ESTABLISHMENT AND IMPLEMENTATION OF GUIDELINES FOR NARRATIVE AUDIO-BASED ROOM-SCALE VIRTUAL REALITY USING PRACTICE-BASED METHODS*

PRESENTED AT THE 2022 INTERNATIONAL CONFERENCE ON AUDIO FOR VIRTUAL AND AUGMENTED REALITY (AVAR 2022)

Constantin Popp
Department of Electrical Engineering
University of York
York, YO10 5DD, UK
constantin.popp@york.ac.uk

Damian T. Murphy
Department of Electrical Engineering
University of York
York, YO10 5DD, UK
damian.murphy@york.ac.uk

August 15, 2022

ABSTRACT

Room-scale Virtual Reality (VR) presents sound designers with new challenges to tell stories with audio in games with player-driven narratives. These challenges arise from the player moving in and interacting with the virtual environment. The paper performs a small scoping review of VR/non-VR games and associated literature to identify issues and solutions to the placement of speech-based audio using practice-based research methods. The review leads to the proposition of design guidelines and strategies for their implementation. The paper advocates that each instance of speech-based audio should be short, interactive, and complemented by non-speech audio. Furthermore, each instance's spatial, interactive, visual, aural, and narrative representation should be considered in combination. The paper also suggests that 3D-binaural audio informed by physics can aid storytelling and make virtual environments player-responsive.

Keywords Storytelling · Guidelines · 3D binaural audio · Sound design · Room-scale VR

1 Introduction

Room-scale Virtual Reality creates new opportunities and challenges for the sound designer to tell stories with audio. This medium gives the player control over the virtual environment's camera that correlates to their head movement and position in space [1]. Therefore, they are free to move and look around [2]. However, the medium's affordance in movement implies interactivity to the user which the storyteller needs to account for [3].

One solution to such interactive player-driven storytelling can be found in environmental narrative games ("walking simulators") [4]. They are often unpopulated, high-fidelity virtual environments and contain traces of human activity [5]. They tell stories through speech-based audio, written artefacts and features of the environments ("environmental storytelling") [5]. They task the player to explore the environment and interpret the presented artefacts, allowing them to pace and interact with the story [6]. However, creating audio that supports and drives this form of interactive storytelling in VR requires new working methodologies [7].

This paper proposes design guidelines and strategies for implementing an audio-focused approach to speech-based audio and environmental storytelling in environmental narrative games presented in room-scale VR. It uses practice-

*Preprint version. Original version to be found on: http://www.aes.org/e-lib/browse.cfm?elib=21841.

based research methods, mainly analysing player-generated screencasts of relevant VR/non-VR games, corresponding literature, and issues encountered in developing the authors' prototype experience. This paper uses the prototype to reflect on the feasibility of the proposed design guidelines. The prototype also serves as their example implementation.

This paper is organised as follows: Section 2 conducts a scoping review with its results culminating in design guidelines. Section 3 presents the implementation of these guidelines based on the authors' prototype room-scale VR experience and highlights found approaches. Section 4 discusses the advantages and disadvantages of these approaches. Section 5 concludes the paper.

2 Embedding Narrator(s)

The authors conducted a scoping review to establish design guidelines for embedding narrators in environmental narrative games. The information sources included two VR and five non-VR games, a literature review, personal experience from playing some of these games, and the authors' VR experiments. The analysed games are *Half-Life: Alyx* [8], *The Under Presents* [9], *Dear Esther* [10], *Tacoma* [11], *Firewatch* [12], *What Remains of Edith Finch* [13], and *Everybody's Gone to the Rapture* [14]. They were mainly analysed via screencasts found on *Youtube.com* [15]. Several screencasts of the same game have been viewed where necessary to identify variations in player interactions. A preliminary analysis focused on the spatial, interactive, and representative features of speech-based audio, revealing the following issues.

