DIGITAL HERITAGE (2025)
S. Campana, D. Ferdani, H. Graf, G. Guidi, Z. Hegarty, S. Pescarin, and F. Remondino (Editors)

Swiss Echoes: An immersive and embodied exploration of a national broadcasting archive

G. Alliata[1], L. Serafin[1], A. Rattinger[1] and S. Kenderdine[1]

[1]Laboratory for Experimental Museology, EPFL, Switzerland

Abstract

This paper answers the need for new modes of access for large audiovisual collections and builds on top of the computational and immersive turn of the cultural sector. The immersive installation Swiss Echoes fosters an embodied and spatialised paradigm in which visitors fly over a topographical 3D map of the Swiss landscape to discover how the voices captured in the national broadcasting archive of the Radio Télévision Suisse talk about local places.

Through a reflection on the design rationale supported by key insights from a user evaluation conducted at our laboratory, three main themes are elucidated. First, the computational augmentation of the archive foregrounds access to a collection of locations rather than videos. Second, the use of an immersive environment and a performative controller fosters a profoundly embodied mode of spatial exploration. Third, sharing the immersive experience between multiple visitors creates a collective form of engagement, where the main visitor interacting becomes a performer and director for the others.

CCS Concepts

• ***Applied computing*** → ***Arts and humanities;*** • ***Human-centered computing*** → ***Visualization;*** • ***Computing methodologies*** → *Information extraction;*

1. Introduction

Audiovisual (AV) archives constitute some of the most significant mnemonic records of the 20th and 21st centuries. As complex media objects, AV materials capture the multi-layered dimensions of culture through the interplay of audio and visual channels. Beyond textual documents and still images, these media encapsulate gestures, expressions, voices, and soundscapes, offering a rich and embodied account of past events, everyday life, and cultural practices. Their unique capacity to convey both explicit and tacit aspects of human experience positions AV archives as critical sources for understanding contemporary history and collective memory [Bru17].

Over the past decades, large-scale digitisation efforts by cultural heritage institutions, ranging from galleries, libraries, archives, and museums (GLAM) to public and private broadcasters, have dramatically expanded the availability of cultural collections [Thy19]. For the particular case of AV archives, Radio Télévision Suisse (RTS) has digitised over 200,000 hours of footage [RTS18], the British Broadcasting Corporation (BBC) holds more than a million recorded hours [Wri17], and the Netherlands Institute for Sound and Vision (NISV) has invested hundreds of millions to digitise Dutch AV heritage [vEKO*17]. The proliferation of born-digital audiovisual content further amplifies this exponential growth. As the scale of these archives increasingly surpasses the capacity of traditional curation and access methods, there is a growing imperative to adopt computational approaches for organising, analysing, and accessing AV materials [KMH21, CBJN21, F*12].

At the same time, this transformation calls for a rethinking of how non-expert audiences access and engage with audiovisual heritage. Conventional text-based interfaces are often inadequate for navigating the semantic complexity of AV material [MO21]. In response, cultural institutions are increasingly turning to immersive and interactive technologies to foster more embodied, affective, and intuitive forms of engagement. Building on Gumbrecht's call for a shift towards a 'culture of presence' [Gum04], this so-called 'immersive turn' in the museum [Kid18, SS23] reflects a broader trend in the GLAM sector: the integration of interactive systems and sensory environments to facilitate situated, experiential encounters with digitised collections. By 'amalgamating cultural heritage archives with interactive cinema to foster novel forms of embodied narrative' [KMH21, p.7], non-expert audiences can more intuitively explore these complex and vast archives.

In this paper, we present *Swiss Echoes*, an interactive and immersive installation designed to explore the television archive from Radio Télévision Suisse (RTS), the Swiss national broadcasting company. The installation builds on prior computational work in which videos were geolocated based on place names automatically extracted from the spoken content [ABRK24, RABK25]. A second semantic layer was added by classifying each video into broad topical categories using a large language model (LLM) applied to the

DOI: 10.2312/dh.20253313

transcripts. The installation runs on the Panorama+, a 360-degree stereoscopic projection system capable of simultaneously accommodating up to twenty visitors.

We focus here on the concept and design rationale of the installation, supported by insights from a user study focusing on the spatial exploration of the collection and informal visitor observations during tours at our laboratory. The evaluation was conducted during the development of the installation with 23 students from a local university. After interacting with the installation in groups of seven to fifteen people, participants completed a post-hoc questionnaire on an iPad with the *muse* tool [KK15]. The survey included the short-form User Engagement Scale [OCH18], the Spatial Presence Experience Scale [HWS*16], custom questions (three of which probed the social experience of multi-user immersion) and basic demographic items, including a self-report of familiarity with Swiss life and culture. While a full report is beyond the scope of this paper, we note that Confirmatory Factor Analysis confirmed strong internal consistency, with high item loadings (> 0.75 for most items) and good discriminant validity (highest factor correlation: 0.27), including for the custom three-items *social* scale. While this custom scale cannot fully capture complex dynamics like social awareness and mutual influence, it remains a useful indicator when paired with informal observations during lab tours.

Through this contribution, we examine how augmenting and mapping an audiovisual collection through computational processes to then explore it in an immersive experience supports novel forms of engagement with large audiovisual archives. After detailing the computational pipeline used to enrich and reorganise the archive, we describe the immersive interface and interaction model developed for the installation *Swiss Echoes*. We then reflect on the design rationale and visitor experience, drawing on insights from the user study. In doing so, we highlight how spatial navigation, embodied interaction, and social dynamics can reshape how audiovisual memory is accessed.

2. Conceptual background

2.1. Computational Augmentation of Audiovisual Archives

Audiovisual (AV) archives pose unique challenges for access and organisation due to their multimodal nature and limited pre-existing metadata [MOvNF20, MO21]. The richness of embedded layers in AV materials, in particular, often cannot be captured verbally and is thus missing from traditional metadata schemas. Recent advances in computational processing, however, have opened new pathways to engage with these materials at scale [Her, OMZ*24], unlocking the varied semantic layers of AV materials. Focusing on the spoken content in particular, scholars have initiated 'distant listening' approaches [Cle16] while broadcasting companies have adopted transcription technologies in their databases [Baz23].

