIVE provides a framework for the investigation and identification of advanced interaction opportunities. The authors propose three top-level categories that need to be addressed through an interdisciplinary design work: Immersion, Coherence, and Entanglement. In the AR/VR research community, these terms have multiple definitions and slightly different meanings. In this work, we refer to the terms as defined in the SIVE framework.

According to Slater's definition of immersion [2], two key concepts are introduced concerning the capture of subjective internal states: plausibility illusion and place illusion. The former determines the overall subjective credibility of a virtual environment; the latter is intended as the quality of a simulation in providing the sensation of "being in a real place" that can be crucial in various cultural heritage experiences, such as museum exhibitions [21]. **Immersion**, hence, can be defined as "the degree in which the range of sensory channels is engaged by the virtual simulation" [22]. This degree measures the technological level and its enactive potential. In mixed reality, audio is fundamental in creating a sense of immersion [23]. Immersive mixed reality should be designed in an egocentric perspective [20, 24], referring to the coordination of the perceptual/cognitive individuality of the user's self, identity, or consciousness with multi-sensory information processed by the technological system.

Coherence concerns the plausibility of the rendering, the interactions, and possible behaviours in the virtual environment in realistic fictional experience [25]. It measures the effectiveness of the sonic interaction design and is concerned with various factors, including subjective expectations [26] and social rules [27].

Entanglement [28] describes the effectiveness of the overall sonic experience in terms of dynamic and mutual adaptations among its key actors. It measures the level of active participation of the user, the technology, and the content in what is called the *locus of agency* [29]. This refers to a meta-environment with technological and digital features in which each actor involved in the experience (including the user and the technological platform) can act. According to Frauenberger's research in entanglement human-computer interaction [30], the design of computers and interaction design cannot be approached directly. Instead, the focus should be facilitating specific configurations that bring about certain phenomena. The term "agency" denotes a performative mechanism that constructs one's sense of self by establishing boundaries. The key principle is the shift from *inter-action* between defined objects to *intra-action* within a phenomenon, where the boundaries between actors are fluidly determined within a system, i.e., a *locus*, similar to the Gibsonian ecological theory of perception [31]. In an egocentric perspective, the *locus of agency* takes shape around the listener or the natural world that is meaningful to them.

2.2 Cultural heritage storytelling

Some notable efforts to come up with guidelines for effective storytelling recently started: according to The Center for Digital Storytelling in Berkeley[1], digital narratives should be designed following seven specific criteria:

- **Point of View** of the author
- **Gift of voice,** and the register of the narration: colloquial, formal, etc.
- **Dramatic Question**, the intrinsic message conveyed by the story
- **Emotional Content**, the emotions that are transmitted by the narrator

[1] https://www.storycenter.org/

- **Power of Soundtrack**, audio elements in the narration
- **Economy**, the amount of content and information conveyed by the narrative
- **Pacing**, the rhythm of the story, in terms of time storytelling structure

Scientific literature has widely adopted criteria whose primary purpose was education (e.g. [32], for the analysis of autobiographical literacy).

With a specific focus on digital storytelling, Meadows [33] highlights four essential component:

- **Perspective**: the perspective that the author wants to convey with the story consisting of emotions, presentation, and the process of encoding/decoding
- **Narrative**: the story that is narrated by the storytelling in the specific medium
- **Interactivity**: a peculiar characteristic of digital media that can be implemented, e.g., with the design of multiple storylines or choices that the user can do
- **Medium**: storytelling message interpretation can be strongly influenced by the type of medium used [34].

No specific framework for digital storytelling evaluation has been found in the literature, especially in the audio domain, but some efforts have been made in this direction. Sitters et al. [35] explore digital storytelling in the health research domain, suggesting that storytelling should be evaluated for its validity in four different ways. Storytelling material should convey emotions (empathic validity), be credible (intersubjective validity), be sound and just (ethical validity), and have stakeholders play an active part in the design process (participatory validity). This last aspect is also underlined by many other studies, including a framework that considers three different dimensions for the design and evaluation [36]: aesthetics, cognition, and sociality. In sound-driven design, Dalle Monache et al. [11] argue that participatory design should involve stakeholders of different fields in order to develop a deeper understanding of the media and its potential applications. Moreover, we would like to emphasize the perspective of Sonic Interaction Design (SID, [37]), which considers sound as a primary channel for conveying not only information and meaning but also aesthetics and emotional qualities in interactive contexts. SID goes beyond mere information transmission. It encloses the wholeness of the user experience, enriching it with auditory cues that enhance immersion and engagement. In this context, storytelling can be used to enhance the overall experience, and the SIVE framework aims to help understand how SID principles contribute to crafting compelling narrative experiences. Furthermore, it is important to underline the differences between music and sound. Indeed, sound encompasses a rich spectrum of elements that contribute to the overall immersive experience, including aesthetic qualities and emotional elements. Since music is an art form with specific canons [38], storytelling through music can be considered a special form of storytelling that can be used to enhance the sonic environment or can be enhanced with other sonic or multimodal elements. In this work, we have recognized music's important role in storytelling and included it as a key component. To maintain the inclusive breadth of our work, however, we have presented it as one of several options available. On the other hand, as specified in the two frameworks described above, music can be an integral part of the immersive storytelling experience.

Lugmayr et al. [39] introduce the concept of *serious storytelling*, namely storytelling designed with a specific purpose other than entertainment where "the narration progresses as a sequence of patterns impressive in quality, relates to a serious context, and is a manner of thoughtful process".

Cultural heritage storytelling can be considered a specific case of serious storytelling with some interesting peculiarities. The most important requirements of storytelling for the cultural heritage domain are preserving the correct reconstruction of historical and artistic

aspects and conveying a non-ambiguous message. Various works use storytelling techniques to enhance visitor experience in cultural heritage sites or museums, e.g., navigating through a site and following a story. In this kind of situation, personal experiences and emotions can influence information comprehension and interpretation [40] and how users infer and interpret a story [39]. In other words, storytelling can be used to convey cultural heritage notions and concepts for educational purposes effectively. The Storytelling creation process itself can also be used to help understand concepts and historical timelines, e.g., in classroom experience [41]. In this case, the storytelling design process should be correctly used to convey cultural heritage knowledge.

Another important aspect in various cultural heritage-related applications, such as tourism and education, is engagement. Storytelling can be a powerful means for transmitting knowledge, but it should be designed in such a way as to correctly fulfill the task, convey the desired message, and avoid errors in the interpretation [12]. In other words, storytelling for cultural heritage can include engaging elements and gamification, however such elements should not be the primary aim. Moreover, Podara et al. [42] suggests that interactive and non-linear storytelling can be useful to elicit engagement, especially in younger people. On the other hand, the amount and type of interactive elements should be carefully calibrated to avoid the risk that users lose interest in the storyline or in the storytelling message [43]. Again, no comprehensive framework can be found to evaluate serious storytelling. However, various tests are available to evaluate specific elements such as engagement, memorability, cognitive load, etc. The method used in the reviewed paper is discussed in Sec. 5.5.

3 Audio tools and technological platforms

The analysis of the reviewed papers shows a very diverse situation in terms of frameworks and technologies used. Thus, it is difficult to identify the main trends and durable tools over the years, especially when considering open-source or reusable solutions without license fees. Most of these technologies do not find practical applications in specific cultural heritage contexts but serve as valuable insights into the landscape of available technologies within cultural heritage but may not directly contribute to identifiable trends or prevalent practices. However, many recent immersive storytelling applications were realized using commercial software (Unity3D[2] or Unreal Engine[3]) or free SDKs (Vuforia, ARKit, AR Foundation[4]) during the design and development stages, while some examples of custom solutions can be found for specific purposes.

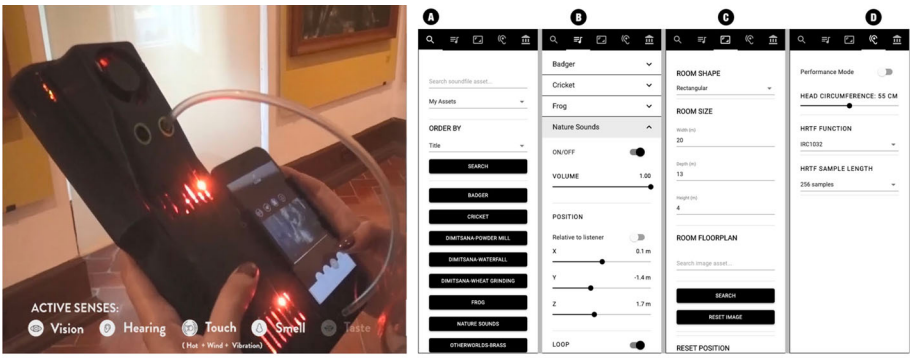
Platforms specifically designed for exploring spaces, providing navigation and spatial audio are described in various studies [46–49]. In particular, AVIE is a system for directional audio in VR theaters [50], while other audio technologies for spatialisation in theatres are included in Beck's compendium [51] along with a historical overview.

Various technologies are available for creating personalized sonic experiences, using both multi-channel speakers for room audio spatialization (Ambisonics, Dolby Atmos, etc.) and headphones for mobile applications. With the latter technologies, head and body movement tracking along with dynamic binaural audio technology are crucial for auralisation, i.e., the mathematical modeling of physical audio sources in space [52] convincingly localizing a sound source for attentional and navigation purposes [53]. User movement can be tracked by

[2] https://unity.com/

[3] https://www.unrealengine.com/en-US/

[4] https://unity.com/unity/features/arfoundation



(a) (b)

Fig. 1 Examples of technological platforms: a) M5SAR platform, audio guide with the elicitation of all the five senses photo © [44] b) screenshot of PlugSonic user app. photo © [45]

various types of sensing devices. For outdoor position tracking, GPS information is used [54] often combined with inertial measurement units (IMU) [46, 55], and gyroscopes [56] are used for determining the indoor precise position and/or head orientation.

Audio is combined in a multimodal interaction through the five senses in the M5SAR platform (Fig. 1-a) [44, 57], by using a custom-built mobile device along with a mobile app that recognizes artwork in a museum. PlugSonic [45] (Fig. 1-b) is a web-based platform for spatial audio storytelling in cultural heritage. Another storytelling design tool is MOGRE [58, 59], a software aimed at the creation of 3D scenarios and stories based on the Ogre[5] graphics engine for children.

Ec(h)o [60, 61] is an engine for audio material classification and real-time search based on user position tracking and preferences. Proper audio material, classified by using a given ontology, is inserted in a database and presented in real time to the user who interacts with a tangible interface (cube) while moving. Practical guidelines and conceptual frameworks are available on cultural experience design. For example, "Adaptive Augmented Reality" (A2R, [62]) is an AR museum guide architecture that recommends content based on gestural and biometric information to elicit interest and engagement during museum visits. Katz et al. [63] suggest that acoustic space is an important aspect to consider in communicating acoustic heritage and producing audio experiences. Moreover, they introduce *Past Has Ears(PHE)*, a hardware/software prototype for the presentation of immersive audio experiences adaptable to multiple platforms. Polaris∼ [64] is a wearable platform creating privacy-respectful audiovisual AR experiences to foster artistic and musical expression. Polaris∼ comprises an open-source AR headset (Project North Star), a pair of bone-conduction headphones, and software developed in Unity and PureData. Meyer [65] suggests guidelines for story design in a spatial context, based on a literature review, suitable for interactive audio-visual narrative and 360° films. In his work, Fujinawa et al. [66] hypothesized that a comfortable sound field could induce different behaviours in human movement and navigation in a physical space. In a laboratory study, users are left free to move in a room where sound is diffused with different pressure levels by motor-controlled moveable speakers in three different audio stimuli conditions (white noise, Jazz music, no sound). By using questionnaires and position tracking, user preferences and time spent in areas with different loudness levels have been analyzed, overall showing longer staying in more quiet areas when audio stimuli were evalu-

[5] https://www.ogre3d.org/

Table 1 Inclusion Criteria

	Description
Inclusion Criteria 1	Article and book chapters related to Audio Virtual and Augmented Reality Storytelling in Cultural Heritage
Inclusion Criteria 2	Work from Journal Paper or Conference Proceedings
Inclusion Criteria 3	Paper, Article or Book written in English

ated as unpleasant. Kenderdine et al. [67] describe the potential application of computational practices within archival and museological domains that could improve the preservation and cataloging of sources, provide new forms of representation and knowledge, and can empower new forms of art. In this direction, *Jazz Luminaries*[6], and interactive installation that display the connection between jazz, blues and Latin artists using a 3D visualization in a fulldome. Other contributions can be found in interactive musical performances such as *Carillon* [7] and *Membrana Neopermeable* [68] that uses a virtual environment used by the performers to interact with virtual instruments. In the supplementary materials, Appendix A provides a historical excursus of platforms and available studies regarding audio features, cultural heritage field, and storytelling capabilities.

4 Methods

While structuring this review, we followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) method [6] as a guide for our work. PRISMA is widely used in diverse research areas as medicine [69], psychology [70], and computer science [71]. The data search and paper selection process is described here.

4.1 Eligibility criteria

Selected papers or books must include research, reviews, or case studies about the use of virtual/augmented/mixed reality for cultural heritage storytelling, specifically referring to audio as a part of the dissertation. Table 1 lists the inclusion criteria. It is worth noting that the author found no work including all the considered topics, probably due to limitations binding the different fields. See, for example, the work in [72], claiming implementations of a mobile virtual environment through a PDA as a future possibility should technological advancements allow for, or "KidsRoom" [73] that describe an AR interactive experience for storytelling not specifically designed for cultural heritage.

4.2 Search strategy

Our research was conducted by performing an extensive search on online publication-related databases: Elsevier Scopus, ACM Digital Library, IEEE Xplore, and Google Scholar. In

[6] https://www.kirellbenzi.com/art/jazz-luminaries

[7] https://ccrma.stanford.edu/~rob/portfolio/carillon/

all these databases, four different clusters of keywords were used to include synonyms and related keywords. Presented below are the four names representing different groupings:

- **Storytelling:** storytelling, narrative, guide, guidance.
- **Audio:** sound, audio, acoustic, acoustics, auditory, sonic, hearing, soundscape, soundscapes, voiceover, narrator.
- **Immersive media:** Virtual Reality, Augmented Reality, VR, AR, Spatialized, Spatialised, Spatialization, Spatialisation, virtual environment, XR, extended reality, immersive media, metaverse, immersive space, immersion, virtual agent.
- **Cultural heritage:** cultural heritage, museum, history, tourism, culture, archaeology, archaeological, historic, conservation, preservation, tangible heritage, intangible heritage, living heritage.

Keywords were connected by using OR inside clusters AND between clusters. Search was performed in Title, Keyword, and Abstract in Scopus, IEEE, and ACM by using their specific language. Due to Google Scholar's limitation in search string length (256 characters), a different string format has been used to include a wider selection of papers. The string formats are specified in (see Appendix B in the supplementary materials).