2.1 Diegetic or non-diegetic narrator

The narrator's voice can be aurally and visually embedded within (diegetic) or external (non-diegetic) to the game world, affecting the player's perception of the voice and its narrative content [16]. The type of diegesis can also directly influence the player's interpretation of the trustworthiness of the transmitted information [16]. *Half-Life: Alyx* shows that an experience can switch between diegetic and non-diegetic representations of the same narrator depending on the game's situation and context [17, see 03:48, 13:32, 18:15]. In this instance, the narrator's acoustic features match the type of presentation: in an avatar-like presentation, its voice is co-located at a specific point within the player's space, whereas in a headset-like presentation, the voice has a limited frequency range ($\approx$600-3500 Hz), is unresponsive to the player's head movement and has its own acoustic space. While the experience does not explicitly say this narrator's voice emanates from a headset, it evokes this association through context, gameplay, the voice's audio fidelity and its acoustic spatial behaviour [17, 03:48, 13:26, 19:41]. This example shows how acoustic features and the spatial behaviour of speech-based audio can perform functions of storytelling.

2.2 Interdependent factors

Sound designers need to consider several interdependent factors in deciding on the narrator's diegesis, such as player posture, the duration of the speech-based audio, "the desired look and feel, scenarios and project objectives" [18]. For example, if speech-based audio were embodied at a location in the game world using 3D audio, the player would have to remain at or near its location to hear it. This placement could become problematic for long texts and standing player postures, as the player could wish to sit to reduce fatigue [19]. If a diegetic voice were to also move in space, the player would have to follow it to hear it. If a voice's audio were non-diegetic, the player could continue performing their activities while listening to it passively for longer, as seen in [20]. This placement could ask the player to diverge their attention between their current activities and their listening ("split-attention effect"), potentially affecting information transfer negatively [21]. *Firewatch* addresses these issues by attaching diegetic voices to the player's avatar and a hand-held walkie-talkie, allowing them to have "long" conversations while exploring the environment.

2.3 Duration

The scoping review suggested that the player's attention may diverge for speech-based audio longer than approximately 25 s. The screencasts show that players can engage in random activities, as seen in [22, 04:00-05:07], [17, 04:13, 19:15] or [23, 10'58], or leave the narration prematurely [23, 17:05]. As described above, the specific problematic duration depends on the game's context/level of interactivity. Longer speech-based audio can be implemented by explicitly integrating or interweaving player interaction [23, 08:04], [17, 08:04-12:54], [24, 00:13-01:30].

2.4 Speech intelligibility

Sound designers must consider the complexities of achieving speech intelligibility in VR. Speech intelligibility depends on a variety of factors, such as "the properties of the interfering signal(s), the number of signals, the spatial configuration of the sources, and the acoustic environment" [25]. These factors may change unpredictably in player-driven/non-linear games [26, chap.2], requiring context-specific, dynamic audio mixing techniques [26, chap.8]. Dry voice-overs may seem beneficial as they could minimise signal fluctuations caused by distance attenuation, head-rotation, and masking effects caused by reverberation. However, they impair the player's ability to aurally localise them in the game world [23, 02:43]. Visual means could aid speech intelligibility, such as subtitles [24], synchronous lip movement of the voice's avatar [27], or the avatar's body language [28]. These methods can be observed in the screencasts [29, 06:58] and [17, 03:30]. However, they may not be available due to project constraints/budgets. Additionally, interactive controls could give players agency over the voice's playback, enabling repetition of misunderstood sections [29, 12:33].

2.5 Triggers

Sound designers need to consider the mechanism that allows players to initiate / trigger speech-based audio, affecting representation and interactivity. This mechanism can be implemented via visible objects the player needs to "touch" [24, 00:53], ensuring that the narration is player-driven. Invisible and inaudible triggers tied to a specific position that the player needs to cross can be used similarly [20, 02:20]. However, they may be difficult for the player to find/identify. Therefore, the authors recommend visible or audible diegetic landmarks, as they can attract/guide the player's attention, as seen in [22, 01:32].