Automatic speech recognition, Named Entity Recognition (NER), and large language models (LLMs) allow for extracting content-level and descriptive metadata that can be used to enrich, structure, and classify vast AV corpora. As several scholars have noted, analysing the actual material of AV archives through computational means allows us to move beyond traditional logics of provenance and original order, focusing instead on what actually is in the records rather than how they were originally filed or described [CBJN21, KMH21]. This shift is particularly relevant for archives that were not initially intended for long-term public access or research, such as the archive of the broadcasting company considered in this paper.

2.2. Embodied Access and Immersive Interfaces

Alongside computational restructuring, cultural institutions are increasingly exploring embodied and immersive modalities for engaging with digitised heritage. As [KMH21] observes, digitisation enables archives to move beyond static systems of storage and retrieval and into the 'social sphere of immersed experience' (p. 6). This transformation resonates with the 'immersive turn' in the museum sector [Kid18, SS23], where spatial presence, gesture-based interaction, and multisensory feedback are mobilised to create more open-ended, exploratory and participatory forms of access. This shift in museological practices emphasises how visitors engage with immersive installations through their bodies and motion habits, resulting in evolving epistemological paradigms [Cal23].

This rethinking of archival engagement finds particular traction in experimental museology, which foregrounds the visitor's role as an agent of interaction within cultural systems [Ken15, KMH21]. Immersive environments are not simply containers for content, but are themselves modes of framing and structuring visitors' engagement with data, because 'memories are not inherent in the archival stock, but are created in the context of reception, through processes of remediation and recontextualisation' [Bru17, p.98]. Especially in contexts where linear or search-based methods are insufficient to support discovery, such as with AV material due to their semantic and multimodal complexity, developing new modes of access for diverse and non-expert audiences is of prime importance [F*12].

The installation *Swiss Echoes* discussed in this paper builds on this lineage, offering an embodied mode of interaction with a semantically structured AV dataset that positions movement, orientation, and auditory proximity as primary means of accessing the archive. By enabling visitors to navigate a 3D topographical map of Switzerland, hear geolocated sound excerpts, and select video clips organised by thematic categories, the system transforms the AV archive into an explorable landscape. Through their interactions, visitors shape their own journey through the audiovisual collection in an immersive and embodied way. They access the archive in a spatial manner that supports a more intuitive mode of navigation than merely accessing specific videos through a traditional search-based interface.

3. Computational augmentation and mapping of the collection

The installation presented in this paper is grounded in the Radio Télévision Suisse (RTS) archive, a large-scale audiovisual collection comprising 200'000 hours of French-language footage. While this collection is accompanied by existing metadata, it is often inconsistent and primarily deals with distribution-related descriptors. There are also semantic descriptors, including locations, but they are at the level of full videos that can extend for several minutes to over an hour, making them unsuitable for interactive or immersive presentation without further segmentation. These descriptors

are also dependent on evolving curatorial practices at RTS over the years, without assurance that the entire archive is treated equally. Therefore, we decided to develop a computational pipeline to automatically extract all named locations specifically mentioned in the audio script [ABRK24], of which we present the final iteration here. This pipeline builds on a custom back-end infrastructure we have developed at the laboratory to ingest and process large AV archives [RABK25].

Spoken language is inherently rich in information (naming places, events, individuals, and themes), but it is difficult to access and analyse at scale without computational tools. To address this gap, we developed a processing pipeline that focuses specifically on operationalising the speech embedded within audiovisual material. This transformation aligns with what [MSC] defines as datafication: rendering phenomena into a quantified format so they can be tabulated, analysed, and recombined. [EV21] further characterises data as potential information in digital form, meaningful only once processed by machines, a concept that aptly describes the status of spoken language in large archives: named locations are at once easily accessible, it suffices to listen to the video, and invisible at scale, considering the hundreds of thousands of hours in this collection.

The pipeline described here transforms the archive into what [KMH21] defines as a 'computational archive'. This transformation brings about three fundamental shifts in the way archives are conceived and engaged with (p. 6). First, the archive transitions from being a static repository optimised for search and retrieval to becoming a dynamic, social environment for immersive experience. Second, the role of the archivist expands beyond technical stewardship to include active cultural mediation, fostered through collaboration across disciplines such as museology, interaction design, and computer science. Third, the digitally processed features of the archive, its spoken content in this case, take on new expressive forms through spatial, temporal, sensory, and aesthetic structures that enable alternative pathways for access and interaction. This transformation is particularly important for AV collections due to the complex nature of their different semantic layers that are usually not captured by traditional metadata [MOvNF20].

The transformation begins with automated speech-to-text transcription using the WhisperX model [RKX*22, BHHZ23], making the spoken content machine-readable. To enable finer-grained interaction, speaker diarization segments the original recordings into shorter clips based on speaker changes. On these transcribed segments, Named Entity Recognition (NER) is applied using spaCy to extract place names [HMLB20]. Each identified location is then geolocated by querying Wikidata for its latitude and longitude. This enables the content to be spatially restructured: instead of navigating the archive by categories or dates, visitors can explore it as a dynamic map, organised by the places referenced in the speech. Note that this first segment of the pipeline is part of a previously published work.

While the full archive comprises over 200,000 hours of material, a first sample of 18.6k hours was processed with the above pipeline for computational costs and processing times. This sample yielded 10k locations, of which 541 were identified in Switzerland. The final subset the installation currently operates on consists of 260 hours of footage, corresponding to 26k clips that had named locations out of a set of 541 identified places in Switzerland. Even though this represents only a tiny fraction of the whole archive, it is still a large enough amount of content for the kind of interactive installation we are aiming for.

To illustrate the result of the geolocation process, Figure 1 shows a two-dimensional map of Switzerland annotated with the extracted locations from the sample dataset. The size of each point reflects the number of clips associated with that location, highlighting the clearly skewed coverage towards bigger cities such as Geneva or Lausanne, even though many small villages do still appear on the map.

Finally, an additional layer of computational augmentation has been implemented to visually structure content within a specific location. This was particularly necessary to aid navigation in the larger cities, where visitors would otherwise have been confronted with thousands of videos without any navigational structure.