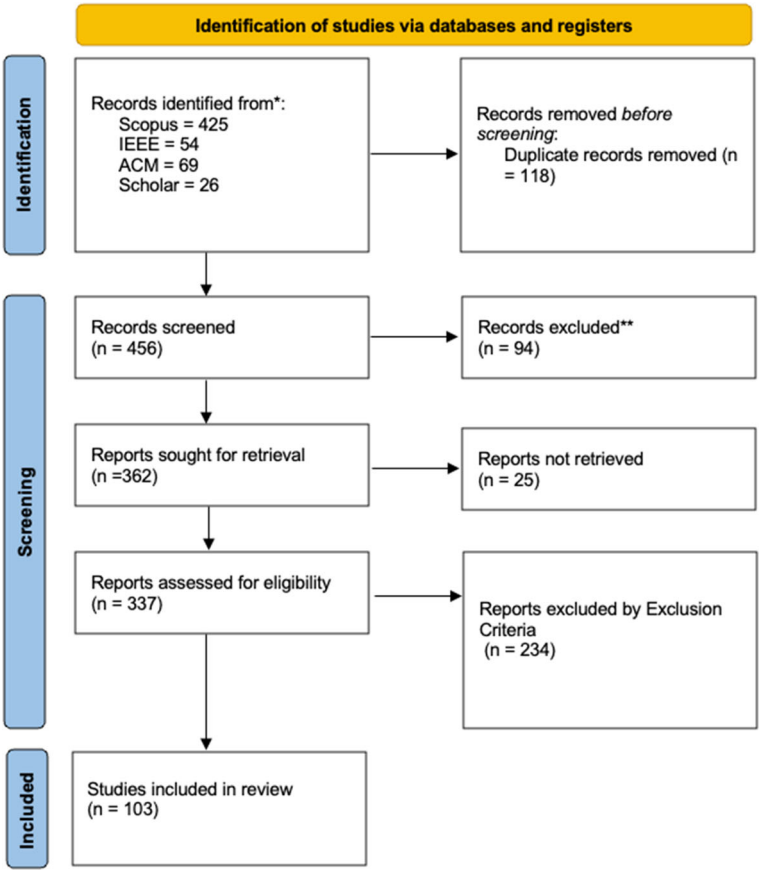


Fig. 2 PRISMA flow diagram

4.3 Study selection

Starting from the previously mentioned inclusion criteria, Fig. 2 depicts and characterizes the PRISMA phases that considered a total of 574 articles selected from 4 scientific databases. Duplicate papers (118 in total) were removed, and the remaining 456 papers were screened by title, abstract, and scanning using the exclusion criteria listed in Table 2. The selected papers must have a strong research focus on cultural heritage and Storytelling, therefore work with different main topics (e.g., technologies that can also be used for cultural heritage, general sound localization studies, generic AR platforms, solutions for urban navigation, etc.) and works in which one topic was absent have been excluded. Since the audio-sensory modality was particularly important for our systematic review, we introduced an ad-hoc classification:

- **Category 1**, Audio First: audio is the leading modality in the paper.
- **Category 2**, Audio and Multimodal Integration: audio is considered together with other modalities (e.g., haptics, video, etc.), but specific information is given.
- **Category 3**, Audio in Multimodal Comparison: audio experience is considered in comparison with others (e.g., Audio vs. Haptic feedback).
- **Category 4**, Multimodal Experience: audio is present, but it is not possible to ascertain its specific audio contributions.
- **Category 5**, Other experience: audio is present but not considered.
- **Category 6**, No Audio: audio is not present.

All papers with level 5 or 6 were excluded since they were of little or no interest for this review.

4.4 Data collection process

4.4.1 Type of immersion

A classification for mixed reality systems has been proposed by Bekele et al. [7], which introduces two different categories for AR systems and three for VR systems. Bekele et al. proposed a general classification based on system flexibility and experience. We exported this classification to auditory information:

- **Outdoor audio AR:** By using GPS or other markerless tracking techniques, people can navigate inside a vast open-air field and obtain location-aware audio information.

Table 2 Exclusion Criteria

	Description
Exclusion Criteria 1	Article or paper not purchasable/unavailable
Exclusion Criteria 2	Articles or papers are classified using the following categories and double-checked by authors. Articles in Category 5 and 6 are excluded from evaluation.
Exclusion Criteria 3	Not about cultural heritage but other aspects are described (e.g., a technology that can also be used for cultural heritage)
Exclusion Criteria 4	No Storytelling is involved
Exclusion Criteria 5	No VR

- **Indoor audio AR**: Indoor tracking needs precise information about head orientation and user in a limited space. The most used techniques are IMU, Gyroscope, and AR toolkits like ARKit[8] or Vuforia[9].
- **Non-Immersive audio VR:** The virtual environment is viewed from a desktop or screen, and audio content is conveyed from a speaker and not spatialized.
- **Semi-Immersive audio VR:** Audio content is conveyed through multichannel or directional speaker systems, usually in rooms with multiple users.
- **Immersive audio VR:** A user is fully immersed in an auralized virtual environment or audio content that is realistically integrated inside a physical environment, with a high level of presence.

4.4.2 Purpose of each study

Mixed reality can provide important support to cultural heritage for different purposes. Different surveys have already been published on this topic [74]. Based on these works, we identify the following main purposes for the use of immersive media technologies:

- **Education:** We identify systems designed with the aim of learning and dissemination, such as mixed reality books for children [75] and audiovisual storytelling experiences [76].
- **Exhibit enhancement:** A work about cultural heritage audio augmentation applied to tour visits [77] and an app for augmentation of a physical diorama [78].
- **Exploration:** Applications or methods for navigating or discovering spaces or contents, such as augmented audio guides [54, 79].
- **Reconstruction:** Re-creation of elements of the past inside today's world as the reconstruction of Fort Sant Jean [80] and AR interviews [81].
- **Virtual museums and interactive installations:** A virtual cultural heritage experience such as virtual drama [55].

4.5 Computer-aided Qualitative Data Analysis Software (CAQDAS)

As part of the research process, we conducted an automatic content analysis on the papers we considered. For the automatic content analysis we utilized *Leximancer*[10], a Computer-aided Qualitative Data Analysis Software (CAQDAS) that extracts statistical properties from a text to identify a list of terms and concepts [82]. The software identifies highly connected concepts and clusters them into higher-level groups, defined as themes.

Leximancer uses an analysis method that examines a unified body of text—in this case, reviewed articles. The program selects a ranked list of emerging lexical terms based on their frequency and co-occurrence usage. These terms are used for creating a thesaurus, which in turn builds a set of classifiers from the text by iteratively extending the seed word definitions. The result is a set of weighted term classifiers known as concepts. The text is then classified using these concepts at a high resolution—typically every three sentences. This produces a concept index for the text and a concept co-occurrence matrix. An asymmetric co-occurrence matrix is created by calculating the concepts' relative co-occurrence frequencies. This matrix produces a two-dimensional concept map using a proprietary clustering algorithm based on

[8] https://developer.apple.com/augmented-reality/

[9] https://developer.vuforia.com/

[10] https://www.leximancer.com/

Table 3 Eligible papers divided by purpose and audio involvement level

Study Purpose	Audio Involvement Level	Eligible Papers Ref.
Education	Audio First	[84, 85]
	Audio and Multimodal Integration	[75, 78, 86–90]
	Audio in Multimodal Comparison	[91–94]
	Multimodal Experience	[95–103]
Exhibition Enhancement	Audio First	[46, 55, 56, 77, 104–110]
	Audio and Multimodal Integration	[61, 111–120]
	Audio in Multimodal Comparison	[121, 122]
	Multimodal Experience	[123–130]
Exploration	Audio First	[54, 131–133]
	Audio and Multimodal Integration	[79, 134–138]
	Audio in Multimodal Comparison	[139]
	Multimodal Experience	[140, 141]
Reconstruction	Audio First	[142–147]
	Audio and Multimodal Integration	[76, 148–150]
	Audio in Multimodal Comparison	no paper available
	Multimodal Experience	[80, 81, 151]
Virtual Museums and Interactive Installations	Audio First	[152–156]
	Audio and Multimodal Integration	[157–166]
	Audio in Multimodal Comparison	[167]
	Multimodal Experience	[168–175]

the spring-force model for the many-body problem [83]. The correctness of each concept in this semantic network generates a third hierarchical dimension, which displays the more general parent concepts at the higher levels.

5 Analysis and results

A total of 103 papers have been considered eligible for the study. Works were read and analyzed to collect useful information about the study and disclose research trends. All contributions, available in Appendix C of the supplementary materials, have been classified based on the purposes of the study. Among all papers, 63 describe a cultural heritage site application, and 40 are theoretical or applied studies (Table 3).

One first important consideration concerns the role of audio components in the context of mixed reality storytelling in cultural heritage. Results in Fig. 3 show that 30 studies out of 103 have been specifically designed to evaluate auditory and 37 studies consider audio in a combined [54] or comparative [176] multimodal perspective. On the other hand, in 27% of the studies, it is not possible to determine a specific contribution of audio since this component is ambiguously considered in combination with video or other elements. However, the interest in auditory evaluation is increasing since the number of papers in the field has grown in the last few years (see Fig. 3). As also noted by Jerald [177], the interest in mixed reality is undoubtedly increasing, also thanks to new technologies and platforms available in the market.

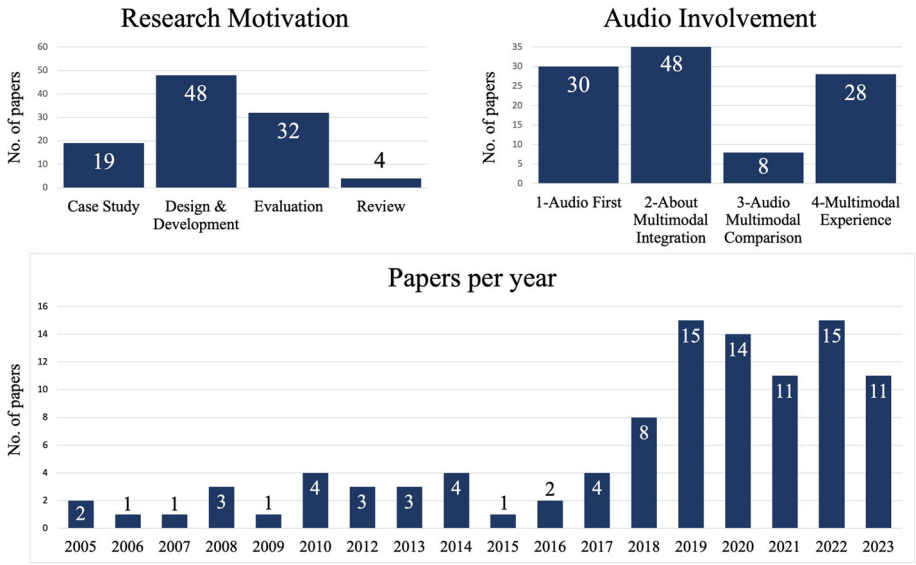


Fig. 3 Count of articles categorized by (top left) research motivation, (top right) audio involvement level, and (bottom) publication year of eligible studies

Figure 4 shows the main themes identified by CAQDAS analysis, and the main concepts considered:

- **Sound**: concepts regarding audio and spatialisation are included in the theme, e.g., *spatial, sound, position, environment, etc.*
- **Story**: represents storytelling and narrative elements including *narrative, emotional, music, media, history, landscape, world, etc.*
- **Experience**: it refers to the overall user experience and is very intersected with both content and narrative terms (such as *immersive, history, natural, create, cultural heritage, digital media*) and technological elements (*technology, interaction, interface*
- **Social**: includes concepts related to cultural heritage spaces and collaborative experiences, e.g., *urban, public, site, city, landscape, performance*
- **User**: centrality and need of personalisation of user in the SIVE experience are underlined in this theme that includes *user, audio, interaction, guide, visit*.

Storytelling is central to the mixed reality experience and requires careful design in order to improve user interest in different aspects (engagement, education, experience time, etc.) while maintaining the characteristics of effectiveness and correctness of information. Content has to be built around the user, following an egocentric perspective, and should be personalised respecting each person's particular traits and special needs. Technology must help make the experience realistic and plausible inside the virtual environment, including using collaborative elements. The main concepts discovered in CAQDAS analysis are discussed in greater detail in the following subsections.

5.1 Sound reproduction and spatialisation

Technology plays an important part in conveying a sense of presence in terms of Immersion, Coherence, and Entanglement dimension, corresponding to *Sound* CAQDAS concepts. A

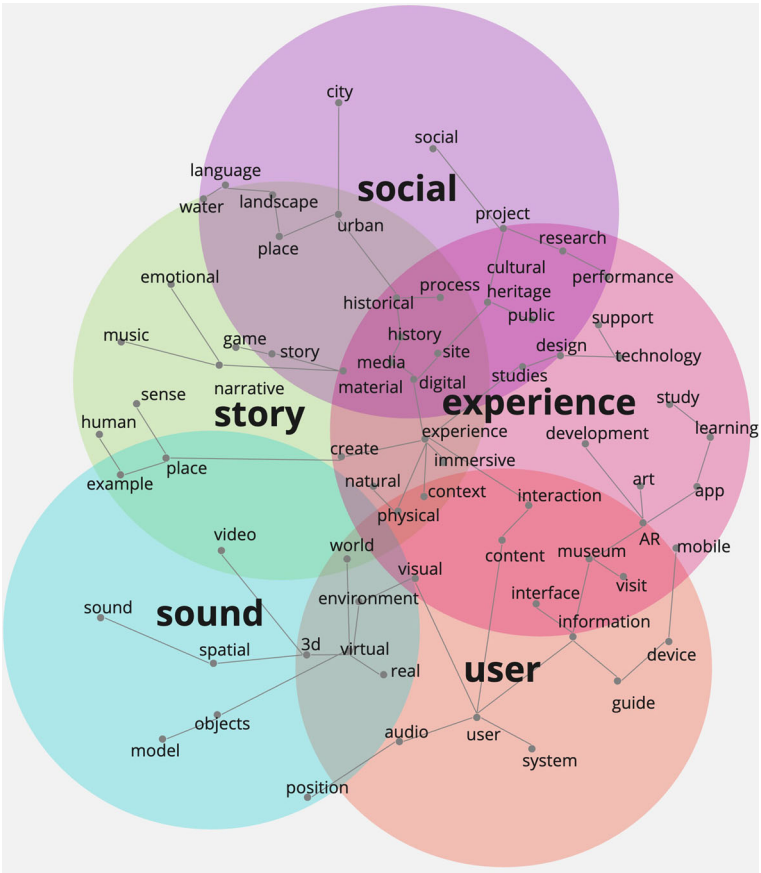


Fig. 4 Topical map of discovered concepts. Clusters are displayed as heated circles

plausible audio rendering requires several types of reproduction devices and spatialisation techniques [178], as it is also underlined by CAQDAS analysis, which includes *spatial* and *position* in the domain, as well as *real*, *virtual* and *place* emphasizing the interplay between all experience elements, whether tangible or not. Moreover, the intersection between *sound* and *user* underlines the importance of an egocentric audio perspective. The devices identified in the reviewed works have been divided into three different categories:

- **Headphones/earphones:** personal reproduction devices with two small speakers available on the market in different dimensions, ear fit (in-ear, outer-ear, over-ear), and connectivity (wired or wireless). Audio headphones can reproduce monophonic sound, diffusing the same audio information in both speakers, stereophonic sound accounting for a horizontally localized source, or binaural sound considering listener acoustics in the so-called head-related transfer functions (HRTF, [179]). Of particular interest here, stereophony conveys a sense of direction, especially if in conjunction with spatialisation models simulating a physical environment [180]. Moreover, audio headphones can isolate totally or partially from the external acoustic environment (audio transparency [181]).
- **Directional speaker:** a particular type of loudspeaker with a narrow acoustic beam. Directional audio speakers provide personal content while maintaining environmental

audio transparency [112], with performance limits for listeners who occupy areas outside a known sweet spot.