2.6 Design guidelines

Following the discussion of the review, the authors propose the following generalised design guidelines for speech-based audio in environmental narrative games presented in room-scale VR:

1. The text of speech-based audio must be separable into a collection of easily understood speech recordings, each lasting less than $\approx$25 s, to aid player agency and reduce cognitive load. Longer texts should explicitly incorporate interactivity.
2. Diegetic audio/visual landmarks or objects should mark locations containing speech-based audio to the player to aid agency and guidance.
3. The player should be able to control/initiate the playback of the speech-based audio to aid their agency.
4. The speech-based audio should be intelligible enough without the support of facial or body animations or subtitles to reduce project complexity if necessary.
5. The effect of the spatial aspect of diegetic/non-diegetic speech should be considered regarding interactivity, the game's scenario, and sound localisation.
6. Several means of storytelling should be used concurrently to contextualise and complement speech-based audio following the definition of environmental narrative games.

3 Prototype Implementation

To validate the design guidelines, the authors implemented them in a prototype room-scale VR experience called *Planet Xerilia*. It is documented in a screencast highlighting this paper's issues [30].

3.1 Project-specific additional design guidelines

Project-specific design guidelines were added to minimise project scope and implement an audio-focused approach to environmental storytelling:

7. As an audio-focused implementation of design guideline 6, audio should provide narrative, emotive detail to support the spoken narration, reduce visual complexity, and shift the player's focus towards the audio.
8. Audio should be formatted exclusively in diegetic, 3D-binaural audio to enable room-scale player exploration. This approach specifies guideline 5.
9. Audio can be physics informed to add narrative detail, e.g., based on the motion of objects in space using a physics system and their spatialisation via 3D-binaural audio, addressing guideline 7.
10. The experience should run standalone on the Meta Quest 2 [31] to facilitate deployment.

3.2 Platform

The prototype is built using Unity [32], FMOD [33] and Google Resonance Audio [34] and contains 3D-binaural audio exclusively. This approach addresses design guidelines 8 and 10. The authors replaced Unity's audio engine with the audio middleware FMOD for ease of use, which runs as a plugin within Unity. Resonance Audio ran as a spatialisation plugin in FMOD and was chosen for its ease of use and efficiency [35]. The proposed approaches could be implemented on other platforms using alternative tools, but this paper limits the discussion to the authors' development platform.

3.3 Navigable Audio

A predominately object-based rendering approach [36] was chosen to enable room-scale exploration of audio (guideline 8) [37]. Monaural sound sources are transformed into sound-fields, taking player position and rotation, the sound source's directivity and interaction with the acoustic environment into account [36]. The sound-fields are then transformed and rotated according to the player's head rotation into a binaural representation for the player's left and right ears [36]. Distance attenuation curves imprint the player position's effect onto the spatialised sound. These curves describe the decrease in a sound's signal level over distance toward the player. They can be constrained via minimum and maximum attenuation settings and affect the impression of realism [38, chap.6]. These curves are critical in how large and loud an object sonically appears to the player [7, chap.5]. These curves can be supplemented with distance-based equalisation to imitate or exaggerate changes in a sound's spectrum over distance, such as to model the high-frequency attenuation of a sound source over time/distance due to air absorption [38, chap.16]. Sound-field rendering [36] was used for diffuse, surrounding ambiences, which were equally attenuated by distance attenuation curves.

3.4 General narrative approach

Planet Xerilia tells its story through eight acoustic environments and a collection of short speech recordings, addressing design guidelines 1 and 7. The story focuses on two scientists trying to save their civilisation before and after a devastating meteor impact. The writer, Helene Vitting, framed the text as brief diary entries made by the story's narrators. This approach facilitated the subdivision of the text into short speech recordings that can be grouped by topic and distributed spatially and diegetically. Each topic is associated with an area made up of (interactive) objects ("props"). Several areas make up an environment, and the environment mirrors the narrator's descriptions and emotional state in a symbolic way via audio-driven environmental storytelling.

3.5 Embedded narrators: monoliths

The narrators are embedded audio-visually in the game world through monoliths (object in the bottom right in Figure 1), complying to design guideline 2/3/5. They feature distinct lighting and colour, making them stand out from their surroundings. They house a single, audio-only speech recording from one of the narrators that the player can trigger at will. An activation and termination sound frames the playback to sonically affirm the player's interaction and the recordings' end. The screencast shows an example interaction at 0:12.