To this end, we employed a large language model (LLM) to further process the previously generated speech-to-text transcripts and assign them to broad thematic categories (such as "politics" or "culture"), leveraging the computational infrastructure built for another project at the laboratory. Rather than enforcing exhaustive taxonomies of the AV material, the goal of this additional semantic layer is to enable thematic browsing within a location with a lightweight visual structuring into cylindrical thematic rings of content, as it will be explained in the next section.

In summary, the computational pipeline described here semantically enriches the audiovisual material, fragmenting the full videos into more digestible clips, precisely geolocating them on a map of Switzerland and assigning them to broad thematic categories. This wealth of 'potential information' [EV21, p.3] in the spoken layer of the AV archive unlocked through computational means serves as the basis for the spatial and semantic exploration of the collection in the immersive installation *Swiss Echoes*, as described in the following section.

4. *Swiss Echoes* - Description of the interactive installation

4.1. Technical description

The immersive installation is hosted in the Panorama+, a large-scale, 360-degree stereoscopic projection environment based on the AVIE system [MSK*07]. This immersive platform places embodiment at the core of the interaction, allowing visitors to physically situate themselves within a spatialised digital archive. By engaging both egocentric (visitor-centred) and allocentric (object-centred) perspectives [Ken15], the Panorama+ facilitates embodied engagement, enabling visitors to orient themselves both within and in relation to the entities they encounter in the virtual world. Unlike today's common museological installations that fully immerse visitors in a virtual world via head-mounted displays, the installation aligns with Dourish's model of embodied interaction, which seeks to augment lived experience rather than replace it [Dou01, CR21]. Here, in particular, visitors explore the collection as a group, therefore augmenting the social nature of a museum visit [FD16], instead of being completely cut off in a 'dislocated place' [SF19].

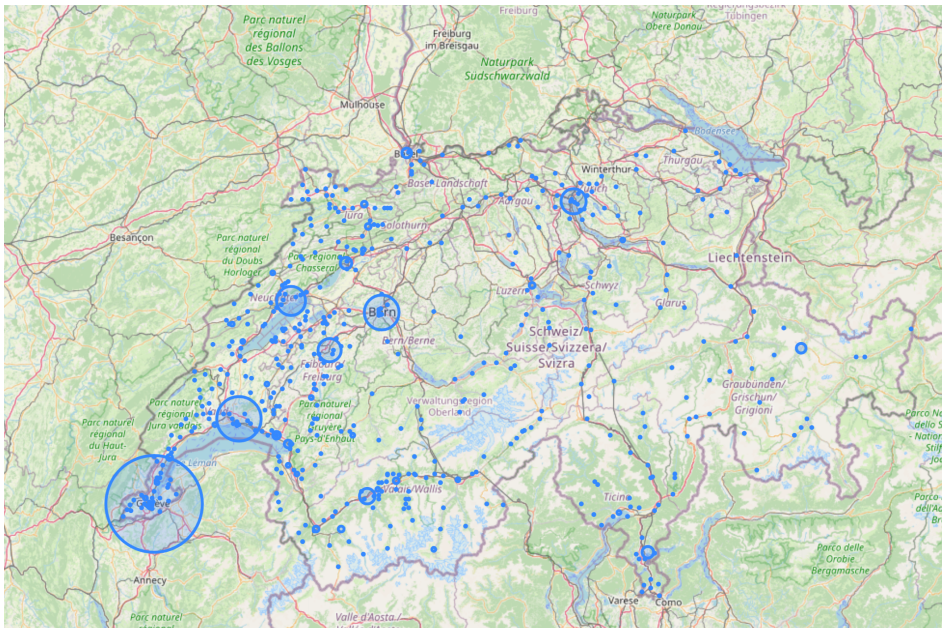

Figure 1: *Geolocated distribution of video clips across Switzerland, based on named locations extracted from the transcribed audio content. The visualisation reveals a concentration of references in the French-speaking regions and major urban centres, reflecting the institutional provenance of the dataset.*

Technically, the Panorama+ comprises five 4K projectors arranged in a 360-degree configuration, as shown in Figure 2. Each projector renders a stereoscopic pair of images, one for each eye, yielding ten simultaneous video outputs. These are powered by a cluster of five high-performance computers, each equipped with two GPUs synchronised via QuadroSync. The resulting display surface measures 2160 pixels in height and approximately 16,000 pixels in width, accounting for the overlaps in the blend zones between projections. Visitors wear active 3D glasses to perceive the stereoscopic depth that characterises the immersive visual experience. Finally, the Panorama+ is complemented by a custom 32-speaker audio array allowing full spatialisation of sounds around the 360-degree display, greatly contributing to the sensory and immersive experience [BS07].

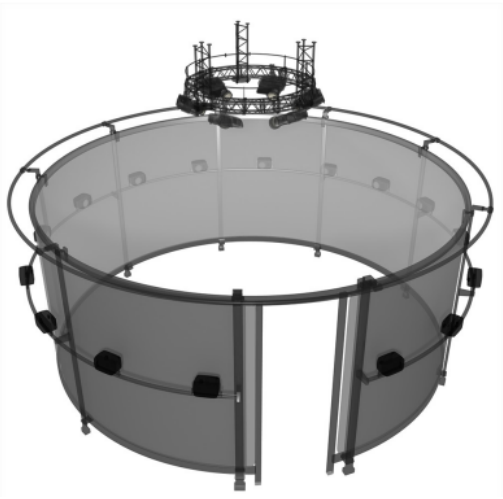

Figure 2: *Schematic of the Panorama+, a 360-degree stereoscopic environment enabling embodied exploration of digital archives through egocentric and allocentric perspectives.*

The application *Swiss Echoes* was developed using Unreal Engine 5, with the nDisplay system employed to distribute rendering across the Panorama+ computing cluster. One node acts as the coordinator, orchestrating the synchronisation of the ten rendering instances, two per projector, to generate a unified virtual environment. The choice of Unreal Engine and nDisplay was driven by the need for high-fidelity rendering across a large-scale, stereoscopic display, while maintaining precise frame synchronisation across multiple projectors. Achieving a coherent 360-degree visual environment is essential for sustaining the illusion of continuous space and supporting embodied forms of navigation. To complement this environment, interaction is mediated through an HTC Vive controller, which enables intuitive, gesture-based input.