– **Multichannel:** a loudspeaker set conveying spatial audio and permitting sound source localisation. Multichannel audio can be composed of similar or different speakers in terms of frequency reproduction (frequency response). In the reviewed papers, some particular multichannel sets have been identified in two works [55, 157]. The former uses four bone conduction speakers installed in a headband to create audio spatialisation and orient the user in a cultural site using GPS and compass sensors. The latter uses a combination of custom-built headphones and room speakers to enable a social experience in a theatre (more information is given later in this chapter).

Information about audio reproduction and spatialisation are available in Fig. 5. In the considered studies, the most widely used audio reproduction devices are headphones/earphones, which, as opposed to other solutions like mobile phone integrated speakers, can limit social/collaborative interaction inside museums or cultural sites [182] and modify the presence and immersion level of the system [183]. However, it is important to highlight that no sound reproduction device is specified in 48 studies (46.7%), even though it could be an important aspect of experience evaluation, especially in studies based on portable devices.

It is worthwhile to note that audio spatialization and algorithm information are still sparsely treated, but interest in the topic is increasing. In fact, only 35 studies involve a technology making use of spatial audio or more than two speakers. Interestingly, 18 such studies have been published recently as in the last two years.

Commercial headphone equalizers based on HRTFs (Unity or Unreal Engine with SteamAudio or FMOD) are used in various works [46, 104, 116, 133, 144–146, 153, 164, 165, 175, 184–186]. Other headphones, commercial or open-source headphone-based rendering models, and commercial or open-source libraries have been considered in reviewed papers: OpenAL [108, 109, 142], Audiokinetic WWise [155], Apple XCode [175], Processing [130] or not specified [56, 110, 131, 143, 150, 159, 174, 187]. Andolina et al. [134] uses a "headset capable of 3D audio".

In works with a multichannel setup, SensiMAR [148] uses 4-speaker Ambisonics 3D audio format, a method for reproducing audio on a spherical spatial sound-field [188], in outdoor Conimbriga archaeological site to recreate the soundscape of ancient Roman activity. "Trying to get trapped in the past" [163] uses Wave Field Synthesis on a linear loudspeaker array to spatially render a virtual theatrical play. "Pigments" [166] use Dolby Atmos for the sonification of *Pigments of Imagination* audiovisual installation. Finally, a Dolby 5.1 system is used for the sonification of Roal Dahl's "The Time Machine" story [162].

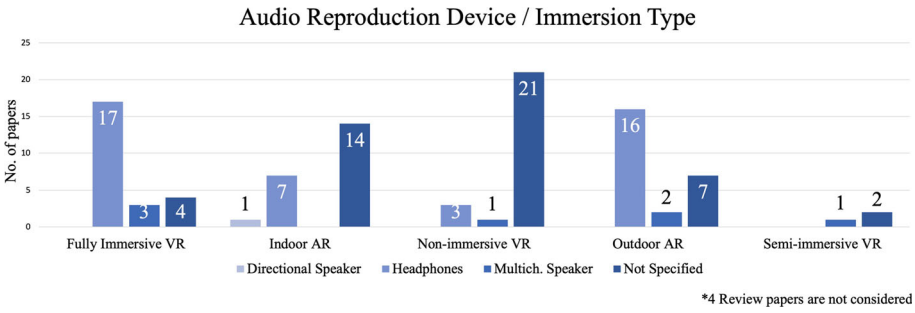


Fig. 5 Count of articles categorized by sound reproduction device and immersion type

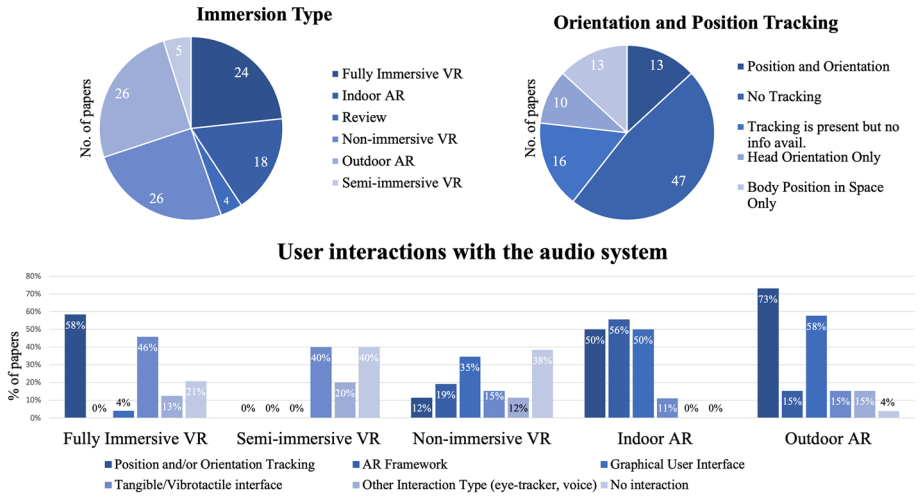


Fig. 6 Distribution of articles by (top left) immersion type, (top right) orientation and position tracking, and (bottom) interaction type

An interesting approach is used in the installation called CAVE [154, 157], a VR theatrical play on nomadic tribes of Northern Europe, whose audio setup includes a combination of custom-built off-ear headphones providing spatial audio and room speakers. The headphones convey sound effects and dialogues, while room speakers and a subwoofer diffuse soundscape and music.

5.2 Interaction with the virtual environment

A correct interaction design and evaluation is fundamental for achieving a high level of coherence and immersion in a system, and in fact, the experience theme resulted from an interaction design activity in 83 out of 92 papers. *Experience* in CAQDAS analysis, along with related concepts such as immersion, support, interaction, and interface, emphasize SIVE as a key area. Fig. 6 classifies the user interactions with the system grouped by tracking method: user body position and head orientation inside the physical or virtual space, object recognition in rooms or sites, physical actions on tangible interfaces, etc.

The most common tracking method (55 papers) selects and personalises audio and video content based on head orientation and position in space, especially in indoor situations. In all outdoor experiences, GPS is used along with compass sensors [55, 56, 79, 108, 109, 122, 135, 136, 138], otherwise tracking is left unspecified [77, 134, 138]. GPS is also used in one interactive artwork described "called Hybrid Gifts" [161] in which a smartphone app permits museum visitors to communicate by geolocated audio messages with peers in order to express and transmit their emotions while looking at paintings in a museum. For a more precise indoor user position estimate, tracking is performed by inertial IMU sensors [46, 121, 123]; otherwise, it is not specified. In VR contexts, user orientation in virtual space is usually tracked using Head-Mounted Displays (HMD) integrating IMU sensors.

Direct interaction with a Graphical User Interface (GUI) on a device screen is widely used (33 papers), especially in AR experiences, often relying on a smartphone as an interaction device, enabling at the same time the use of a wide range of built-in sensors such as IMUs and

cameras as well as internet connectivity. It is important to note that no immersive or semi-immersive experience uses GUI interaction, preferring user movement tracking or tangible controllers.

Computer vision algorithms enabled by AR frameworks (ArKit, ARCore, Vuforia, OpenAL, etc.) are used to recognize objects and elements in space and to interact with an audio virtual environment in 22 works. AR is used to present context- and position-aware touristic information [111, 123, 134, 139, 148, 185] and to navigate in an AR diorama by using a portable device (smartphone or tablet) [78]. Various physical books [75, 88, 96, 97, 97] have been digitally augmented by using devices that recognize book pages and show related content or QR codes [111, 127, 129].

In *War children* [81], stories narrated by WWII eyewitnesses can be watched inside the user's room through a smartphone screen. Huang et al. [168] describe a series of video games for touristic purposes, one of which asks visitors to find a specific image inside a cultural site by using AR to get a game reward.

23 studies use tangible devices, such as console games or VR headset controllers [86, 128, 133, 144, 145, 149, 152, 159, 165, 168, 175, 186], haptic interfaces [158] or custom-built devices. In *ec(h)o* [61], users indicate a particular preference for presented audio content by rotating a wooden cube with coloured faces whose position is detected by a camera. A similar approach is used [169] in a comparison test about movement in a virtual museum, with four different interaction methods on movement in the virtual environment and control condition over an audio guide playback. During the experiment, users can control audio by rotating a playing card (Fig. 7-a), whose movement is recognized by using the Vuforia AR framework. A preference for controlling movements directly was identified.

In an interactive installation, Kenderdine [160] presents a custom-built console to interact with, Kortbek et al. [112] uses a three-step staircase with pressure sensors to detect visitors stepping. Moreover, Geronazzo et al. [153] implemented a custom-built pipe controller (Fig. 7-b) with a button and a 9-axis IMU in order to control movement in an audio virtual environment. On the other hand, Andolina et al. [134] uses a haptic vest with different mounted sensors (Fig. 7-c) and actuators in a navigation task inside a city. The preliminary study compares different interaction/feedback modalities (Visual AR, haptic-audio guidance, and the two combined), showing a preference for the haptic-audio modality. Finally, *The Time Machine* [79, 135, 136] uses vibrotactile feedback together with audio produced by a smartphone for navigating a cultural heritage site. The audio-vibrotactile feedback provides user information about the distance from a specific position of the site.

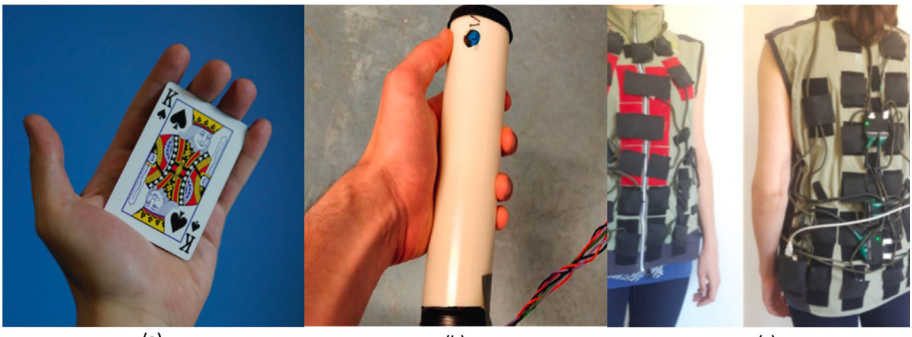


Fig. 7 Examples of tangible devices for interaction: a) AR-controlled playing card. photo ©[169] b) playback control device in [153] c) haptic vest with built-in sensors. photo ©[134]

Other interesting multimodal interactions were described in 8 studies. Two of them use different eye-tracking techniques to detect the user's gaze and present information about the artwork or artistic building the user is looking at in a virtual environment. In particular, Kelling et al. [167] uses eye position estimation based on an IMU sensor of an HMD, Sanchez et al. [130] uses gaze to control the playback of audio elements, such as effect or noises, while reading an children's tales book, and Kwok et al. [113] use eye-tracking glasses. In a fourth study [158], one of the author's interactive installations uses voice and pitch tracking to induce vibration in objects inside a virtual environment based on voice intensity. Voice is also used in an AR experience [185] to receive information based on the narration.

5.2.1 Users with special needs

In reviewing papers, special attention has been paid to SIVE accessibility for cultural heritage experience users with special needs. Unfortunately, available solutions often provide limited accessibility for the unnatural or insufficient interaction methods proposed therein [111]. This is also noticeable in CAQDAS analysis that presents no concepts related to this theme (such as *special needs*, **impairment**, *accessibility*, etc.). Notwithstanding, users with disabilities, especially low vision, can benefit from audio technologies, and some contributions can be found in audio storytelling for cultural heritage as well. In the following paragraphs, we would like to mention five research projects that provide meaningful examples of the inclusive role of audio in such a context. *The Time Machine* [79, 135, 136] is a touristic guide that provides vibrotactile feedback to convey distance and directional information about interesting sites, so as to keep focus on the environment and help navigation of visually impaired people and elderly users. *MuSA* [111] is an application specifically designed for people with low vision, which reads information about art pieces in a museum and then renders it through AR. By receiving an image from a smartphone camera pointing to a specific artwork, the application recognizes various elements and then provides speech information as well as augments the image with colour optimizations and zooms on its particulars. *Greta* [131] is a mobile application that provides audio descriptions of films in cinemas specifically designed for people with low vision. The preliminary evaluation study shows that the use of voiceovers improves enjoyment and immersion but also affects engagement. *Trying to Get Trapped in the Past* [163] is a virtual drama whose narrative relies on spatial audio content, with visually impaired people in the sonic interaction designer's mind. In the recent work by Kelly [119], the author describes the early stage design of a sonification aimed at improving the accessibility of Irish cultural heritage sites.

5.3 Storytelling and personalisation

As underlined by the CAQDAS analysis (concept **Story**, *user* and *experience*), narrative structure (*narrative*, *history*, *landscape*) and the level of sonic interaction between user and audio virtual environment (*sense*, *experience*, *process*) are key aspects to consider when dealing with cultural heritage storytelling in mixed realities. The intersection between story and experience themes highlights that stories should be *immersive*, and the *context* is important for the narration. To enhance the sense of presence in mixed reality, especially in the entanglement dimension, storytelling should be designed to be interactive and non-linear in order to allow active participation of both users and narrations eliciting higher engagement and immersion [189]. Non-linearities in narrative can be achieved by using different techniques

such as inverting the chronological order of events, creating different parallel storylines, and dynamically modifying the narrative.

The majority of studies use non-linear storytelling (67 out of 92 papers) to let users freely explore the augmented physical environment or the virtual space, conversely, 35 papers follow a linear narrative. No specific information about storytelling architecture can be found in the three remaining papers. An overview of storytelling personalisation features is given in Table 4. A particularly relevant work [105] exploits a dual non-linear structure for encouraging a collaborative approach. As soon as they start the visit of the St. Fagan Historical Museum in Wales, users have to choose between two possible partial storylines. Information about the discarded storyline can be obtained only by asking and discussing it with other users. Moreover, artificial intelligence (AI) was employed [159] to adapt storytelling in a VR video game story and, in *Exhibot* [108, 109], to generate storylines about the history of the central square in Heraklion(Greece) based on user position and orientation and third party content services. Surprisingly, the previously cited works were the only contributions using AI techniques, although a large number of works can be found in other interactive and non-linear storytelling contexts. For instance, Riedl et al. [190] present a review of AI techniques in Computer games storytelling, Hernandez et al. [191] discusses the application of AI for eliciting emotions during storytelling, Pisoni et al. [192] introduces AI techniques for accessibility in cultural heritage including interactive storytelling.

Position and head/body orientation are often tracked and used to select a storyline. In this context, content can be pre-determined or personalized in different aspects (point of view in space, content presentation order, dynamically generated elements, and directions) by a user's movement or explicit command. City tour maps are dynamically generated [54, 135, 136] by using the user's navigation path, along with historical and touristic insights about buildings and monuments nearby and at the destination. In *Cave* [157], linear storytelling is spatialised according to the user's specific position inside the room. In the SARIM system [184], sound zones are associated with audio samples coherent with the specific exhibited art piece or historical device.

In terms of audio content, we identify three main high-level categories: (synthetic or real) speech, environmental sounds, and music (Fig. 8). Since voice is the most used material,

Table 4 Content personalisation and personalisation method in non-linear storytelling

What	How
Soundscape	User model detected by user position and orientation in space [121] User position and/or head orientation [56] Interaction with the virtual environment and AI techniques [165]
Virtual Environment and Story	User interaction with tangible interface [86, 158, 160] Interaction with voice [86, 160]
Directions	User position and/or head orientation [79, 128, 132, 135, 136]
Storyline	Interaction with HMD headset controller [108, 109, 116, 128, 159, 164, 185] AI content generation [108, 109] Interaction with a tangible device [61]
Content created by users	User interaction with tangible interface [152] Collaborative experience [161, 171]

especially in tourist or educational experiences, sound effects are used to create soundscapes and personalized content. In Ec(h)o [61], audio soundscape and content are based on user preferences selected by choosing the orientation of a physical cube. In a work by Fu et al. [56], the audio soundscape is dynamically generated based on the user's position. On the other hand, music is used mainly in theatrical or linear storytelling experiences.