3.5.1 Storytelling through audio effects processing and sound materials

The monolith's physical appearance is augmented sonically through exaggerated audio processing and the connotation of sound materials. This strategy can be seen as an example of audio providing narrative detail of a prop, implementing design guideline 7. The monolith emits an oscillating hum, reminiscent of a humming loudspeaker. The hum is based on two slightly saturated sine waves tuned to 52 and 156 Hz. The sine waves occupy minimal spectral space to minimise them masking surrounding sounds or the narrators. The activation/termination sound is based on a short noise burst and a blip-like square wave. The latter is sourced from a pushed button of a humidity meter. The speech recordings are first band-limited to 370-4000 Hz, then tube-saturated, resampled to 7060 kHz, quantised to 6 bits, ring-modulated (20% depth at 4500 Hz), compressed, and again band-limited to 500-14000 Hz. This audio processing manifests itself through clearly audible artefacts and thus shifts awareness to the recording medium, aiming to evoke the impression of an old, digital recording device.

3.5.2 Intelligibility

To maximise the intelligibility of the speech recordings (design guideline 4), the authors implemented a combination of techniques capitalising on 3D audio as recommended in [39]. This approach is audible in the screencast at 0:45. The recordings remain immobile in space, are distanced from loud environmental sounds wherever possible, and radiate evenly in all directions. The shape of the distance attenuation curve associated with the speech source remains

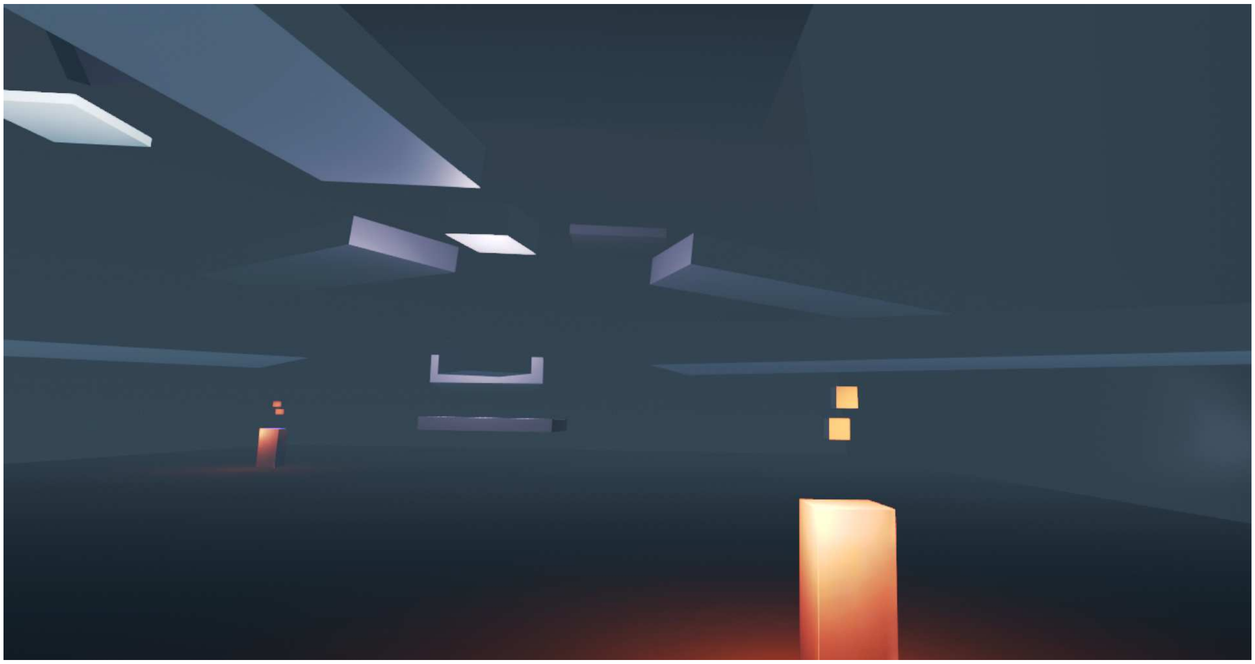

Figure 1: Example screenshot depicting two monoliths (front bottom right, back left) and two three-armed rotors (centre). The two rotors represent this area's non-interactive main props.