4.2. Global Layout: Geolocated Content

The organisation of content within the installation follows a two-level spatial-semantic mapping, rooted in the computational structuring described in Section 3. At the global level, each video clip is geolocated according to place names referenced in its transcribed speech. These locations are visualised on a 3D topographical model

of Switzerland, built based on topographical data from Swisstopo, Switzerland's national mapping agency [**?**], and labelled by name. At each location, the corresponding clips, represented by flying cubes with slightly rounded corners that loosely resemble CRT televisions, are arranged into rising columns that visually indicate from a distance where archival content is located.

Visitors fly over this map using the HTC Vive controller. By pointing in a direction and holding the rear trigger, they glide forward, navigating the landscape at a controllable speed, in an elevated position. This design fosters a direct, embodied relationship between the visitor's gestures and the system's responses, aligning physical pointing with spatial orientation in the virtual environment. As they fly near a location, they begin to hear spatialized audio excerpts drawn from clips associated with that place, thus hearing the mentioned place name. These sound cues, serving as previews to the geolocated AV content, are positioned directionally around the 360-degree space thanks to the 32-speaker array, encouraging embodied turning, listening, and orienting toward points of interest.

Therefore, as visitors are navigating the 3D virtual world, they are provided with both visual and aural cues. On the one hand, the location names at ground level and the Swiss topography visually guide them on the 3D map, while the rising spires of floating video cubes intuitively indicate where archival content is located. On the other hand, the directional sound excerpts support serendipitous discoveries, prompting visitors to turn towards a location they are passing by as they get curious about what they just heard.

When a visitor selects a location by pointing and pressing the front button of the HTC Vive Controller, the system automatically flies to the selected location and transitions to a local exploration mode, presenting a cylindrical wall of CRT televisions filled with video thumbnails (Figure 4). This marks the second level of the spatial-semantic structure.

4.3. Local Layout: Thematic Cylinders

Each selected location is visualised as a cylindrical wall of CRT televisions, where each horizontal ring of screens corresponds to one of the thematic categories previously assigned via LLM classification. The clips are arranged radially around the visitor, and a single clip may appear in multiple rings if it belongs to more than one category. Because the number of clips per location varies widely (from a few in small villages to thousands in cities like Geneva or Zurich), the wall of televisions is populated modularly: clips are duplicated to meet a minimum threshold or randomly sampled when too many are available.

This cylindrical structure supports embodied, spatial browsing: visitors physically turn to face different directions and look upward or downward to explore thematic layers. The form makes the density and diversity of clips at each location visually legible, while also encouraging discovery through bodily movement rather than linear scanning. The thematic rings do not serve as strong interpretative tools (in fact, the label indicating the category is deliberately designed in a non-prominent way) but are more intended to lightly guide navigation.

Visitors engage with the cylindrical wall of video thumbnails using the HTC Vive controller, which functions as an aural torch. By pointing at a specific screen, they dynamically reveal the corresponding audio, allowing for preliminary auditory inspection of clips before committing to playback. This mode of interaction enables visitors to listen directionally as they explore the spatial arrangement of content, aligning auditory focus with gestural orientation.

When a visitor decides to engage more deeply with a clip, they press the main front button on the controller to activate video playback. Up to ten videos can be played simultaneously, with only the clip currently under inspection emitting sound. This constraint is governed by real-time performance requirements, ensuring a responsive and coherent audiovisual experience. The design thus encourages embodied and selective engagement, supporting an exploratory rhythm of scanning, listening, and curating one's own audiovisual trajectory through the space.

The two-level spatial-semantic structure and layered sound interaction invite visitors to build their own exploratory pathways, dynamically weaving together fragments of the AV archive through sight and sound. This design avoids imposing a linear view of the collection. Instead, it offers a situated orientation within a complex and semantically rich dataset, with an interface that invites visitors to discover relationships between place and theme, between individual clips and broader cultural contexts. Each location becomes not only a geographic anchor but also a semantic container, within which the visitor can browse across layers of meaning.

5. Design rationale and discussion

5.1. Epistemological Shift: Navigating a Collection of Locations

One of the most consequential outcomes of the computational pipeline described in Section 3 is the epistemological shift it introduces in how visitors engage with the RTS audiovisual archive. Rather than beginning with predefined metadata categories such as genre, date, or broadcast title, visitors are instead presented with a geospatial interface in which named locations, extracted from the spoken content itself, serve as the primary entry points into the archive.

This shift effectively transforms the archive into a collection of places, through which visitors access its contents. The places themselves do not originate from curatorial metadata but are instead surfaced by the archive's own language: extracted from within the transcribed speech of the audiovisual recordings. In doing so, the pipeline foregrounds the spatial logic already embedded in the material and reorganises the archive to make this logic visible and navigable. Audiovisual collections are indeed particularly rich repositories of 'the culture that people practise as part of their daily lives' [Kur04]. This dimension of intangible cultural heritage is often deeply embedded in place, reflecting everyday life, local festivities, and regionally specific forms of expression.

Crucially, this spatialization does not just support interaction: it already constructs knowledge. It acts as what digital humanist scholar Johanna Drucker has termed a 'diagrammatic interface'

Figure 3: *Visitors engaging with Swiss Echoes in the Panorama+ system. The large-scale 360-degree stereoscopic projection enables shared immersive exploration of a geolocated audiovisual archive, with content dynamically arranged over a 3D topographical map.*

Figure 4: *Inside a selected location, visitors encounter a cylindrical wall of video thumbnails arranged by semantic category. Using the controller, visitors can activate clips and selectively reveal sound, curating their own audiovisual experience.*

[Dru22]. Simply looking at the 2D plot of geolocated clips above (Figure 1) reveals a clear pattern in the distribution: references cluster in the French-speaking regions of Switzerland, and especially around major urban centres such as Geneva and Lausanne. This outcome reflects the original institutional provenance of the material, which comes from the French-language division of the national broadcaster. In this way, the spatial structuring does not just visualise the archive but also reveals its linguistic and institutional biases, and does so through a computational logic rather than an interpretive overlay. This is then reflected in the immersive experience with the rising columns of video cubes placed at each location that vividly mark the bigger cities such as Lausanne or Geneva.