5.4 Collaborative experiences

Collaborative and social experiences are important, especially for improving the level of Immersion and Entanglement with both technology and the user's peers: in different cultural heritage subfields, such as music or multimedia production, collaboration elicitation among users is crucial [193]. CAQDAS analysis underlined that *social* is one of the key concepts found in the reviewed works. It is strictly connected with *story* theme (*place*, *history*, *urban*), which emphasizes that when narratives occur in public places, users can benefit from collaboration, and with *experience* that together with the previous theme suggest the possibility of new *digital media* that can include collaborative narratives. In our review, 10 works consider collaborative and social aspects. The majority of them were ad hoc installations in which users can interact in real/virtual shared spaces with tangible interfaces [160], voice intensity and pitch [158], or movement [142]. In [112], social interaction is fostered by a particular physical room setup in which directional speakers mounted on the ceiling deliver personal information about a physical museum without isolating visitors from environmental sounds. Hättich and Schweizer underline the importance of cinemas in fostering sociality [131]. In *This Land AR* [171], three users can interact with virtual musical instruments implemented in mobile devices to create a collaborative audio performance. Moreover, storytelling can be efficiently designed for eliciting discussion during or after the experience as in *Traces* [105] artistic installation/touristic guide, where different complementary linear narratives are used to enhance discussion among different visitors.

Another remarkable point of view [91] states that the use of audio guides or other mediation devices during a cultural heritage visit has the twofold effect of enhancing learning performance and diminishing social interaction. This effect was particularly evident when

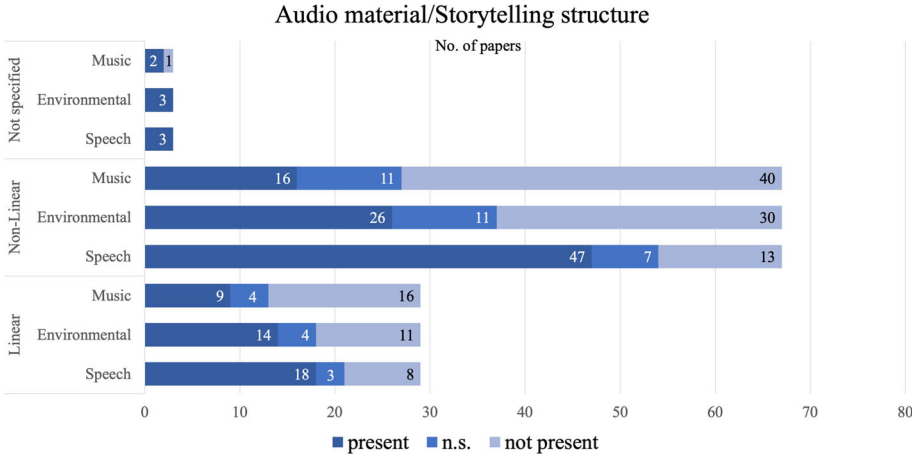


Fig. 8 Distribution of articles by audio material/storytelling structure

wearable devices such as headphones were employed, facilitating the creation of personal virtual environments without any collaboration support.

Previous works suggest that the perception of a virtual space changes with the proposed experience, and can help achieve different specific purposes. A careful design of sonic interactions is a powerful tool for inducing collaboration in a virtual environment and has to be encouraged as a mean of enhancing immersion in shared virtual spaces and providing different communication channels.

The concept of proxemics, defined by psychologists as the set of implicit social rules of interpersonal distance among people, has well-known cognitive effects [194] and can help in describing immersive sonic experiences in shared spaces.

5.5 Quality of the experience and evaluation

Evaluation of different aspects is a fundamental task for creating an effective, immersive, and engaging experience. Fig. 6 shows the immersion type in the considered papers, holding that a wide variety of experimental protocols have been used to evaluate system usability, engagement, immersion, learnability, emotional content, cognitive load, and educational aspects, and most studies rely on questionnaires and semi-structured interviews. From the authors' point of view, the proposed user tests show an overall good acceptance level and engagement, although it is difficult to obtain quantitative information due to differences in the experimental procedure and measured features. CAQDAS analysis does not present any reference to the evaluation of user experience or interaction among the concepts represented, probably because of the heterogeneity of the measurement instruments and traits measured. However, some common traits can be identified in current research.

In order to test User Experience, validated questionnaires are used in many works. The NASA-TLX[11] questionnaire is used [79, 113, 136] to evaluate technology in terms of Mental and Physical Demand Performance, Effort, and Frustration and has been proven valuable to test user's technology acceptance levels. System Usability Scale (SUS, [195]) is a 10-item questionnaire and has been used [54, 111, 113, 148] for testing system usability. User Experience Questionnaire[12](UEQ) is a 26-item questionnaire used [113, 169] to evaluate the attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty of the technology. Both questionnaires have been adopted in order to avoid fatigue in users while obtaining comparable and valid results. Simulation Sickness Questionnaire for Cybersickness (SSQC, [196]) has been used for evaluation in the different study conditions [169]. Narrative Engagement Scale (NES, [197]) is used [131] to evaluate the narrative engagement difference elicited by an audio description of films in people with low vision. In [153], User experience in a spatialized narration task is evaluated in terms of immersion and elicited emotions. Users were asked to navigate in an audio-virtual spatialised soundscape composed of several different virtual rooms while listening to a story by using a tangible controller with different degrees of control over movement, audio spatialisation, and audio playback. In the study, the Immersive Response Questionnaire (IMX[13]) was used to evaluate immersion and the Discrete Emotion Questionnaire (DEQ, [198]) to measure emotional content, along with Heart Rate with a non-intrusive device (a wearable band) showing an increased immersion and emotional content in conditions with spatial audio and controlled playback.

[11] Available at https://humansystems.arc.nasa.gov/groups/tlx/

[12] Available at https://www.ueq-online.org/

[13] Available at https://innermetrix.com/products/assessments/

In addition to Geronazzo et al., other studies use biosignals for measuring emotional parameters, particularly arousal. Jurica et al. [106] compare Electro-Dermal Response (EDR) with a custom questionnaire for measuring arousal in soundscape navigation tasks inside a cultural heritage site, highlighting a better accuracy of the questionnaire for this specific task. On the other hand, Mansilla et al. [170] uses EDR to measure the emotional content of *Virtual Acousmatic Storytelling Environment* (VASE) virtual play, showing a correlation between EDR and emotional content increase. Biosignals are confirmed to be useful for emotion measurement and are suggested by the authors of this paper to be promising for use within tests employing auditory stimuli.

Moreover, mixed reality technologies have been proven to enhance performance in the education domain applied to cultural heritage in various studies [75, 76, 78, 91, 93, 95, 96]. In the majority of these studies, a custom questionnaire designed by the researchers involved in the study has been proposed to participants before and after test sessions to evaluate the differences in terms of mechanical and meaningful information memorization for educational purposes.

6 General discussion

Most of the reviewed studies in the cultural heritage domain are related to tourism and museums. In particular, audio guides are augmented with personalized content, non-linear storytelling, user localisation, and positional tracking to (i) propose an enhanced cultural heritage experience fitting the user preferences, (ii) elicit interest and engagement in users, and (iii) foster cultural heritage dissemination. In some applications, tourist information and street directions are shown, helping tourists orient themselves within a cultural site or city. In others, visitors were immersed in audio AR in which synthetic sonic environments aimed at conveying emotional engagement, re-creating historical events, or presenting position-based narratives.

Notwithstanding the vast amount of literature about narrative and storytelling, to our surprise, we found very little information about narrative content and structure. Practically speaking, the mandatory inclusion of all the identified clusters (*storytelling, immersive media, audio, cultural heritage*) identified the intersection of all these only. Accordingly, this issue suggests the necessity to design new methods for a combined analysis of immersive and non-linear narrative experiences in interdisciplinary contexts. Since none of the reviewed papers explicitly considered a systematic approach to digital narratives, our work would also provide different perspectives for sonic interaction designers, practitioners, and technologists.

In our opinion, the criteria described in Sec. 2.2 are useful in defining a starting point for advanced sonic interactions in immersive storytelling. In the following, we suggest some promising research directions based on the previous guiding principles.

- **Perspective and Point of View.** To provide an immersive experience, most narratives should be designed using an egocentric audio approach [20], in which sonic interactions have to be defined in terms of personalized listener-virtual environment relations. The term egocentric refers to the perceptual reference system for the acquisition of multisensory information in immersive mixed reality environments, as well as the sense of subjectivity and perceptual/cognitive individuality that shape the self, identity, or consciousness of the user. Such a level of personalization should avoid the mediating action of the immersive technology, which might result in a break in presence that can hardly be restored after a pause [199]. These cognitive illusions depend, for example, on the level of hearing training and familiarity with a stimulus/sound environment, and they should

be taken into account from an egocentric perspective to create an immersive, coherent, and entangled experience.

- **Gift of voice.** When speech is employed, the narrative is usually held in the third person for informative purposes or city directions and in the first person for conveying emotional content, such as witness experiences or event reconstructions [77, 81]. Popp [133] suggests that the narrator's role can influence the trustworthiness of the conveyed information of the message. In the analyzed studies, no information can be found about the narrator register. However, we suggest that the narration register should be personalized according to the user's psychological state and cultural background to create a more immersive and user-specific experience.
- **Narrative, Dramatic Question, Emotional Content, and Power of Soundtrack.** From the perspective of the conveyed message, various works [105, 106, 153, 156] are specifically designed for emotion elicitation. Music is mainly used in experience with artistic purposes and linear storytelling. Non-musical elements, such as speech and sound effects, are more common in reviewed works and are sometimes personalized or dynamically generated based on the user's position or interaction. Again, sonic interactions in virtual environments mediated by physical devices or GUIs help modify the users' internal psychological state during personal or collective situations, resulting in entangled experiences. Emphatic technologies have a pivotal role in such modulation, tracing fluid boundaries between humans and technologies [200], and eliciting internal emotional states following the user's expectations and emotions.
- **Economy, Pacing and Interactivity.** In many works, the total time and pacing of the experience depend on the interest and involvement of the user, who seems to prefer a higher degree of control in terms of movement and interaction with the virtual environment and storytelling structure. Storytelling personalisation based on user interaction with the sonic or visual virtual environment, as well as non-linear narrative structure, provide promising evidence of eliciting users' interest in cultural heritage and enhancing dissemination and education [92, 97]. Such a complexity might be managed by AI agents, i.e., non-human entities capable of interacting with ecological behaviors [201], that would be able to predict the user's intentions in an intelligent environment for storytelling. Their ability to monitor listeners' behavioural responses could balance users' expectations and cognitive capabilities to adapt and modulate specific interactions and events [202]. More importantly, AI algorithms have the potential to encourage the exploration of different and meaningful paths within a non-linear narrative tailored to the user's needs.
- **Medium**: In the reviewed works, the immersive medium in use was clearly specified and the experience was usually designed considering its peculiarities. Based on the reality-virtuality continuum proposed by Milgram and Kishino [203], the level of isolation and the combination of real and synthetic elements can differ in audio mixed realities. Fig. 9 illustrates a continuum in the context of audio-specific domain, where different levels of isolation range from a completely real environment (a) to a completely isolated virtual environment. Passing through different degrees of audio augmented environment allows users to experience synthesized audio elements in the real world by using headphones or other hearing devices with a high degree of audio transparency [204]. Skarbez et al. [26] state that it is impossible to completely avoid conflict between sensory information originating in the real environment and sensory information originating in a virtual or augmented environment. Therefore, every VR experience is actually a combination of virtual (e.g., computer-generated stimuli) and real elements (e.g., the feeling of gravity), hence ending up with a mixed reality. In this perspective, the medium to consider in

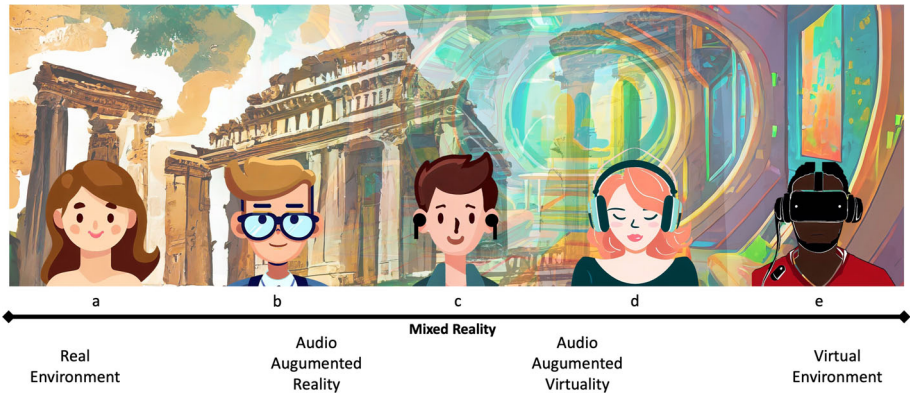


Fig. 9 Milgram and Kishino's Mixed Reality on the Reality-Virtuality Continuum [203] with audio reproduction devices. Made with Microsoft Image Creator and Adobe Firefly

both an AR and a VR experience should be studied as part of a unique mixed reality continuum, thus simplifying the design work and the production of guidelines [205].

6.1 Sonic interactions - notable mentions

In the realm of sonic interaction, tangible and haptic devices are widely employed, especially in non-linear storytelling experiences. Interaction within the virtual environment is used by the system to select a specific storyline and guide the user in the virtual or augmented world. To foster a more natural interaction, several works [112, 134, 153, 160, 169] have adopted custom-built tangible devices that are usually used as the main element during the interactive experience and are usually rated as engaging by users.

Despite the few works reviewed, we would like to mention eye tracking as another promising opportunity to (i) provide information in an ecological manner during interaction with the virtual environment (e.g., [113, 167]), (ii) make the experience accessible in the presence of users with physical impairments([135]), (iii) augment the experience with multisensory elements such as the environmental sound of tactile feedback (e.g., [130]).

Focusing on interactive installations only, we extracted some interesting insights regarding socialization and collaboration aspects. With the help of particular audio setup configurations such as extra-aural headphones [157] or directional speakers [112], or by using standard headphones with custom sensors [142], researchers recreated a social experience taking place in an augmented shared space. From the technological point of view, it was difficult to retrieve specific information about the audio speaker/headphones setup and spatialisation models, especially in studies involving mobile phones and tablets. For the sake of repeatability in research, more technical audio specifications are thus needed to evaluate immersion and to create design guidelines for cultural heritage augmented audio storytelling platforms. This is even more relevant because studies including cognitive evaluations of emotional content, engagement, interest, usability, learnability, acceptability, cognitive load, etc., lead to better results for audio mixed reality compared to static audio guides and onsite information panels [96, 206]. Educational aspects also benefit from augmented books and virtual AR exhibitions in different contexts: children's narratives, travel guides, and interactive installations. In particular, children show interest in this specific technology, hence opening room for future research, even though ethical issues about technology addiction must be considered.