constant between 0 and 2 m, falls quickly after 2 m and becomes inaudible by 7 m. This shape removes distance attenuation below 2 m. It allows the player to listen to the speech recording at a comfortable distance, reduces signal level fluctuations caused by player movement, and lessens the masking effects of reverberation. The quick fall-off makes distance attenuation audible and limits aural overlap with other speech sources. A further dip in the frequency spectrum of an environment's louder sounds around 1.5-3 kHz, as recommended in [26, chap.4], clears spectral space for the speech's consonants. The reverb's level and decay time are adjusted to balance speech intelligibility with the space's dimensions and narrative connotation.

3.5.3 Interactivity and agency

Several strategies aim to enable the player's agency regarding the monoliths, addressing design guidelines 1-3. The speech recordings are between 4 and 22 s. They are actively triggered by the player via a point-and-click action or a collision with the player's virtual hands, giving them control over if, when, and how often a recording is heard. The recording's short audible range allows the player to withdraw quickly, providing an intuitive way of "stopping" the playback. The monolith's distinct visuals form attraction points so that the player can consciously seek or avoid them. Their hum is kept subtle to facilitate avoidance; otherwise, the player could feel forced to trigger the speech playback to stop the noise.

3.6 Audio-based environmental storytelling

Audio-driven environmental storytelling is implemented in *Planet Xerilia* through the connotation of sounds, applied audio processing, movement in space and reverberation using features of 3D-binaural audio and the game engine's physics system, addressing design guidelines 7-9.

3.6.1 Distance-attenuation curves and audio processing

The environment around the monoliths harnesses distance attenuation curves, audio processing, and the sounds' motion in space to allude to the narrator's feelings symbolically. For example, in the screencast at 2:08 and within the context of Figure 1, the monolith's narrator implicitly expresses the stress caused by guiding their people to survive while their situation deteriorates. Two vertically stacked oversized rotors, which slowly rotate in opposite directions, form this area's main props. Each of the three blades per rotor features engine-like sounds, and is based on processed versions of a close-up recording of a dehumidifier's compressor. The processing aims to enhance the source recordings' loud, aggressive, oscillating, and electronic character for emotional impact. Combined with the rotor's motion and customised distance attenuation curve, the sounds form wave-like gestures that repeatedly engulf and encroach the player from their left and right. This aural behaviour aims to create the sensation of tension and dissonance.

3.6.2 The physics system creates complex, narrative sound gestures

Planet Xerilia harnesses Unity's physics system and 3D-binaural audio to create complex, interactive sound gestures for environmental storytelling purposes. For example, in the screencast at 3'20, large white cuboids ("crystals") move chaotically toward the player from the sky. Invisible forces or the player push the crystals to move in random directions. Invisible springs constrain and force the crystals to bounce between the player and the environment. Each force/push action is mapped to a muffled, rocket-like sound with its pitch and frequency response affected by the Doppler effect, the associated velocity, and player distance. Collision sounds of rocks, gravel, and a resampled, saturated synthesised kickdrum sonify the crystal's collision. The crystal's impact velocity is mapped to parameters of the gravel and rock sounds. Their sounds suggest the materiality of the floor and the crystals. The authors exaggerated the strength of the Doppler effect (by 500%) and the impact collisions to over-accentuate the crystal's velocity and force, making it appear larger than life and (potentially) threatening.

3.6.3 Reverberation

The combination of mapping impact velocities to sound and physics-informed reverberation can also be used for environmental storytelling. Resonance Audio's reverberation effect aims to reproduce the reverb tail of a simulated environment "in a manner consistent with a corresponding real-world environment" [40] by taking a room's surface materials and overall dimension into account. The resultant reverb tail can be adjusted via correction factors (brightness, size, gain, reflectivity, time) [41]. In the screencast at 0:31, impact collisions from water droplets continuously excite the area's reverberation. The surface materials of the room's reverb simulation are set to brick-like on all sides and created in combination with the room size (30 by 31.375 by 5 m) and a reverb time-scale factor of 0.86 an estimated decay time of 7.45 s at 500 Hz (T30 measurement). The authors matched the surface materials implied by the water droplets' collision with the surface materials used in the reverberation to avoid causing aural contradictions. The water droplets and the reverb tails together aim to evoke the impression of a cavernous, stone-like, wet space.