Each location on the 3D map thus acts as a semantic anchor, a portal that grants access to the audiovisual material associated with

it. The visitor is not initially browsing a catalogue of videos but instead encountering a cartographic visualisation of speech references, letting the density and geography of language lead them into the content. The memory of place becomes the structure through which the memory of audiovisual material is accessed.

This dynamic was clearly evident during tours organised at the laboratory to showcase the installation in its development phase. Among Swiss audiences in particular, it was common to observe visitors intentionally navigating toward familiar locations, using both the realistic topography of the 3D map and the labels identifying named places to orient themselves. These behaviours illustrate how spatial familiarity and place-based memory can shape visitors' exploratory trajectory through a computationally augmented archive and contribute to fostering an engaging experience.

These informal observations are further supported by results from a post-experience user study. The User Engagement Scale (UES) [OCH18] showed an overall engagement score of 3.87 ± 0.13, with scores on the four sub-scales larger than 4 except for Focused Attention (3.33 ± 0.17), potentially due to the presence of multiple visitors in the immersive space. Furthermore, respondents rated their familiarity with Swiss life and culture on a scale from 1 to 5 (Swiss Culture hereafter). While no statistically significant correlations emerged, likely due to the small sample size (N = 23) and the overall high Swiss Culture score (3.91 ± 0.19, with only 6 respondents scoring 3 or less), some meaningful trends were observed. Notably, Swiss cultural familiarity showed a positive correlation with both the Reward subscale of the UES ($\rho = 0.18, \alpha = 0.417$) and the custom social engagement scale ($\rho = 0.25, \alpha = 0.248$), both computed using Spearman's rank correlation. These trends suggest that visitors with stronger local cultural knowledge were more likely to find the experience rewarding and socially engaging, particularly when encountering familiar places within the archive.

This interpretation is echoed in the open-ended responses from several participants (that we have translated from French), which vividly reflect how place recognition fostered affective resonance and a sense of personal connection. One participant (Age 21) described it as "really interesting to be able to visit one's own village through history", while another (Age 23) noted that the installation "makes us want to discover the corners of our own country." Several participants emphasised the emotional impact of encountering familiar places: "Seeing all these archives of places where I grew up… I felt a kind of nostalgia in that regard" (Age 21). Another remarked, "It's great! Since we all come from different places, it makes you want to go see your place of origin" (Age 20). Finally, one participant explicitly reflected on the system's conceptual premise, describing it as "interesting to see an archive organised not diachronically, but geographically" (Age 28). Together, these responses reinforce the idea that spatial familiarity does not simply facilitate orientation but actively shapes the visitor's cognitive and emotional trajectory through the archive.

Furthermore, this transformation exemplifies a broader epistemological shift introduced by computational approaches to heritage access. By surfacing machine-extracted patterns, geolocated references in this case, archival material is no longer navigated through curatorial hierarchies or linear taxonomies, but through latent relational structures inherent to the recordings themselves. In *Swiss Echoes*, the result is a form of situated access, in which knowledge is assembled spatially and incrementally, through movement and proximity.

This spatial reconfiguration does not replace the underlying archival structure but instead introduces an additional mode of discovery, one that aligns with exploratory behaviours rather than targeted search, particularly well-suited to diverse, non-expert audiences [LBLM13]. While some visitors were observed to markedly draw on their knowledge of Swiss geography and personal connections to the land to orient themselves within the virtual landscape, others were guided by the visible columns of AV content. One respondent (Age 21) indicated that "it was quite moving for [them] to see all these archives of places where [they] grew up. [They] felt a kind of nostalgia in that regard", while another (Age 22) reported that it was "a great way to learn more about Swiss culture." Furthermore, the audio excerpts visitors would hear as they passed by a location often sparked curiosity. In this configuration, the archive is no longer encountered as a static repository of videos, but as a landscape of spoken memory, an environment through which visitors move, listen, and create their own interpretations as the AV material resonates with their personal cultures. The result is an embodied journey across a collection of places that now serve as interfaces to historical experience.

5.2. Embodied and spatial navigation of the AV archive

Building on the spatial reorganisation of the archive, the installation is designed to support a mode of access grounded in embodied movement and spatial orientation. Instead of retrieving items via text-based search or hierarchical lists, visitors physically navigate the archive by flying over a landscape of geolocated voices and turning their bodies to engage with layers of audiovisual content thanks to their kinaesthetic sense [GJGR18] (Figure 3).

This 'kinaesthetic perception' [KI22] of the virtual landscape is further reinforced by the pointing gestures visitors perform with the HTC Vive Controller. Indeed, the wand-like device allows visitors to navigate the virtual landscape by simply pointing in any direction and pressing the rear trigger. Therefore, the interaction draws on bodily orientation and spatial perception to support an intuitive and continuous form of movement. The pointing gesture reinforces visitors' sense of agency and proprioception, the sense of bodily posture and orientation, within the system. Following Dourish's conception of embodied interaction augmenting lived experience rather than replacing it [Dou01, CR21], visitors' bodily presence becomes an active locus of engagement.

The design of this embodied navigation encourages a fluid, engaging and playful mode of exploration, as confirmed by multiple responses to the open-ended question in the survey. Moving through the virtual environment was described with words like "fun", "interesting", "playful" and "stimulating". One participant (Age 23) described it as "a very enriching immersive experience for discovering the archive in a playful and interactive way, which makes you want to explore your part of the country," while another (Age 21) shared that "the playful side made [them] want to fully dive into the experience." Visitors are not guided by specific objectives or pathways. Instead, they move at their own pace, guided

by rising columns of video content and spatialized audio cues that grow louder as they approach a cluster of clips. These directional sounds function as auditory markers and previews of the AV content, inviting visitors to turn, approach, and investigate, much like following a voice in a physical environment.