6.2 Research directions

It is evident from our perspective that entanglement in human-computer interaction is highly desirable, as suggested by Frauenberger [28]. Mid-long-term guidelines should integrate technological aspects and content to pave the way for truly immersive storytelling. In particular, our theoretical framework suggests a binding role of the auditory component in its fluidity of perception-action within virtual environments through an ecological perspective [207]. This means that the listener might enact cultural heritage experiences in exploring the virtual environments [208]. Accordingly, we should consider an embodied, environmentally situated perceiver where sensory and motor processes are enabled by technology and inseparable from exploratory action in a narrative space. It is also important to note that in the field of immersive storytelling for cultural heritage, the new methods introduced by AI may open up new possibilities and design in terms of ease of use, involvement, and accessibility. To better understand the possibilities enabled by a proper design of SIVE, some potential research directions are illustrated below.

6.2.1 Emotional museum

In [209], Perry et al. argues that the state of the art of virtual museums doesn't take full advantage of all available possibilities in terms of interaction between participants, and emotional and personal development. An interesting scenario can be the use of audio cultural heritage storytelling to create a virtual museum or augmented cultural heritage site specifically designed to convey emotional content through the exploration of the environment. This kind of experience permits new forms of knowledge dissemination for educational purposes or entertainment. Specific guidelines for cultural heritage audio storytelling can be a useful starting point for the design of more effective storytelling that conveys emotional historical reconstruction in an effective and engaging way.

Challenges: Some historical and cultural experiences can contain a significant emotional impact. In this context, the SIVE framework can embrace the Research through Design (RtD) approach [210], helping in developing methods to identify the emotions to be expressed and to treat with respect and care for the arousal and respect of the users. Research through Design combines design practice and inquiry to better understand complex scenarios. It involves iteratively developing prototypes that allow the capture of emerging patterns. This approach recognizes that design practices are not just meant for implementation or creativity but also valuable methods for conducting research. Such practices could be a valid methodology for exploring interactions and possibilities with non-human agents, e.g., the virtual narrator [211]. The authors recently applied this research perspective in designing an augmented reality audioguide for museums [212].

6.2.2 Immersive analytics

Especially with the new possibilities made available by Artificial Intelligence, immersive media, such as mixed reality and audio mixed reality can be used as a new methodology for research, knowledge analysis and decision-making [213]. *Immersive analytics* consists of using interaction with the immersive media to support analytical reasoning and decision making, providing multimodal interfaces allowing users to immerse themselves in data [214]. A similar approach is already employed in sonification, which consists in transforming data information through sound and auditory features in order to facilitate communication and

interpretation. On the other hand, immersive analytics aim to enhance a bidirectional and entangled interaction between the user and the virtual environment [215].

In the literature, few tools [216] exist for immersive analytics of visual media only. However, cultural heritage research can benefit from the auditory component. Many musical sources are available, especially in fields related to audio and music, such as musicology and musical aesthetics but also by analyzing historical audio sources along with written documents. In this context, audio storytelling can be used twofold. Firstly, it helps researchers orient through the vast amount of audio material available by creating specific story paths, e.g., with particular themes or timelines based on historical facts. Secondly, implementing tools to create cultural heritage audio storytelling and use the design process as a reasoning method for knowledge discovery. Moreover, the use of AI can foster collaboration between the user and the virtual environment in an entangled experience [20].

Challenges: In this type of entangled experience, proper verification and processing of sources are crucial to avoid any misleading storylines that may be generated or introduced by interaction with the virtual environment and AI's hallucinations [217]. Moreover, integrating diverse sources of audio data, as well as designing a convincing interactive experience in mixed reality, can be complex and require robust data management and the integration of multiple 3D user interactions.

6.2.3 Archaeoacoustics, virtual musical instruments, historical voices, and personalities

The analysis of the acoustic of ancient historical places is used by many cultural heritage researchers to better understand different historical aspects of everyday living[14]. In a work by Ciaburro et al. [218], acoustics properties (reverberation time) of an Ancient Roman catacomb have been measured and analyzed in order to better understand the reason for the choice of the specific prayer space by the ancient population. Similarly, Warusfel and Emerit [219] used available historical documents to simulate the acoustics of an ancient Egyptian temple of Dendara dedicated to Hathor, the goddess of music, love, and joy in order to investigate the role of sound in worship ceremonies. Moreover, the reconstruction of virtual ancient or modern musical instruments is widely used to study the main traits of instrument sound throughout history [220]. Again, in this scenario, audio storytelling can help reconstruct historical events and ancient spaces in virtual spaces with natural acoustics and sound for research purposes, education and entertainment, e.g., in museums and cultural heritage sites. Historical talks and documents played or read immersed in a virtual/augmented reconstructed audio environment could help understand the role of space in history and better comprehend the emotions elicited by historical figures.

Challenges: This type of scenario requires advanced simulation and reconstruction techniques that rely on technology that may not always be accessible or sufficiently advanced to represent historical acoustic environments accurately. Especially in VR and AR experiences, this can affect the correct representation and perception of spaces (e.g., positions and distances [9]), limiting or altering the planned experience. Considering the evolution of sound environments and listening habits in historical/architectural reconstructions is essential. For example, in the given location, external noises such as car traffic not present at the reconstructed period, could affect the perception of the experience. Another example is the modification of the tuning of musical instruments, such as that introduced by J.S. Bach in the 17th century [221]. Using the previous configuration, which was conventional for listeners of the time, might sound out of tune to users.

[14] https://pro.europeana.eu/post/discover-the-potential-of-3d-and-heritage-acoustics

6.2.4 Universal fruition, multimodality

Multimodal interaction is widely used in museums and cultural heritage sites for entertainment and education but often has few or no audio elements [222], an important feature in many experiences designed for fruition by users with special needs [223]. It is worthwhile to notice that audio storytelling has been proven to be an important accessibility medium for people with disabilities. Visually impaired people highly benefit from visiting museums and cultural heritage sites for historical and touristic information access and urban navigation [224]. Moreover, storytelling was employed in connection with sound spatialisation and 3D elements in the therapy and rehabilitation of cognitively impaired children for facilitating child-therapist interaction [225]. Audio storytelling can deeply help users with physical disabilities but also users with cognitive disabilities in cultural heritage knowledge fruition, improving attention and engagement [226].
Challenges: The addition of too many multimodal elements could confuse the user with special needs and make it complicated to maintain attentional focus on the topic [43]. Moreover, ensuring the sustainability of multimodal interaction systems involves continuous maintenance, content updates, and technical support can be challenging. This may be problematic when moving from prototyping to implementation and ultimately to maintenance due to limited budget and organisational and political issues.

6.3 Limitation

The review process of our study may be limited due to the intrinsic constraints of our search method, i.e., the specific keyword clusters we selected for analysis. Although we carefully chose these clusters employing a multidisciplinary approach, the proposed selection might have restricted the results. Moreover, there may be some specific terminologies related to the study that we did not consider, which could have affected the retrieved works.

Although we used specific scientific repositories universally considered highly reliable, they may have missed some important studies. On the other hand, it is essential to keep in mind that the field of cultural heritage is strictly related to cultural heritage institutions such as museums and public or private foundations. These institutions may have restrictions on content production and delivery for research purposes. They already have a corpus of content and narratives related to artworks and historical sites, including written guides, audio guides, commentaries, and other multimedia products. However, such sources could not be diffused due to copyright issues or political constraints.

Finally, the creation of interesting and engaging storylines and material for various platforms becomes difficult because of the costs associated with the change of the medium, due to the fast-paced technological advancements that are not financially viable in the long run, and the lack of standard technologies and frameworks.

7 Conclusion

This paper systematically reviewed works on audio mixed reality storytelling for cultural heritage by analyzing recent results, identifying promising trends, discussing peculiar works, and proposing directions for future research. Moreover, a brief overview of platforms and applications specifically designed for augmenting audio in cultural heritage storytelling has been presented and discussed.

Audio mixed reality technologies were mainly used in cultural heritage for tourism and education, especially in mobile solutions. Information about audio setups was often incomplete. Nevertheless, audio mixed reality appeared to be a promising technology for the field, and its implementation in a broader context should be examined more closely by sonic interaction designers, especially concerning its potential to convey immersive experiences to users.

We exploited the scientific context of SIVE in analyzing the reviewed works from a more comprehensive point of view to highlight common aspects and valuable examples for engaging storytelling experiences for cultural heritage in terms of immersion, coherence, and entanglement. Audio storytelling was widely used in cultural heritage and mixed reality with a *de facto* dichotomy between technical discussions and narrative content. In particular, the last aspect seems weakly connected to identifying each sound source and structure and even more to the relationship between storytelling and immersion in terms of experience, enjoyment, and learnability. Nonetheless, personalised and non-linear storytelling, along with sonic interaction, strongly contribute to a complex interrelation towards the creation of a sense of presence in virtual environments. User experience evaluations, mainly based on questionnaires and semi-structured interviews, showed an overall good acceptance level and proved cultural heritage storytelling in immersive environments engaging. Users preferred interactive experiences over controlled ones, obtaining a higher level of overall entanglement and coherence, which can also be fostered by collaborative experiences.

The knowledge acquired with this overview and analysis suggests that a joint effort among sonic interaction designers coming from different backgrounds would be desirable, with the aim of connecting a multiplicity of viewpoints into a shared research agenda.

Funding Open access funding provided by Università degli Studi di Udine within the CRUI-CARE Agreement.

Data Availibility Statement Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Bassano C, Barile S, Piciocchi P, Spohrer JC, Iandolo F, Fisk R (2019) Storytelling about places: Tourism marketing in the digital age. Cities 87:10–20. https://doi.org/10.1016/j.cities.2018.12.025
2. M. Slater, M.V. Sanchez-Vives, Enhancing our lives with immersive virtual reality, Frontiers in Robotics and AI **3** (2016). https://doi.org/10.3389/frobt.2016.00074
3. Cecotti H (2022) Cultural heritage in fully immersive virtual reality. Virtual Worlds 1(1):82–102. https://doi.org/10.3390/virtualworlds1010006

4. Balbi B, Marasco A (2022) Co-designing cultural heritage experiences for all with virtual reality: a scenario-based design approach, Umanistica Digitale No. 11, rights and values in the Digital Age. https://doi.org/10.6092/ISSN.2532-8816/13686
5. Ceuterick M, Ingraham C (2021) Immersive storytelling and affective ethnography in virtual reality. Review of Communication 21(1):9–22. https://doi.org/10.1080/15358593.2021.1881610
6. Lee SJ (2016) A review of audio guides in the era of smart tourism. Information Systems Frontiers 19(4):705. https://doi.org/10.1007/s10796-016-9666-6
7. Bekele MK, Pierdicca R, Frontoni E, Malinverni ES, Gain J (2018) A survey of augmented, virtual, and mixed reality for cultural heritage. Journal on Computing and Cultural Heritage 11(2):1. https://doi.org/10.1145/3145534
8. Regia-Corte T, Marchal M, Cirio G, Lécuyer A (2012) Perceiving affordances in virtual reality: influence of person and environmental properties in perception of standing on virtual grounds. Virtual Reality 17(1):17–28. https://doi.org/10.1007/s10055-012-0216-3
9. Privitera AG, Fontana F, Geronazzo M (2023) In: Proceedings of the 15th Biannual Conference of the Italian SIGCHI Chapter (Association for Computing Machinery, New York, NY, USA), CHItaly '23. https://doi.org/10.1145/3605390.3605422
10. Broderick J, Duggan J, Redfern S (2018) In: 2018 IEEE Games, Entertainment, Media Conference (GEM) (IEEE). https://doi.org/10.1109/gem.2018.8516445
11. Monache SD, Misdariis N, Özcan E (2022) Semantic models of sound-driven design: Designing with listening in mind. Design Studies 83:101134. https://doi.org/10.1016/j.destud.2022.101134
12. Dimoulas CA (2022) Cultural heritage storytelling, engagement and management in the era of big data and the semantic web. Sustainability 14(2):812. https://doi.org/10.3390/su14020812
13. Vert S, Andone D, Ternauciuc A, Mihaescu V, Rotaru O, Mocofan M, Orhei C, Vasiu R (2021) User evaluation of a multi-platform digital storytelling concept for cultural heritage. Mathematics 9(21):2678. https://doi.org/10.3390/math9212678
14. J. Yang, A. Barde, M. Billinghurst, Audio augmented reality: A systematic review of technologies, applications, and future research directions, Journal of the Audio Engineering Society **70**(10), 788 (2022). https://doi.org/10.17743/jaes.2022.0048
15. de Villiers Bosman I, Buruk OO, Jørgensen K, Hamari J (2023) The effect of audio on the experience in virtual reality: a scoping review, Behaviour & Information Technology pp. 1–35. https://doi.org/10.1080/0144929x.2022.2158371
16. Youngblood G (2020) Expanded cinema. Meaning Systems (Fordham University Press)
17. Bederson BB (1995) In: Conference companion on Human factors in computing systems - CHI '95 (ACM Press), CHI '95. https://doi.org/10.1145/223355.223526
18. Marti P, Rizzo A, Petroni L, Tozzi G, Diligenti M (1999) Adapting the museum: a non-intrusive user modeling approach (Springer Vienna), p. 311–313. https://doi.org/10.1007/978-3-7091-2490-1_34
19. Popovici DM, Morvan S, Maisel E, Tisseau J (2003) In: Proceedings. 2003 International Conference on Cyberworlds (IEEE Comput. Soc), CW-03. https://doi.org/10.1109/CYBER.2003.1253489
20. Geronazzo M, Serafin S (2023) Sonic interactions in virtual environments. Human-Computer Interaction Series. https://doi.org/10.1007/978-3-031-04021-4
21. Doornbusch P, Kenderdine S (2004) Presence and sound; identifying sonic means to "be there"., Consciousness Reframed 2004
22. Kim G, Biocca F (2018) Immersion in Virtual Reality Can Increase Exercise Motivation and Physical Performance (Springer International Publishing), p. 94–102. https://doi.org/10.1007/978-3-319-91584-5_8
23. Kern AC, Ellermeier W (2020) Audio in vr: Effects of a soundscape and movement-triggered step sounds on presence. Frontiers in Robotics and AI 7. https://doi.org/10.3389/frobt.2020.00020
24. Varela FJ, Thompson ET, Rosch E (1992) The embodied mind. The MIT Press (MIT Press, London, England)
25. Atherton J, Wang G (2020) Doing vs. being: A philosophy of design for artful vr. Journal of New Music Research **49**(1), 35–59. https://doi.org/10.1080/09298215.2019.1705862
26. Skarbez R, Smith M, Whitton MC (2021) Revisiting milgram and kishino's reality-virtuality continuum. Frontiers in Virtual Reality 2. https://doi.org/10.3389/frvir.2021.647997
27. Hall ET, Birdwhistell RL, Bock B, BohannanP, Diebold AR, Durbin M, Edmonson MS, Fischer JL, Hymes D, Kimball ST, Barre WL, Frank Lynch SJ, McClellan JE, Marshall DS, Milner GB, Sarles, HB, Trager GL, Vayda AP (1968) Proxemics [and comments and replies]. Curr Anthropol 9(2/3):83. http://www.jstor.org/stable/2740724
28. Frauenberger C (2019) Entanglement HCI the next wave? ACM Transactions on Computer-Human Interaction 27(1):1. https://doi.org/10.1145/3364998