4 Discussion

This paper has presented some of the sound design strategies used to implement the proposed design guidelines for audio-driven storytelling in *Planet Xerilia*. The implementation revealed the advantages and disadvantages of these strategies.

4.1 Physics informed sounds tell stories

Game data mapped to sound, such as an object's sounding motion, can support narrative, responsive, audio-driven environments. This approach can also be considered a form of (interactive) sonification [42]. For example, depending on the settings, the physics system approximates the object's motion in line with the player's expectations based on their previous experiences ("intuitive physics") [43]. Sound can support this expectation through plausible aural cues which allow the player to deduce the object's physical properties [44], such as mass, material, speed, and distance, and thus predict possible interaction outcomes [43], as shown in the screencast at 01:31-01:48. These aural cues provide narrative detail to the objects seen. However, they should conform to the player's expectations [45] that are shaped by (media) experiences [46] and can be tuned for emotional impact or intelligibility, as seen in this paper's exaggerated application of the Doppler effect or customisation of distance attenuation curves. Garner describes this strategy as skewing aural realism towards something that is emotionally more evocative [47, chaps.1.4.4].

4.2 Agency enables stories

Enabling interactivity through physics-informed interaction, such as collisions or binaural audio, allows the player to directly experience aspects of a story through performing motions and (inter-) actions that cause objects and environments to respond, as seen in the screencast at 03:30-03:35. Their response in sound and motion also acknowledges the player's presence in the environment, reinforcing believability [48]. Fostering player agency and environmental storytelling makes the designer a story enabler: to enable the player "to engage with and perceive their own unique version of the story" [49].

4.3 Challenges

The physics-informed, diegetic, audio-focused approach to storytelling also presents challenges. For example, mapping data such as collision velocities to sound requires a large pool of sound files, as each velocity requires material-dependent mappings [50]. *Planet Xerilia* groups these velocities into 4 to 5 levels and blends transitions between them to reduce complexity. It also limits velocity-to-sound mapping to the story's essential objects. Procedural audio synthesising sound effects in real-time based on algorithms [51] may solve these limitations.

Finding matching mappings between the physical properties of objects heard and seen proved challenging. The sound designers needed time to identify and address a plethora of underlying issues that potentially could cause contradictions perceived between, for instance, an object's aural and visual distance towards the player or the motion/energy implied by sounds and their visual representation.

4.4 Aural limitations of diegetic narration

The combination of diegetic implementation of speech without a humanoid avatar, lack of subtitles, and Resonance Audio's architecture restrict the use of audio processing and sound placement to ensure intelligibility. The lack of a user-accessible audio mix bus system in Resonance Audio [33, chap.15] requires sound designers to rely on alternative methods to ensure speech intelligibility. Subtitles or an avatar's body language could have eased the problem but were unavailable due to project constraints. Instead, the authors used audio-only solutions, such as positioning the speech-based audio away from loud sound sources, freeing spectral space via equalisation, and reducing the signal level of reverberation. The reverb's lack of directionality [52] also required mitigating strategies to avoid obscuring sound localisation.

5 Summary

The paper has proposed design guidelines and strategies for their implementation for embedding speech-based audio and environmental storytelling in room-scale VR narrative environmental games. These guidelines were deduced from a scoping review, adapted, and tested via the development of an audio-focused implementation of a room-scale VR environmental narrative game. The guidelines proved helpful in guiding the development process. In addition, the authors showed that a physics-informed, 3D-binaural audio approach to environmental storytelling could be beneficial to player involvement but adds additional challenges.

Future work would investigate the narrative effectiveness of the design guideline's implementation in *Planet Xerilia* in a small user study.