Upon selecting a location, the visitor enters a cylindrical wall of screens, a spatial container displaying all clips associated with that place (Figure 4). These are arranged in horizontal rings, each corresponding to a thematic category derived from prior classification. Visitors can activate up to ten videos by clicking on screens. However, audio is not played automatically: sound is revealed dynamically by pointing at a screen, turning the controller into an aural torch. As in the global navigation on the map, the interaction with the HTC Vive Controller draws on visitors' proprioception and kinaesthetic perception: visitors direct their attention by physically aligning their gestures with the content they wish to hear. This navigation model aligns with core principles of embodied cognition, in which perception and action are deeply interwoven [VTR92]. Visitors don't merely access the archive: they move through it and inhabit it, listen directionally, and physically curate their own audiovisual journey as they explore the panoramic wall of televisions by turning around.

The user study further supports the behaviours observed during tours organised at the laboratory. As shown in Figure 5, the installation received moderately high scores on the Spatial Presence Experience Scale, measuring the sense of 'being there' [HWS*16,SW97,SBB*22]. In particular, visitors seem to 'perceive possible actions within the media environment rather than their real environment [HWS*16, p.8] (mean score of 3.54 ± 0.19 on the Possible Actions subscale). Additionally, when asked whether "exploring the archive in this manner felt like a journey," participants strongly agreed (4.17 ± 0.23), affirming the installation's capacity to support self-directed and dynamic exploration.

Figure 5: *Boxplots of visitor agreement on the Spatial Presence Experience Scales (Likert scale 1-5). Responses indicate moderately high self-location and slightly better perceived action possibilities.*

Overall, the affordances of both the Panorama+ and the HTC Vice Controller enable an exploratory and situated mode of archival access that directly draws on visitors' bodily presence in the immersive space of the installation. The goal is not to retrieve predefined content but to cultivate a dynamic relationship with the AV material through bodily engagement and spatial discovery. In line with Gumbrecht's notion of a 'culture of presence' [Gum04] and the 'immersive turn' of the museum [Kid18, SS23], the installation asks visitors to step into the archive and directly engage with the collective memory embedded in it, engaging with the spoken content in a sensory and spatial manner. This contrasts with traditional archival interfaces, where visitors typically interpret content from a distance, mediated and already interpreted through lists or metadata, rather than experiencing it physically and perceptually.

5.3. Single-User, Multi-Spectator Dynamics

Although the installation is designed for one visitor to actively interact with it at a time, it is rarely experienced in isolation. Designed for a public and situated space, the Panorama+ accommodates multiple viewers who witness the experience together. During the user study, for instance, visitors engaged in groups of between 7 and 15 persons. In this configuration, adopting the terminology proposed in [Ree11], one visitor holds the HTC Vive controller and assumes the role of an active *participant*, while others become *spectators*, observing the interaction from a 'third-person perspective' [MM18]. This distinction is not rigidly enforced throughout the experience: visitors effortlessly shift between roles by handing each other the controller.

This asymmetry makes the social dimension of the experience evident. The participant is not only navigating the archive for themselves, but also performing the act of exploring it for others, choosing flight paths, selecting locations, activating clips, and revealing sound. In doing so, they act as 'both cameraperson and editor', in the words of media artist Jeffrey Shaw [Sha03, p.23]. This is reinforced by the controller that requires *performative gestures* clearly visible to other spectators to perform manipulations of the virtual space. Therefore, when the participant turns and points in another direction, they prompt other visitors to also turn around, *directing* them.

Spectators, in turn, may suggest directions, ask questions, or point out content they see on the screen, influencing the participant's path through the archive. In this way, the installation supports a collaborative mode of interaction, in which agency is not confined to the controller-holder alone. The act of exploration becomes a socially negotiated experience, blending individual control with group curiosity. During the tours organised at the laboratory, it was indeed common to observe spectators instructing the participant where to go next.

This social aspect is also supported by the results of the user study. As shown in Figure 6, most respondents perceived they were exploring the archive with other visitors, that they were aware of their presence while interacting and that this presence of other visitors influenced them. These results suggest a strong sense of co-presence and social awareness. Visitors, as they assume the role of the *participant*, enact what [DH08] described as 'performed perception': they are aware that their gestures of attention and curiosity become visible, interpretable, and shared amongst all visitors. With a mean aggregated score of 4.28 ± 0.14, this custom social scale seems to indicate that the installation successfully foregrounds exploration of the AV collection as a shared and social act.

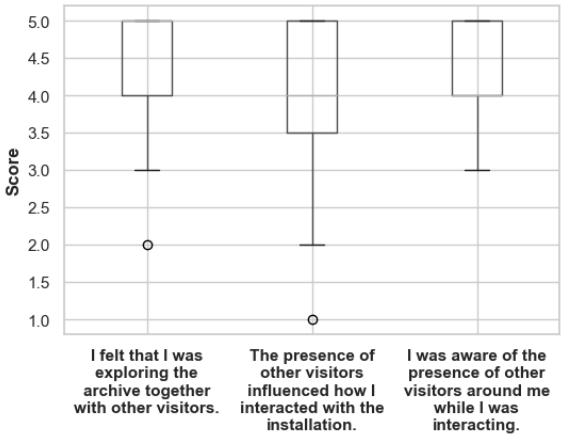

Figure 6: *Boxplots of visitor agreement on social presence items. Responses indicate awareness of and influence from other visitors during the experience.*

This model blurs the boundary between interaction and exhibition, positioning the participant simultaneously as 'operator of the system', 'performer for other people present' and 'spectator of the action in [their] immediate surroundings' [DH08, p.31]. The act of navigating the audiovisual archive becomes a mediated interface for collective engagement, where perception is externalised and shared. During the tours, visitors were seen naturally forming groups around the active participant, fostering a kind of collective interaction even though there was a single controller. Unlike solitary digital interfaces, which often render exploration invisible, this installation turns the process of accessing memory into something publicly legible, a trajectory of gaze, gesture, and sound that unfolds across a shared spatial canvas.

Thanks to the shared immersive space and the performative controller employed, the interaction and who performs it become part of the experience, expanding beyond the visual and aural inputs of the system and thus underscoring the relational and affective nature of heritage interaction in public and situated settings.

6. Conclusion

Swiss Echoes shows how combining a computational pipeline to unlock the spoken content of audiovisual material with immersive and interactive technologies fosters an intuitive and engaging mode of access to discover a broadcasting national archive.