29. Geronazzo M (2022) Egocentric audio in the digital twin of virtual environments. In: 2022 IEEE 2nd International Conference on Intelligent Reality (ICIR) (IEEE). https://doi.org/10.1109/icir55739.2022.00017
30. Frauenberger C (2019) Entanglement HCI the next wave? ACM Transactions on Computer-Human Interaction 27(1):1. https://doi.org/10.1145/3364998
31. Gibson JJ (2014). The Ecological Approach to Visual Perception: Classic Edition (Psychology Press). https://doi.org/10.4324/9781315740218
32. Poletti A (2011) Coaxing an intimate public: Life narrative in digital storytelling. Continuum 25(1):73. https://doi.org/10.1080/10304312.2010.506672
33. Meadows M (2002) Pause & Effect: The Art of Interactive Narrative (Pearson Education)
34. McLuhan M (1964) Understanding media: the extensions of man. The MIT Press (MIT Press, London, England)
35. Sitter KC, Beausoleil N, McGowan E (2020) Digital storytelling and validity criteria. International Journal of Qualitative Methods 19:160940692091065. https://doi.org/10.1177/1609406920910656
36. Grove N (2012) (ed.), Using Storytelling to Support Children and Adults with Special Needs (Routledge, London, England)
37. Rocchesso D, Serafin S, Behrendt F, Bernardini N, Bresin R, Eckel G, Franinovic K, Hermann T, Pauletto S, Susini P, Visell Y (2008) In: CHI '08 Extended Abstracts on Human Factors in Computing Systems (ACM), CHI '08. https://doi.org/10.1145/1358628.1358969
38. Pierce JR (1999) In: The Psychology of Music (Second Edition), ed. by D. Deutsch, second edition edn., Cognition and Perception (Academic Press, San Diego), pp. 1–23. https://doi.org/10.1016/B978-012213564-4/50002-0. https://www.sciencedirect.com/science/article/pii/B9780122135644500020
39. Lugmayr A, Sutinen E, Suhonen J, Sedano CI, Hlavacs H, Montero CS (2016) Serious storytelling - a first definition and review. Multimedia Tools and Applications 76(14):15707. https://doi.org/10.1007/s11042-016-3865-5
40. Ojala V (2022) Chernobyl dreams: investigating visitors' storytelling in the chernobyl exclusion zone. International Journal of Tourism Cities. https://doi.org/10.1108/ijtc-04-2022-0094
41. Vlachos E, Holck JP, Jensen MK (2022) Narrating the Story of a Digitized Old Historical Map (Springer International Publishing), p. 296–303. https://doi.org/10.1007/978-3-031-06391-6_39
42. Podara A, Giomelakis D, Nicolaou C, Matsiola M, Kotsakis R (2021) Digital storytelling in cultural heritage: Audience engagement in the interactive documentary new life. Sustainability 13(3):1193. https://doi.org/10.3390/su13031193
43. Forceville C (2017) Interactive documentary and its limited opportunities to persuade, Discourse. Context & Media 20:218. https://doi.org/10.1016/j.dcm.2017.06.004
44. Rodrigues JMF, Ramos CMQ, Pereira JAR, Sardo JDP, Cardoso PJS (2019) Mobile five senses augmented reality system: Technology acceptance study. IEEE Access 7:163022. https://doi.org/10.1109/access.2019.2953003
45. ComunitáM Gerino A, Lim V, Picinali L (2021) Design and evaluation of a web- and mobile-based binaural audio platform for cultural heritage. Appl Sci 11(4):1540. https://doi.org/10.3390/app11041540
46. Kaghat FZ, Cubaud P (2010) Fluid interaction in audio-guided museum visit: Authoring tool and visitor device. International Symposium on Virtual Reality p. Archaeology and Intelligent Cultural Heritage, VAST. https://doi.org/10.2312/VAST/VAST10/163-170
47. D'Auria D, Di Mauro D, Calandra DM, Cutugno F (2014) Interactive headphones for a cloud 3D audio application. In: 2014 Ninth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, pp. 437–442. https://doi.org/10.1109/3PGCIC.2014.145
48. Kabisch E (2007) A periscope for mobile discovery and narrative, CC2007 -Seeding Creativity: Tools, Media, and Environments pp. 259–260
49. Salo K, Giova D, Mikkonen T (2016) Backend infrastructure supporting audio augmented reality and storytelling. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Notes in Bioinformatics 9735:325
50. McGinity M, Shaw J, Kuchelmeister V, Hardjono A, Favero DD (2007) Avie: A versatile multi-user stereo 360° interactive vr theatre
51. Beck S (2009) The immersive computer-controlled audio sound theater: history and current trends in multi-modal sound diffusion, SIGGRAPH 2009: Talks p. 1
52. Summers JE (2008) Auralization: Fundamentals of acoustics, modelling, simulation, algorithms, and acoustic virtual reality. The Journal of the Acoustical Society of America 123(6):4028. https://doi.org/10.1121/1.2908264
53. Geronazzo M, Tissieres JY, Serafin S (2020) In: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (IEEE). https://doi.org/10.1109/ICASSP40776.2020.9053873

54. Boletsis C, Chasanidou D (2018) In: Proceedings of the 10th Nordic Conference on Human-Computer Interaction (ACM), NordiCHI'18. https://doi.org/10.1145/3240167.3240243
55. Kratky A (2019), Walking in the head: Methods of sonic augmented reality navigation, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics
56. Fu Y, Byrne D, Shea L (2021) Evoking the post-industrial landscape memories through spectrality and mixed reality soundscapes
57. Sardo J, Pereira J, Veiga R, Semião J, Cardoso P, Rodrigues J (2018) A portable device for five sense augmented reality experiences in museums. WSEAS Transactions on Environment and Development 14:347
58. Santiago A, Sampaio P, Fernandes L (2014) Mogre-storytelling: Interactive creation of 3d stories, Proceedings -2014 16th Symposium on Virtual and Augmented Reality. SVR 2014:190–199
59. Santiago A, Sampaio P, Fernandes L, Martins V (2014) A digital approach to storytelling with mogre, 15th International Conference on Intelligent Games and Simulation pp. 104–111
60. Hatala M, Wakkary R, Kalantari L (2005) Rules and ontologies in support of real-time ubiquitous application. Journal of Web Semantics 3(1):5. https://doi.org/10.1016/j.websem.2005.05.004
61. Hatala M, Wakkary R (2005) Ontology-based user modeling in an augmented audio reality system for museums. User Modeling and User-Adapted Interaction 15(3–4):339–380. https://doi.org/10.1007/s11257-005-2304-5
62. Damala A, Stojanovic N (2012) Tailoring the adaptive augmented reality (a2r) museum visit: Identifying cultural heritage professionals' motivations and needs, 11th IEEE International Symposium on Mixed and Augmented Reality 2012-Arts, Media, and Humanities Papers pp. 71–80
63. Katz BFG, Murphy D, Farina A (2020) The Past Has Ears (PHE): XR Explorations of Acoustic Spaces as Cultural Heritage (Springer International Publishing), p. 91–98. https://doi.org/10.1007/978-3-030-58468-9_7
64. Bilbow S (2022) In: NIME 2022 (PubPub). https://doi.org/10.21428/92fbeb44.8abb9ce6
65. Meyer S (2016) Right, left, high, low narrative strategies for non-linear storytelling, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics
66. Fujinawa E, Sakurai S, Izumi M, Narumi T, Houshuyama O, Tanikawa T, Hirose M (2015) Induction of human behavior by presentation of environmental acoustics. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics 9172:582
67. Kenderdine S, Mason I, Hibberd L (2021) Computational Archives for Experimental Museology (Springer International Publishing), p. 3–18. https://doi.org/10.1007/978-3-030-83647-4_1
68. Rhodes C (2022) In: NIME 2022 (PubPub, 2022), NIME. https://doi.org/10.21428/92fbeb44.6e17eaf5
69. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JPA, Clarke M, Devereaux PJ, Kleijnen J, Moher D (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 339(jul21 1), b2700. https://doi.org/10.1136/bmj.b2700
70. Vaquerizo-Serrano J, Pablo GSD, Singh J, Santosh P (2021) Catatonia in autism spectrum disorders: A systematic review and meta-analysis, European Psychiatry **65**(1). https://doi.org/10.1192/j.eurpsy.2021.2259
71. den Hengst F, Grua EM, el Hassouni A, Hoogendoorn M (2020) Reinforcement learning for personalization: A systematic literature review. Data Science 3(2):107. https://doi.org/10.3233/ds-200028
72. Kusunoki F, Sugimoto M, Hashizume H (2002) Toward an interactive museum guide system with sensing and wireless network technologies. In: Proceedings. IEEE International Workshop on Wireless and Mobile Technologies in Education (IEEE Comput. Soc), WMTE-02. https://doi.org/10.1109/WMTE.2002.1039228
73. Bobick AF, Intille SS, Davis JW, Baird F, Pinhanez CS, Campbell LW, Ivanov YA, Schütte A, Wilson A (1999) The kidsroom: A perceptually-based interactive and immersive story environment. Presence: Teleoperators and Virtual Environments f8(4):369–393. https://doi.org/10.1162/105474699566297
74. Marto A, Gonçalves A, Melo M, Bessa M (2022) A survey of multisensory VR and AR applications for cultural heritage. Computers & Graphics 102:426. https://doi.org/10.1016/j.cag.2021.10.001
75. Tomi A, Rambli D (2013) An interactive mobile augmented reality magical playbook: Learning number with the thirsty crow, Procedia Computer Science pp. 123–130
76. Domínguez P (2016) Audiovisual didactic storytelling through virtual reconstruction: The iberian temple of la alcudia in elche. 2016 International Symposium on Computers in Education, SIIE 2016: Learning Analytics Technologies

77. Dow S, Lee J, Oezbek C, Maclntyre B, Bolter J, Gandy M (2005) Exploring spatial narratives and mixed reality experiences in oakland cemetery, Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology pp. 51–60
78. Harrington M (2020) Connecting user experience to learning in an evaluation of an immersive, interactive, multimodal augmented reality virtual diorama in a natural history museum amp; the importance of story, 2020 6th International Conference of the Immersive Learning Research Network (iLRN) pp. 70–78
79. Rassmus-Gröhn K, Szymczak D, Magnusson C (2013) The time machine -an inclusive tourist guide application supporting exploration, Lecture Notes in Computer Science (including subseries Lecture Notes in ArtiRicial Intelligence and Lecture Notes in Bioinformatics 7989:127
80. Fischnaller F (2018) Virtual journey through the history of fort saint jean, marseille (vj-fsj project) : Case study: New media exhibit at the musée des civilisations de l'europe et de la méditerranée (mucem), 3rd Digital Heritage International Congress (DigitalHERITAGE) held jointly with 2018 24th International Conference on Virtual Systems Multimedia pp. 1–8
81. Zimmer C, Ratz N, Bertram M, Geiger C (2018) War children: Using ar in a documentary context, 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct) pp. 390–394
82. Smith AE, Humphreys MS (2006) Evaluation of unsupervised semantic mapping of natural language with leximancer concept mapping. Behavior Research Methods 38(2):262. https://doi.org/10.3758/bf03192778
83. Chalmers M, Chitson P (1992) In: Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '92 (ACM Press). https://doi.org/10.1145/133160.133215
84. Björken-Nyberg C (2020) Hearing, seeing, experiencing: Perspective taking and emotional engagement through the vocalisation of jane eyre, heart of darkness and things fall apart. International Journal of Language Studies 14:63
85. Greer A (2017) Murder, she spoke: the female voice's ethics of evocation and spatialisation in the true crime podcast. Sound Studies 3(2):152. https://doi.org/10.1080/20551940.2018.1456891
86. Bertacchini F, Bilotta E, Carini M, Gabriele L, Pantano P, Tavernise A (2014) Learning in the smart city: A virtual and augmented museum devoted to chaos theory. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics 7697:261
87. Boylorn RM (2021) Visual voices and aural (auto)ethnographies: the personal, political, and polysemic value of storytelling and/in communication. Rev Commun 21(1):1. https://doi.org/10.1080/15358593.2021.1905870
88. Dünser A, Hornecker E (2007) An observational study of children interacting with an augmented story book, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics 4469:305
89. Gardoni F, Mojetta F, Sorrentino C, Etzi R, Gallace A, Bordegoni M, Carulli M (2020) Raising awareness about the consequences of human activities on natural environments through multisensory augmented reality: Amazon rainforest and coral reef interactive experiences. Computer-Aided Design and Applications 18(4):815. https://doi.org/10.14733/cadaps.2021.815-830
90. Hug D (2010) Investigating narrative and performative sound design strategies for interactive commodities, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics
91. Chang KE, Chang CT, Hou HT, Sung YT, Chao HL, Lee CM (2014) Development and behavioral pattern analysis of a mobile guide system with augmented reality for painting appreciation instruction in an art museum. Computers & Education 71:185. https://doi.org/10.1016/j.compedu.2013.09.022
92. Kitajima M, Shimizu S, Nakahira K (2017) Creating memorable experiences in virtual reality: Theory of its processes and preliminary eye-tracking study using omnidirectional movies with audio-guide, 3rd IEEE International Conference on Cybernetics pp. 1–8
93. Yu SJ, Sun JCY, Chen OTC (2017) Effect of AR-based online wearable guides on university students' situational interest and learning performance. Universal Access in the Information Society 18(2):287. https://doi.org/10.1007/s10209-017-0591-3
94. Tzi-Dong Ng J, Hu X, Que Y (2022) Towards multi-modal evaluation of eye-tracked virtual heritage environment. In: LAK22: 12th International Learning Analytics and Knowledge Conference (Association for Computing Machinery, New York, NY, USA), LAK22, p. 451–457. https://doi.org/10.1145/3506860.3506881
95. Al-Imamy SY (2019) Blending printed texts with digital resources through augmented reality interaction. Education and Information Technologies 25(4):2561. https://doi.org/10.1007/s10639-019-10070-w
96. Corrigan-Kavanagh E, Scarles C, Revill G (2019) Augmenting travel guides for enriching travel experiences. e-Review of Tourism Research 17:334