6 ACKNOWLEDGMENT

This work is supported by the UK Arts and Humanities Research Council (AHRC) XR Stories Creative Industries Cluster project, grant no. AH/S002839/1.

Firelight Technologies Pty Ltd and Unity Technologies kindly provided non-commercial licences for research purposes as part of this research project.

References

[1] R. Shewaga, A. Uribe-Quevedo, B. Kapralos, and F. Alam, "A comparison of seated and Room-Scale virtual reality in a serious game for epidural preparation," *IEEE Transactions on Emerging Topics in Computing*, vol. 8, pp. 218–232, Jan. 2020.

[2] R. Aylett and S. Louchart, "Towards a narrative theory of virtual reality," *Virtual Real.*, vol. 7, pp. 2–9, Dec. 2003.

[3] R. Pausch, J. Snoddy, R. Taylor, S. Watson, and E. Haseltine, "Disney's aladdin: first steps toward storytelling in virtual reality," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, SIGGRAPH '96, (New York, NY, USA), pp. 193–203, Association for Computing Machinery, Aug. 1996.

[4] F. Zimmermann and C. Huberts, "From walking simulator to ambience action game," *Press Start*, vol. 5, pp. 29–50, Aug. 2019.

[5] J. Habgood, D. Moore, S. Alapont, C. Ferguson, and H. van Oostendorp, "The REVEAL educational environmental narrative framework for PlayStation VR," in *Proceedings of the 12th European conference on games based learning* (M. Ciussi, ed.), pp. 175–183, Academic Conferences and Publishing International Limited, Sept. 2018.

[6] H. Smith and M. Worch, "What happened here? – environmental storytelling." Game Developers Conference, 2010.

[7] S. Schütze and A. Irwin-Schütze, *New Realities in Audio: A Practical Guide for VR, AR, MR and 360 Video*. Boca Raton: CRC Press, 1st edition ed., Feb. 2018.

[8] Valve Corporation, "Half-Life: Alyx." https://www.half-life.com/en/alyx/, Mar. 2020. Accessed: 2022-1-21.

[9] Tender Claws, "The under presents." https://tenderclaws.com/theunderpresents, Nov. 2019. Accessed: 2022-4-25.

[10] Sumo Digital Ltd, "Dear esther." https://www.thechineseroom.co.uk/games/dear-esther, 2012. Accessed: 2022-4-13.

[11] Fullbright, "Fullbright presents tacoma." https://tacoma.game/, Feb. 2017. Accessed: 2022-4-4.

[12] Campo Santo, "Firewatch." https://www.firewatchgame.com/, 2016. Accessed: 2022-3-11.

[13] Giant Sparrow, "What remains of edith finch." http://edithfinch.com/, Apr. 2017. Accessed: 2022-4-26.

[14] Sumo Digital Ltd, "Everybody's gone to the rapture." https://www.thechineseroom.co.uk/games/everybodys-gone-to-the-rapture, Nov. 2015. Accessed: 2022-4-26.

[15] Google LLC, "YouTube." https://www.youtube.com/, Feb. 2005. Accessed: 2022-4-22.

[16] A. Çamci, "Exploring the effects of diegetic and non-diegetic audiovisual cues on decision-making in virtual reality," in *Proceedings of the 16th Sound & Music Computing Conference* (Isabel Barbancho, Lorenzo J. Tardón, Alberto Peinado, Ana M. Barbancho, ed.), pp. 195–201, 2019.

[17] MKIceAndFire, "HALF LIFE ALYX gameplay walkthrough part 1." https://www.youtube.com/watch?v=WnzjqoSA8dk, Mar. 2020. Accessed: 2022-3-11.

[18] C. Summers and M. Jesse, "Creating immersive and aesthetic auditory spaces in virtual reality," in *2017 IEEE 3rd VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*, pp. 1–6, Mar. 2017.

[19] M. R. Chester, M. J. Rys, and S. A. Konz, "Leg swelling, comfort and fatigue when sitting, standing, and sit/standing," *Int. J. Ind. Ergon.*, vol. 29, pp. 289–296, May 2002.