Through the presentation of the installation concept and with support from key insights derived from a user study, three main themes are addressed. First, the computational augmentation of the Radio Télévision Suisse audiovisual archive with the mentioned locations in the spoken content is framed as a shift to a collection of places, drawing on visitors' personal connections to the local landscape to foster a recontextualised collective memory.

Second, the affordances of the Panorama+ and the HTC Vive Controller entail embodied and spatial access to the archive, dynamically revealed as visitors fly over the 3D map of Switzerland. Once arrived to a specific location, visitors' bodily presences are leveraged as well to selectively unveil video sounds and select individual clips.

Third, the presence of multiple visitors at the same time in the immersive space and the strong performative aspect of pointing at the screen to navigate and select content turn the active participant into a performer in front of spectators and into a director of the shared experience of discovering the AV archive in this manner.

7. Acknowledgments

This research is supported by the Swiss National Science Foundation through a Sinergia grant for the interdisciplinary project *Narratives from the Long Tail: Transforming Access to Audiovisual Archives*, led by co-author Prof. Sarah Kenderdine (grant number CRSII5_198632). The authors are grateful to the Radio Télévision Suisse for access to their archive as part of the *Narratives* project. The authors also wish to thank colleague Kirell Benzi for the LLM-based extraction of topic categories. Finally, the deepest gratitude is offered to Prof. Olivier Lugon and his class of students for participating in the user evaluation and offering valuable feedback.

References

[ABRK24] ALLIATA G., BENZI K., RATTINGER A., KENDERDINE S. I. B.: Ai-driven workflows for unlocking switzerland's collective memory: Distant listening of the rts archive. In *DARIAH Annual Event-Workflows: Digital Methods for Reproducible Research Practices in the Arts and Humanities* (2024). 1, 3

[Baz23] BAZÁN-GIL V.: Artificial intelligence applications in media archives. *Profesional de la información 32*, 5 (Sept. 2023). doi:10.3145/epi.2023.sep.17. 2

[BHHZ23] BAIN M., HUH J., HAN T., ZISSERMAN A.: Whisperx: Time-accurate speech transcription of long-form audio. *INTERSPEECH 2023* (2023). 3

[Bru17] BRUNOW D.: Curating access to audiovisual heritage: Cultural memory and diversity in european film archives. *Image [&] Narrative 18*, 1 (2017), 97–110. 1, 2

[BS07] BLESSER B., SALTER L.-R.: Spaces speak, are you listening. *Experiencing aural architecture 232* (2007). 4

[Cal23] CALISE A.: Inhabiting the museum: A history of physical presence from analog to digital exhibition spaces. *AN-ICON. Studies in Environmental Images [ISSN 2785-7433] 2*, II (Dec. 2023). doi:10.54103/ai/19907. 2

[CBJN21] COLAVIZZA G., BLANKE T., JEURGENS C., NOORDEGRAAF J.: Archives and ai: an overview of current debates and future perspectives. *ACM Journal on Computing and Cultural Heritage (JOCCH) 15*, 1 (2021). 1, 2

[Cle16] CLEMENT T. E.: Towards a rationale of audio-text. *DHQ: Digital Humanities Quarterly 10*, 3 (2016). 2

[CR21] CAFARO F., ROBERTS J.: Theoretical Foundations Embodiment. In *Data through Movement: Designing Embodied Human-Data Interaction for Informal Learning*, Cafaro F., Roberts J., (Eds.), Synthesis Lectures on Visualization. Springer International Publishing, Cham, 2021, pp. 16–32. doi:10.1007/978-3-031-02610-2_3. 3, 7

[DH08] DALSGAARD P., HANSEN L. K.: Performing perception—staging aesthetics of interaction. *ACM Transactions on Computer-Human Interaction (TOCHI) 15*, 3 (2008), 1–33. 8, 9

[Dou01] DOURISH P.: *Where the action is*. MIT press Cambridge, 2001. 3, 7

[Dru22] DRUCKER J.: Diagrammatic Interface. *Interface Critique*, 4 (2022), 17–22. doi:10.11588/ic.2023.4.93405. 6

[EV21] ESCOBAR VARELA M.: *Theater as Data: Computational Journeys into Theater Research*. University of Michigan Press, 2021. doi:10.3998/mpub.11667458. 3

[F*12] FOSSATI G., ET AL.: Found footage filmmaking, film archiving and new participatory platforms. *Found Footage. Cinema Exposed. Amsterdam: Amsterdam University Press/EYE Film Institute Netherlands* (2012), 177–184. 1, 2

[FD16] FALK J. H., DIERKING L. D.: *The museum experience revisited*. Routledge, 2016. 3

[GJGR18] GARNER JR S. B., GARNER J., RENE: *Kinesthetic Spectatorship in the Theatre*. Springer, 2018. 7

[Gum04] GUMBRECHT H. U.: *Production of presence: What meaning cannot convey*. Stanford University Press, 2004. 1, 8

[Her] HERAS D. C.: *Cinema and Machine Vision*. Edinburgh University Press. arXiv:AvAMEQAAQBAJ. 2

[HMLB20] HONNIBAL M., MONTANI I., LANDEGHEM S. V., BOYD A.: spaCy: Industrial-strength natural language processing in Python, 2020. URL: https://spacy.io. 3

[HWS*16] HARTMANN T., WIRTH W., SCHRAMM H., KLIMMT C., VORDERER P., GYSBERS A., BÖCKING S., RAVAJA N., LAARNI J., SAARI T., GOUVEIA F., MARIA SACAU A.: The Spatial Presence Experience Scale (SPES). *Journal of Media Psychology 28*, 1 (Jan. 2016), 1–15. doi:10.1027/1864-1105/a000137. 2, 8

[Ken15] KENDERDINE S.: Embodiment, entanglement, and immersion in digital cultural heritage. *A new companion to digital humanities* (2015), 22–41. 2, 3

[KI22] KWON J., IEDEMA A.: Body and the Senses in Spatial Experience: The Implications of Kinesthetic and Synesthetic Perceptions for Design Thinking. *Frontiers in Psychology 13* (Apr. 2022), 864009. doi:10.3389/fpsyg.2022.864009. 7