97. He X, Hong Y (2020) The effect of augmented reality on the memorization in history and humanities education. Advances in Intelligent Systems and Computing 1217 AISC pp. 769–776
98. Laskari I (2019) Creating algorithmic audio-visual narratives through the use of augmented reality prints. Technoetic Arts 17(1):25–31. https://doi.org/10.1386/tear_00003_1
99. Salerno I (2014) Sharing memories and "telling" heritage through audio-visual devices. participatory ethnography and new patterns for cultural heritage interpretation and valorization, Visual Ethnography . https://doi.org/10.12835/ve2014.2-0035
100. Salihbegovic F (2020) The encounter with the real: What can complicite's theatre performance the encounter teach us about the future of vr narratives?, Body, Space & Technology **19**(1):125. https://doi.org/10.16995/bst.336
101. Terracciano A (2018) Zelige door on golborne road: Exploring the design of a multisensory interface for arts, migration and critical studies, CEUR Workshop Proceedings pp. 152–161
102. Kritikos Y, Mania K (2022) Interactive historical documentary in virtual reality, 2022 International Conference on Interactive Media. Smart Systems and Emerging Technologies (IMET) pp. 1–8
103. Sheremetieva A, Romanv I, Frish S, Maksymenko M, Georgiou O (2022) Touch the story: An immersive mid-air haptic experience, 2022 International Conference on Interactive Media, Smart Systems and Emerging Technologies (IMET) pp. 1–3
104. Fakhour M, Azough A, Kaghat FZ, Meknassi M (2020) A cultural scavenger hunt serious game based on audio augmented reality, Advances in Intelligent Systems and Computing 1102 AISC:1–8
105. Huws S, John A, Kidd J (2018) Evaluating the affective dimensions of traces-olion: a subtle mob at st fagans national museum of history, wales, 3rd Digital Heritage International Congress (DigitalHERITAGE) held jointly with 2018 24th International Conference on Virtual Systems Multimedia (VSMM 2018) pp. 1–8
106. Jurica D, Matija P, Mladen R, Marjan S (2019) Comparison of two methods of soundscape evaluation, 2nd International Colloquium on Smart Grid Metrology (SMAGRIMET) pp. 1–5
107. Chaurasia H, x Majhi M (2022) Sound design for cinematic virtual reality: A state-of-the-art review. Lecture Notes in Networks and Systems 391:357
108. Tsepapadakis M, Gavalas D, Koutsabasis P (2022) 3D Audio + Augmented Reality + AI Chatbots + IoT: An Immersive Conversational Cultural Guide. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 13445 LNCS:249
109. Tsepapadakis M, Gavalas D (2023) Are you talking to me? An Audio Augmented Reality conversational guide for cultural heritage. Pervasive and Mobile Computing 92:101797. https://doi.org/10.1016/j.pmcj.2023.101797
110. Paterson J, Kadel O (2023) Audio for extended realities: A case study informed exposition, Convergence p. 13548565231169723. https://doi.org/10.1177/13548565231169723. Publisher: SAGE Publications Ltd
111. Ahmetovic D, Bernareggi C, Keller K, Mascetti S (2021) Musa: artwork accessibility through augmented reality for people with low vision, Proceedings of the 18th International Web for All Conference pp. W4A–2021
112. Kortbek K, Grønbaek K (2008) Interactive spatial multimedia for communication of art in the physical museum space, Proceedings of the 16th ACM international conference on Multimedia pp. 609–618
113. Kwok T, Kiefer P, Schinazi V, Adams B, Raubal M (2019) Gaze-guided narratives: Adapting audio guide content to gaze in virtual and real environments, Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems pp. 1–12
114. Dima M, Maples H (2021) Affectual dramaturgy for augmented reality immersive heritage performance, Body, Space & Technology **20**(1):25. https://doi.org/10.16995/bst.368
115. Giariskanis F, Kritikos Y, Protopapadaki E, Papanastasiou A, Papadopoulou E, Mania K (2022) The augmented museum: A multimodal, game-based, augmented reality narrative for cultural heritage, IMX 2022 -Proceedings of the 2022 ACM International Conference on Interactive Media Experiences pp. 281–285
116. Zaal T, Akdag Salah AA, Hürst W (2022) Toward inclusivity: virtual reality museums for the visually impaired. In: 2022 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR) , pp. 225–233. https://doi.org/10.1109/AIVR56993.2022.00047. ISSN: 2771-7453
117. Sylaiou S, Gkagka E, Fidas C, Vlachou E, Lampropoulos G, Plytas A, Nomikou V (2023) In: Proceedings of the 2nd International Conference of the ACM Greek SIGCHI Chapter (ACM, Athens Greece), pp. 1–5. https://doi.org/10.1145/3609987.3610008
118. Eames A (2019) In: Proceedings of the 17th International Conference on Virtual-Reality Continuum and its Applications in Industry (ACM, Brisbane QLD Australia), pp. 1–2. https://doi.org/10.1145/3359997.3365736

119. Kelly C (2023) In: 2023 Immersive and 3D Audio: from Architecture to Automotive (I3DA) (IEEE, Bologna, Italy), pp. 1–9. https://doi.org/10.1109/I3DA57090.2023.10289582
120. Sun L (2023) In: 2023 Immersive and 3D Audio: from Architecture to Automotive (I3DA) (IEEE, Bologna, Italy), pp. 1–8. https://doi.org/10.1109/I3DA57090.2023.10289367
121. Kaghat FZ, Azough A, Fakhour M, Meknassi M (2020) A new audio augmented reality interaction and adaptation model for museum visits. Computers & Electrical Engineering 84:106606. https://doi.org/10.1016/j.compeleceng.2020.106606
122. Scott L (2017) Creating opera for mobile media: Artistic opportunities and technical limitations, Proceedings - 14th International Symposium on Pervasive Systems, Algorithms and Networks, I-SPAN 2017, 11th International Conference on Frontier of Computer Science and Technology, FCST 2017 and 3rd International Symposium of Creative Computing, ISCC 2017 2017-November. https://doi.org/10.1109/ISPAN-FCST-ISCC.2017.86
123. Gimeno J, Portalés C, Coma I, Fernández M, Martínez B (2017) Combining traditional and indirect augmented reality for indoor crowded environments. a case study on the casa batlló museum. Computers & Graphics 69:92. https://doi.org/10.1016/j.cag.2017.09.001
124. Pietroni E (2012) An augmented experiences in cultural heritage through mobile devices, 18th International Conference on Virtual Systems and Multimedia pp. 117–124
125. Rizvic S, Sadzak A, Hulusic V, Karahasanovic A (2012) In: Proceedings of the 28th Spring Conference on Computer Graphics (Association for Computing Machinery, New York, NY, USA), SCCG '12, p. 109–116. https://doi.org/10.1145/2448531.2448545
126. Russo A, Cosentino R, De Lucia MA, Guidazzoli A, Cohen GB, Liguori MC (2013) Fishnaller F (2013) In. Digital Heritage International Congress (DigitalHeritage) 1:441–441. https://doi.org/10.1109/DigitalHeritage.2013.6743776
127. Sorrentino F, Spano L, Scateni R (2015) Superavatar: Children and mobile tourist guides become friends using superpowered avatars, Proceedings of 2015 International Conference on Interactive Mobile Communication Technologies and Learning pp. 222–226
128. Kritikos Y, Giariskanis F, Protopapadaki E, Papanastasiou A (2023) E. Papadopoulou, K. Mania, in Proceedings of the 2023 ACM International Conference on Interactive Media Experiences (ACM, Nantes France), pp. 199–204. https://doi.org/10.1145/3573381.3597028
129. Krupa F (2023) In: Proceedings of the 2023 ACM International Conference on Interactive Media Experiences (ACM, Nantes France), pp. 376–378. https://doi.org/10.1145/3573381.3597223
130. Sanchez S, Dingler T, Gu H, Kunze K (2016) In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (ACM, San Jose California USA), pp. 1459–1466. https://doi.org/10.1145/2851581.2892353
131. Hättich A, Schweizer M (2020) I hear what you see: Effects of audio description used in a cinema on immersion and enjoyment in blind and visually impaired people. British Journal of Visual Impairment 38(3):284. https://doi.org/10.1177/0264619620911429
132. Boletsis C, Chasanidou D (2018) Smart tourism in cities: Exploring urban destinations with audio augmented reality, ACM International Conference Proceeding Series pp. 515–521
133. Popp C, Murphy D (2022) Establishment and implementation of guidelines for narrative audio-based room-scale virtual reality using practice-based methods, Audio Engineering Society Conference: AES 2022 International Audio for Virtual and Augmented Reality Conference
134. Andolina S, Hsieh YT, Kalkofen D, Nurminen A, Cabral D, Spagnolli A, Gamberini L, Morrison A, Schmalstieg D, Jacucci G (2021) Designing for mixed reality urban exploration. Interaction Design and Architecture(s) 48:33
135. Magnusson C, Rassmus-Gröhn K, Szymczak D (2014) Exploring history: A mobile inclusive virtual tourist guide, Proceedings of the NordiCHI 2014: The 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational pp. 69–78
136. Szymczak D, Rassmus-Gröhn K, Magnusson C, Hedvall PO (2012) A realworld study of an audio-tactile tourist guide, Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services pp. 335–344
137. Raheb KE, Kougioumtzian L, Stergiou M, Petousi D, Katifori A, Servi K, Kriezi V, Vraka V, Merakos S, Charkiolakis A, Venieri F, Boile M, Ioannidis Y (2022) Designing an Augmented Experience for a Music Archive: What does the Audience Need Beyond the Sense of Hearing? Journal on Computing and Cultural Heritage 15(4):1. https://doi.org/10.1145/3528366
138. Guntarik O, Davies H, Innocent T (2023) Indigenous Cartographies: Pervasive Games and Place-Based Storytelling. Space and Culture p. 12063312231155348. https://doi.org/10.1177/12063312231155348. Publisher: SAGE Publications Inc

139. Javornik A, Kostopoulou E, Rogers Y, gen Schieck AF, Koutsolampros P, Moutinho AM, Julier S (2018) An experimental study on the role of augmented reality content type in an outdoor site exploration. Behaviour & Information Technology 38(1):9. https://doi.org/10.1080/0144929x.2018.1505950
140. Melki H (2021) Stage-directing the virtual reality experience: developing a theoretical framework for immersione literacy. International Journal of Film and Media Arts 6(2). https://doi.org/10.24140/IJFMA.V6.N2.08
141. Bresler Z (2023) Pop Music Diegesis and the 360° Video, Popular Music and Society **46**(5). https://doi.org/10.1080/03007766.2023.2272680
142. Heller F, Knott T, Weiss M, Borchers J (2009) Multi-user interaction in virtual audio spaces, Conference on Human Factors in Computing Systems -Proceedings pp. 4489–4494
143. Butterworth A (2022) Beyond sonic realism: a cinematic sound approach in documentary 360° film. Studies in Documentary Film 16(2):156. https://doi.org/10.1080/17503280.2022.2048234. Type: Article
144. Popp C, Murphy DT (2022) Establishment and implementation of guidelines for narrative audio-based room-scale virtual reality using practice-based methods. In: Proceedings of the AES International Conference, vol. 2022-August, vol. 2022-August, pp. 95 – 104. Type: Conference paper
145. Popp C, Murphy DT (2022) Creating Audio Object-Focused Acoustic Environments for Room-Scale Virtual Reality. Applied Sciences (Switzerland) 12(14). https://doi.org/10.3390/app12147306. Type: Article
146. Warp R, Zhu M, Kiprijanovska I, Wiesler J, Stafford S, Mavridou I (2022) In: 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct) (IEEE, Singapore, Singapore), pp. 262–265. https://doi.org/10.1109/ISMAR-Adjunct57072.2022.00058
147. Briggs NC, Buckley J, Chesworth D, Coyne T, Farr A, Harper L, Ho X, Heyns AL, Leber S, Melo Zurita MdL, Raby O (2023) Listen - Look up! Listen - Look down! Experiencing the counter-city through a sonic and augmented reality experience of urban undergrounds in southeast Melbourne. Cities 142:104513. https://doi.org/10.1016/j.cities.2023.104513
148. Marto A, Melo M, Goncalves A, Bessa M (2021) Development and evaluation of an outdoor multisensory AR system for cultural heritage. IEEE Access 9:16419. https://doi.org/10.1109/access.2021.3050974
149. Ogle J, Duer Z, Hicks D, Fralin S (2020) First world war tunnel warfare through haptic vr, ACM SIGGRAPH 2020 Immersive Pavilion pp. 1–2
150. Mulvany GT (2022) In: CHI PLAY 2022 - Extended Abstracts of the 2022 Annual Symposium on Computer-Human Interaction in Play, pp. 291 – 296. https://doi.org/10.1145/3505270.3558374. Type: Conference paper
151. Marques B, McIntosh J, Carson H (2019) Whispering tales: using augmented reality to enhance cultural landscapes and indigenous values. AlterNative: An International Journal of Indigenous Peoples 15(3):193. https://doi.org/10.1177/1177180119860266
152. Bargsten J (2020) Narrative and spatial design through immersive music and audio, Proceedings -2020 IEEE Conference on Virtual Reality and 3D User Interfaces 2020:396
153. Geronazzo M, Rosenkvist A, Eriksen DS, Markmann-Hansen CK, Køhlert J, Valimaa M, Vittrup MB, Serafin S (2019) Creating an audio story with interactive binaural rendering in virtual reality. Wireless Communications and Mobile Computing 2019:1. https://doi.org/10.1155/2019/1463204
154. Gospodarek M, Genovese A, Dembeck D, Corinne, Roginska A, Perlin K (2019) Sound design and reproduction techniques for co-located narrative vr experiences. Audio Engineering Society Convention 147
155. Steynberg A (2020) Using reverse interactive audio systems (rias) to direct attention in virtual reality narrative practices: A case study, International Conference on Interactive Digital Storytelling pp. 353–356
156. Williams D, Daly I (2021) Neuro-curation: A case study on the use of sonic enhancement of virtual museum exhibits. Audio Mostly 2021:121–125
157. Layng K, Perlin K, Herscher S, Brenner C, Meduri T (2019) Cave: making collective virtual narrative. ACM SIGGRAPH 2019 Art Gallery pp. 1–8
158. Barrass S (2008) Creative practice-based research in interaction design. Computers in Entertainment 6(3):1. https://doi.org/10.1145/1394021.1394026
159. Cutler L, Darnell E, Dirksen N, Tucker A, Stafford S, Oh E, Nagpal A, Lee E, Ladd N (2021) From quest to quill: pushing the boundaries of VR storytelling in baobab's Baba Yaga and Namoo. In: ACM SIGGRAPH 2021 Production Sessions (ACM), SIGGRAPH '21. https://doi.org/10.1145/3446368.3452126
160. Kenderdine S (2010) Immersive visualization architectures and situated embodiments of culture and heritage, Proceedings of the International Conference on Information Visualisation pp. 408–414
161. Koleva B, Spence J, Benford S, Kwon H, Schnädelbach H, Thorn E, Preston W, Hazzard A, Greenhalgh C, Adams M, Farr JR, Tandavanitj N, Angus A, Lane G (2020) Designing hybrid gifts. ACM Transactions on Computer-Human Interaction 27(5):1. https://doi.org/10.1145/3398193

162. Lopez M, Pauletto S (2010) The sound machine: A study in storytelling through sound design, Proceedings of the 5th Audio Mostly -A Conference on Interaction With Sound
163. Struck G, Böse R, Spierling U (2008) Trying to get trapped in the past -exploring the illusion of presence in virtual drama. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics 5334:114
164. Vasquez Gomez JC (2022) In: Extended Abstracts of the 2022 Annual Symposium on Computer-Human Interaction in Play (Association for Computing Machinery, New York, NY, USA), CHI PLAY '22, pp. 103–106. https://doi.org/10.1145/3505270.3558341. Event-place: Bremen, Germany
165. Erol Z, Zhang Z, Özgünay E, Ray L (2022) Sound of(f): Contextual storytelling using machine learning representations of sound and music. Social-Informatics and Telecommunications Engineering 422:332
166. Gmeiner TM, Murakami E (2023) In: 29th ACM Symposium on Virtual Reality Software and Technology (ACM, Christchurch New Zealand). https://doi.org/10.1145/3611659.3617201
167. Kelling C, Karhu J, Kauhanen O, Turunen M, Väätäjä H, Lindqvist V (2018) Implications of audio and narration in the user experience design of virtual reality, ACM International Conference Proceeding Series pp. 258–261
168. W.H. Huang, H.M. Chiao, W.H. Huang (2018) Innovative research on the development of game-based tourism information services using component-based software engineering, Advances in Science. Technology and Engineering Systems Journal 3(1):451. https://doi.org/10.25046/aj030155
169. Li Y, Tennent P, Cobb S (2019) Appropriate control methods for mobile virtual exhibitions, Duguleană M, Carrozzino M, Gams M, Tanea I (eds) VR Technologies in Cultural Heritage pp. 165–183
170. Mansilla W (2006) Interactive dramaturgy by generating acousmêtre in a virtual environment. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics **4161**:90
171. Matthias P, Billinghurst M, See Z (2019) This land ar: An australian music and sound xr installation a transmedia storytelling approach, Proceedings -VRCAI 2019: 17th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry
172. Nicholas M, Daffara S, Paulos E (2021) Expanding the design space for technology-mediated experiences, DIS 2021 -Proceedings of the 2021 ACM Designing Interactive Systems Conference: Nowhere and Everywhere pp. 2026–2038
173. Rzayev R, Karaman G, Wolf K, Henze N, Schwind V (2019) The effect of presence and appearance of guides in virtual reality exhibitions, ACM International Conference Proceeding Series pp. 11–20
174. J. Cui, F. Dong, W. Zhang, in 2020 International Conference on Innovation Design and Digital Technology (ICIDDT) (IEEE, Zhenjing, China, 2020), pp. 165–169. https://doi.org/10.1109/ICIDDT52279.2020.00037
175. Fan Z, Dai R, Xu Y, Wang X, Chen S (2021) in Proceedings - 2021 2nd International Conference on Intelligent Design, ICID 2021, pp. 120 – 128. https://doi.org/10.1109/ICID54526.2021.00032. Type: Conference paper
176. Chong HT, Lim CK, Rafi A, Tan KL, Mokhtar M (2021) Comprehensive systematic review on virtual reality for cultural heritage practices: coherent taxonomy and motivations. Multimedia Syst 28(3):711–726. https://doi.org/10.1007/s00530-021-00869-4
177. Jerald J (2015). The VR Book (Association for Computing Machinery). https://doi.org/10.1145/2792790
178. Brinkmann F, Lindau A, Weinzierl S (2017) On the authenticity of individual dynamic binaural synthesis. The Journal of the Acoustical Society of America 142(4):1784. https://doi.org/10.1121/1.5005606
179. Xie B (2013) Head-related transfer function and virtual auditory display, 2nd edn. J Ross Publishing, Boca Raton, FL
180. Raghuvanshi N, Gamper H (2022) Interactive and Immersive Auralization (Springer International Publishing), p. 77–113. https://doi.org/10.1007/978-3-031-04021-4_3
181. Välimäki V, Franck A, Rämö J (2015) Assisted listening using a headset: Enhancing audio perception in real, augmented, and virtual environments. IEEE Signal Processing Magazine 32:92
182. D'Auria D, Di Mauro D, Calandra DM, Cutugno F (2014) Caruso: interactive headphones for a dynamic 3D audio application in the cultural heritage context. In: Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration (IEEE IRI 2014), pp. 525–528. https://doi.org/10.1109/IRI.2014.7051934
183. Tse A, Jennett C, Moore J, Watson Z, Rigby J, Cox A (2017) Was i there?: Impact of platform and headphones on 360 video immersion, Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems pp. 2967–2974
184. Kaghat F, Azough A, Fakhour M (2018) Sarim: A gesture-based sound augmented reality interface for visiting museums, 2018 International Conference on Intelligent Systems and Computer Vision, ISCV 2018 pp. 1–9

185. Giariskanis F, Kritikos Y, Protopapadaki E, Papanastasiou A, Papadopoulou E, Mania K (2022) In: IMX 2022 - Proceedings of the 2022 ACM International Conference on Interactive Media Experiences , pp. 281 – 285. https://doi.org/10.1145/3505284.3532967. Type: Conference paper
186. Kritikos Y, Mania K (2022) In: 2022 International Conference on Interactive Media, Smart Systems and Emerging Technologies (IMET), pp. 1–8. https://doi.org/10.1109/IMET54801.2022.9929500
187. Chaurasia HK, Majhi M (2022) Sound Design for Cinematic Virtual Reality: A State-of-the-Art Review, Lecture Notes in Networks and Systems **391**:357. Type: Conference paper
188. Fellgett P (1973) Ambisonic reproduction of sound. Electronics and Power 19(20):492. https://doi.org/10.1049/ep.1973.0597
189. Bostan B, Marsh T (2012) Fundamentals of interactive storytelling. Summer 3(8):19. https://doi.org/10.5824/1309-1581.2012.3.002.x
190. Riedl M, Thue D, Gomez-Martín M (2011) Game ai as storytelling, Artificial Intelligence for Computer Games pp. 125–150
191. Hernandez P, Bulitko S, St V, Hilaire E (2021) Emotion-based interactive storytelling with artificial intelligence. AIIDE 10:146
192. Pisoni G, Díaz-Rodríguez N, Gijlers H, Tonolli L (2021) Human-centered artificial intelligence for designing accessible cultural heritage. Applied Sciences 11(2):870. https://doi.org/10.3390/app11020870
193. Deacon T, Barthet M (2022) Spatial design considerations for interactive audio in virtual reality. Sonic Interaction in Virtual Environment
194. Teneggi C, Canzoneri E, di Pellegrino G, Serino A (2013) Social modulation of peripersonal space boundaries. Current Biology 23(5):406. https://doi.org/10.1016/j.cub.2013.01.043
195. Grier RA, Bangor A, Kortum P, Peres SC (2013) The system usability scale. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 57(1):187. https://doi.org/10.1177/1541931213571042
196. Sevinc V, Berkman MI (2020) Psychometric evaluation of simulator sickness questionnaire and its variants as a measure of cybersickness in consumer virtual environments. Applied Ergonomics 82:102958. https://doi.org/10.1016/j.apergo.2019.102958
197. Busselle R, Bilandzic H (2009) Measuring narrative engagement. Media Psychology 12(4):321–347. https://doi.org/10.1080/15213260903287259
198. Harmon-Jones C, Bastian B, Harmon-Jones E (2016) The discrete emotions questionnaire: A new tool for measuring state self-reported emotions. PLOS ONE 11(8):e0159915. https://doi.org/10.1371/journal.pone.0159915
199. Slater M, Brogni A, Steed A (2003) in Presence 2003: The 6th annual international workshop on presence, vol. 157
200. Piumsomboon T, Lee Y, Lee G, Dey A, Billinghurst M (2017) Empathic mixed reality: Sharing what you feel and interacting with what you see, 2017 International Symposium on Ubiquitous Virtual Reality (ISUVR) pp. 38–41
201. Parisi GI, Kemker R, Part JL, Kanan C, Wermter S (2019) Continual lifelong learning with neural networks: A review. Neural Networks 113:54. https://doi.org/10.1016/j.neunet.2019.01.012
202. Cadet LB, Chainay H (2020) Memory of virtual experiences: Role of immersion, emotion and sense of presence. International Journal of Human-Computer Studies 144:102506. https://doi.org/10.1016/j.ijhcs.2020.102506
203. Milgram P, Kishino F (1994) A taxonomy of mixed reality visual displays. IEICE TRANSACTIONS on Information and Systems 77(12):1321
204. Valimaki V, Franck A, Ramo J, Gamper H, Savioja L (2015) Assisted listening using a headset: Enhancing audio perception in real, augmented, and virtual environments. IEEE Signal Processing Magazine 32(2):92. https://doi.org/10.1109/msp.2014.2369191
205. Skarbez R, Smith M, Whitton M (2023) It is time to let go of virtual reality. Communications of the ACM 66(10):41. https://doi.org/10.1145/3590959
206. Al-Imamy SY (2019) Blending printed texts with digital resources through augmented reality interaction. Education and Information Technologies 25(4):2561. https://doi.org/10.1007/s10639-019-10070-w
207. Zahorik P, Jenison RL (1998) Presence as being-in-the-world. Presence: Teleoperators and Virtual Environments 7(1):78. https://doi.org/10.1162/105474698565541
208. Varela FJ, Thompson ET, Rosch E (1992) The embodied mind. The MIT Press (MIT Press, London, England)
209. Perry S, Roussou M, Economou M, Young H, Pujol L (2017) Moving beyond the virtual museum: Engaging visitors emotionally, 23rd International Conference on Virtual System & Multimedia (VSMM) pp. 1–8

210. Gaver W, Krogh PG, Boucher A, Chatting D (2022) Emergence as a feature of practice-based design research. In: Proceedings of the 2022 ACM Designing Interactive Systems Conference (Association for Computing Machinery, New York, NY, USA), DIS '22, pp. 517–526. https://doi.org/10.1145/3532106.3533524
211. Wiltse H (2020) Relating to Things: Design, Technology and the Artificial (Bloomsbury Publishing). Google-Books-ID, SyTXDwAAQBAJ
212. Privitera AG, Geronazzo M (2024) Designing Sonic Interactions in Intelligent Reality with Egocentric Audio Technologies. Routledge Handbook on Sound Design (to appear) (Routledge, Taylor & Francis Group)
213. Skarbez R, Polys N, Ogle J, North C, Bowman D (2019) Immersive analytics: Theory and research agenda, Frontiers in Robotics and AI 6
214. Chandler T, Cordeil M, Czauderna T, Dwyer T, Glowacki J, Goncu C, Klapperstueck M, Klein K, Marriott K, Schreiber F, Wilson E (2015) Immersive analytics, Big Data Visual Analytics (BDVA) p. 7314296
215. Hermann T (2008) In: Proceedings of the 14th International Conference on Auditory Display (International Community for Auditory Display)
216. Koebel K, Agotai D, undefinedöltekin A (2020) Exploring cultural heritage collections in immersive analytics: Challenges, benefits, and a case study using virtual reality, The International Archives of the Photogrammetry. Remote Sensing and Spatial Information Sciences XLIII-B4-2020: 599–606. https://doi.org/10.5194/isprs-archives-XLIII-B4-2020-599-2020
217. Athaluri SA, Manthena SV, Kesapragada VSRKM, Yarlagadda V, Dave T, Duddumpudi RTS (2023) Exploring the boundaries of reality: Investigating the phenomenon of artificial intelligence hallucination in scientific writing through chatgpt references. Cureus. https://doi.org/10.7759/cureus.37432
218. Ciaburro G, Berardi U, Iannace G, Trematerra A, Puyana-Romero V (2020) The acoustics of ancient catacombs in southern italy. Building Acoustics 28(4):411. https://doi.org/10.1177/1351010x20967571
219. Warusfel O, Emerit S (2021) In: 2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA) (IEEE). https://doi.org/10.1109/I3DA48870.2021.9610973
220. Tahvanainen H, Matsuda H, Shinoda R (2019) Numerical simulation of the acoustic guitar for virtual prototyping. Proceedings of ISMA 2019:13–17
221. Lehman B (2005) Bach's extraordinary temperament: Our rosetta stone-1. Early Music 33(1):3
222. Cossou L, Louison C, Bouchigny S, Ammi M (2018) In: Proceedings of the 1st International Conference on Digital Tools & Uses Congress - DTUC '18 (ACM Press), DTUC '18. https://doi.org/10.1145/3240117.3240129
223. Montagud M, Orero P, Matamala A (2020) Culture 4 all: accessibility-enabled cultural experiences through immersive VR360 content. Personal and Ubiquitous Computing 24(6):887. https://doi.org/10.1007/s00779-019-01357-3
224. Montuwy A, Cahour B, Dommes A (2017) Visual, auditory and haptic navigation feedbacks among older pedestrians. Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services pp. 1–8
225. Barbieri T, Bianchi A, Sbattella L (2004) Minus-two: Multimedia, sound spatialization and 3d representation for cognitively impaired children. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics 3118: 1054
226. di Furia M, Guarini P, Finestrone F (2022) In: Proceedings of the Third Workshop on Technology Enhanced Learning Environments for Blended Education , pp. 10–11
227. Pareto L, Snis UL (2006) Understanding users with reading disabilities or reduced vision: Toward a universal design of an auditory, location-aware museum guide. International Journal on Disability and Human Development f5(2). https://doi.org/10.1515/ijdhd.2006.5.2.147
228. Hamilton R (2007) Maps and legends: Fps-based interfaces for composition and immersive performance, International Computer Music Conference pp. 344–347
229. Hamilton R (2008) Maps and legends: Designing fps-based interfaces for multiuser composition, improvisation and immersive performance. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics 4969:478
230. Haladova Z, Boylos C (2012) Held at the international multidisciplinary modeling and simulation multiconference. 11th International Conference on Modeling and Applied Simulation 2012:I3M
231. Nagao S, Naemura T (2012) Saion: Selective audio image reproduction system using multiple hyper directional loudspeakers, ACM SIGGRAPH 2012 Posters
232. Tallig A, Hardt W, Eibl M (2013) Border crosser a robot as mediator between the virtual and real world. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics 8014:411

233. Markert M, Heitjohann J, Geelhaar J (2014) Phonorama: mobile spatial navigation by directional stereophony, Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services pp. 609–611
234. Cubillo J, Martín S, Castro M, Díaz G, Colmenar A, Botički I (2015) A learning environment for augmented reality mobile learning. Proceedings -Frontiers in Education Conference
235. Pérez Y, Berres S, Rodríguez E, Rodríguez S, Antúnez G, Mercado A, Soledad M, Jara C, Ulloa M (2015) Usability principles for the design of virtual tours, Proceedings -21st International Congress on Modelling and Simulation. MODSIM 2015:1876–1881
236. Pozzebon A, Calamai S (2015) Smart devices for intangible cultural heritage fruition, Digital Heritage pp. 333–336
237. Sprung G, Egger A, Nischelwitzer A, Strohmaier SS (2018) Virest -storytelling with volumetric videos. CEUR Workshop Proceedings pp. 54–59
238. Darcy D, Brandmeyer A, Graff R, Swedlow N, Crum P (2019) Methodologies for assessment of speech and audio for optimized quality of experience, Proceedings of the International Congress on Acoustics pp. 6129–6136
239. Du R, Li D, Varshney A (2019) Experiencing a mirrored world with geotagged social media in geollery, Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems pp. 1–4
240. Indans R, Hauthal E, Burghardt D (2019) Towards an audio-locative mobile application for immersive storytelling. KN - Journal of Cartography and Geographic Information 69(1):41. https://doi.org/10.1007/s42489-019-00007-1
241. Mathioudakis G, Klironomos I, Partarakis N, Papadaki E, Anifantis N, Antona M, Stephanidis C (2021) Supporting online and on-site digital diverse travels. Heritage 4(4):4558
242. Constantinou S, Pamboris A, Alexandrou R, Kronis C, Zeinalipour-Yazti D, Papadopoulos H, Konstantinidis A (2022) EnterCY: a virtual and augmented reality tourism platform for Cyprus. In: Proceedings - IEEE International Conference on Mobile Data Management, vol. 2022-June (2022), vol. 2022-June, pp. 314 – 317. https://doi.org/10.1109/MDM55031.2022.00069. Type: Conference paper
243. Butterworth A (2022) Beyond sonic realism: a cinematic sound approach in documentary 360° film. Studies in Documentary Film 16:156
244. Mulvany G (2022) Because the night -immersive theatre for digital audiences: Mapping the affordances of immersive theatre to digital interactions using game engines, Extended Abstracts of the 2022 Annual Symposium on Computer-Human Interaction in Play pp. 291–296
245. Popp C, Murphy D (2022) Creating audio object-focused acoustic environments for room-scale virtual reality, Applied Sciences (Switzerland) p. 12
246. Vasquez Gomez JC (2022) In: Extended Abstracts of the 2022 Annual Symposium on Computer-Human Interaction in Play (Association for Computing Machinery, New York, NY, USA), CHI PLAY '22, p. 103-106. https://doi.org/10.1145/3505270.3558341
247. Tzi-Dong Ng J, Hu X, Que Y (2022) In: LAK22: 12th International Learning Analytics and Knowledge Conference (Association for Computing Machinery, New York, NY, USA), LAK22, pp. 451–457. https://doi.org/10.1145/3506860.3506881. Event-place: Online, USA
248. Sheremetieva A, Romanv I, Frish S, Maksymenko M, Georgiou O (2022) In: 2022 International Conference on Interactive Media, Smart Systems and Emerging Technologies (IMET), pp. 1–3. https://doi.org/10.1109/IMET54801.2022.9929479

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.