[Kid18] KIDD J.: Immersive heritage encounters. *The Museum Review 3*, 1 (2018). 1, 2, 8

[KK15] KOCSIS A., KENDERDINE S.: I sho u: An innovative method for museum visitor evaluation. In *Digital Heritage and Culture: Strategy and Implementation*. World Scientific, 2015, pp. 245–259. 2

[KMH21] KENDERDINE S., MASON I., HIBBERD L.: Computational archives for experimental museology. In *International Conference on Emerging Technologies and the Digital Transformation of Museums and Heritage Sites* (2021), Springer, pp. 3–18. 1, 2, 3

[Kur04] KURIN R.: Safeguarding intangible cultural heritage in the 2003 UNESCO convention: a critical appraisal. *Museum International 56*, 1-2 (2004), 66–77. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1350-0775.2004.00459.x, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1350-0775.2004.00459.x, doi:https://doi.org/10.1111/j.1350-0775.2004.00459.x. 5

[LBLM13] LOPATOVSKA I., BIERLEIN I., LEMBER H., MEYER E.: Exploring requirements for online art collections. *Proceedings of the American Society for Information Science and Technology 50*, 1 (2013), 1–4. 7

[MM18] MUL G., MASSON E.: Data-Based Art, Algorithmic Poetry: Geert Mul in Conversation with Eef Masson. *TMG Journal for Media History 21*, 2 (Nov. 2018), 170–186. doi:10.18146/2213-7653.2018.375. 8

[MO21] MASSON E., OLESEN C. G.: Digital Access as Archival Reconstitution: Algorithmic Sampling, Visualization, and the Production of Meaning in Large Moving Image Repositories. URL: https://journals.openedition.org/signata/3011, doi:10.4000/signata.3011. 1, 2

[MOvNF20] MASSON E., OLESEN C. G., VAN NOORD N., FOSSATI G.: Exploring digitised moving image collections: The SEMIA project, visual analysis and the turn to abstraction. *DHQ: Digital Humanities Quarterly*, 4 (2020). 2, 3

[MSC] MAYER-SCHÖNBERGER V., CUKIER K.: *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, first mariner books edition ed. An Eamon Dolan Book. Mariner Books, Houghton Mifflin Harcourt. 3

[MSK*07] MCGINITY M., SHAW J., KUCHELMEISTER V., HARDJONO A., FAVERO D. D.: AVIE: A versatile multi-user stereo 360° interactive VR theatre. In *Proceedings of the 2007 Workshop on Emerging Displays Technologies: Images and beyond: The Future of Displays and Interacton* (New York, NY, USA, Aug. 2007), EDT '07, Association for Computing Machinery, pp. 2–es. doi:10.1145/1278240.1278242. 3

[OCH18] O'BRIEN H. L., CAIRNS P., HALL M.: A practical approach to measuring user engagement with the refined user engagement scale (ues) and new ues short form. *International Journal of Human-Computer Studies 112* (2018), 28–39. 2, 7

[OMZ*24] OIVA M., MUKHINA K., ZEMAITYTE V., KARJUS A., TAMM M., OHM T., METS M., CHÁVEZ HERAS D., CANET SOLA M., JUHT H. H., ET AL.: A framework for the analysis of historical newsreels. *Humanities and Social Sciences Communications 11*, 1 (2024), 1–15. 2

[RABK25] RATTINGER A., ALLIATA G., BENZI K., KENDERDINE S.: AI-Driven Metadata Extraction and Semantic Search for Audiovisual Archives. In *Archiving 2025* (2025). 1, 3

[Ree11] REEVES S.: *Designing Interfaces in Public Settings: Understanding the Role of the Spectator in Human-Computer Interaction*. Human-Computer Interaction Series. Springer London, London, 2011. doi:10.1007/978-0-85729-265-0. 8

[RKX*22] RADFORD A., KIM J. W., XU T., BROCKMAN G., MCLEAVEY C., SUTSKEVER I.: Robust speech recognition via large-scale weak supervision, 2022. URL: https://arxiv.org/abs/2212.04356, arXiv:2212.04356. 3

[RTS18] RTSARCHIVES: Le nouveau site rtsarchives, 2018. URL: https://www.rts.ch/archives/5919889-le-nouveau-site-rtsarchives.html. 1

[SBB*22] SLATER M., BANAKOU D., BEACCO A., GALLEGO J., MACIA-VARELA F., OLIVA R.: A Separate Reality: An Update on Place Illusion and Plausibility in Virtual Reality. *Frontiers in Virtual Reality 3* (June 2022). doi:10.3389/frvir.2022.914392. 8

[SF19] SAKER M., FRITH J.: From hybrid space to dislocated space: Mobile virtual reality and a third stage of mobile media theory. *New Media & Society 21*, 1 (Jan. 2019), 214–228. doi:10.1177/1461444818792407. 3

[Sha03] SHAW J.: *Future Cinema: The Cinematic Imaginary after Film*. MIT Press, 2003, ch. Introduction. 8

[SS23] SHEHADE M., STYLIANOU-LAMBERT T.: *Presence, Museums, and Immersive Technologies*. Taylor & Francis, Aug. 2023. doi:10.4324/9781003334316-2. 1, 2, 8

[SW97] SLATER M., WILBUR S.: A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments 6*, 6 (Dec. 1997), 603–616. doi:10.1162/pres.1997.6.6.603. 8

[Thy19] THYLSTRUP N. B.: *The politics of mass digitization*. MIT Press, 2019. 1

[vEKO*17] VAN EXEL T., KELLER P., OOMEN J., BRINKERINK M., SWAGEMAKERS W., KEIJSER L.: Images of the past. 7 years of images for the future, Mar. 2017. URL: https://publications.beeldengeluid.nl/pub/498. 1

[VTR92] VARELA F. J., THOMPSON E., ROSCH E.: *The Embodied Mind*. MIT Press, 1992. 8

[Wri17] WRIGHT R.: The future of television archives - digital preservation coalition, 2017. URL: https://www.dpconline.org/blog/wdpd/the-future-of-television-archives. 1