[20] Father, "Dear esther: Landmark edition - full game walkthrough gameplay & ending (no commentary playthrough)." https://www.youtube.com/watch?v=AZcCUJz_47c, Feb. 2017. Accessed: 2022-4-22.

[21] S. Kalyuga and J. L. Plass, "Evaluating and managing cognitive load in games," in *Handbook of Research on Effective Electronic Gaming in Education*, pp. 719–737, IGI Global, 2009.

[22] P. B. Horror Gaming, "Everybody's gone to the rapture: Full playthrough." https://www.youtube.com/watch?v=2vifO3_gjiM, Jan. 2016. Accessed: 2022-4-22.

[23] MrCoolPotato, "The under presents oculus quest 2 GAMEPLAY (no commentary)." https://www.youtube.com/watch?v=QPRoycuvefg, Oct. 2020. Accessed: 2022-4-22.

[24] Father, "What remains of edith finch - full game walkthrough gameplay & ending (no commentary longplay)." https://www.youtube.com/watch?v=q-fj8PhdWeI, Apr. 2017. Accessed: 2022-4-22.

[25] A. W. Bronkhorst, "The cocktail party phenomenon: A review of research on speech intelligibility in Multiple-Talker conditions," *Acta Acustica united with Acustica*, vol. 86, no. 1, pp. 117–128, 2000.

[26] G. Zdanowicz and S. Bambrick, *The Game Audio Strategy Guide: A Practical Course*. Routledge, Dec. 2019.

[27] M. Bourguignon, M. Baart, E. C. Kapnoula, and N. Molinaro, "Lip-Reading enables the brain to synthesize auditory features of unknown silent speech," *J. Neurosci.*, vol. 40, pp. 1053–1065, Jan. 2020.

[28] R. B. Church, S. Ayman-Nolley, and S. Mahootian, "The role of gesture in bilingual education: Does gesture enhance learning?," *Int. J. Biling. Educ. Biling.*, vol. 7, pp. 303–319, Aug. 2004.

[29] GrumpyFox, "TACOMA longplay." https://www.youtube.com/watch?v=5nM4Pk8uW5E, Mar. 2020. Accessed: 2022-4-4.

[30] C. Popp and D. T. Murphy, "Video examples from: Establishment and implementation of guidelines for narrative audio-based room-scale virtual reality using practice-based methods," Apr. 2022.

[31] Facebook Technologies, LLC., "Oculus device specifications." https://developer.oculus.com/resources/oculus-device-specs/, 2021. Accessed: 2021-11-29.

[32] Unity Technologies, "Unity - manual: Unity user manual 2020.3 (LTS)." https://docs.unity3d.com/Manual/, Dec. 2021. Accessed: 2021-1-12.

[33] Firelight Technologies Pty Ltd, "FMOD studio user manual 2.02." https://www.fmod.com/resources/documentation-studio?version=2.02, Dec. 2021. Accessed: 2022-1-24.

[34] M. Gorzel, A. Allen, I. Kelly, J. Kammerl, A. Gungormusler, H. Yeh, and F. Boland, "Efficient encoding and decoding of binaural sound with resonance audio," in *2019 AES International Conference on Immersive and Interactive Audio*, Audio Engineering Society, 2019.

[35] R. Gould, "Let's test: 3D audio spatialization plugins." https://designingsound.org/2018/03/29/lets-test-3d-audio-spatialization-plugins/, Mar. 2018. Accessed: 2021-1-12.

[36] D. R. Méndez, C. Armstrong, J. Stubbs, M. Stiles, and G. Kearney, "Practical recording techniques for music production with Six-Degrees of freedom virtual reality," *Journal of the Audio Engineering Society*, p. 5, 2018.

[37] C. Popp and D. T. Murphy, "Creating audio Object-Focused acoustic environments for Room-Scale virtual reality," *Applied Sciences*, vol. 12, p. 7306, July 2022.

[38] C. Robinson, *Game Audio with FMOD and Unity*. Routledge, Mar. 2019.

[39] Facebook Technologies, LLC, "Mixing for virtual reality | oculus developers." https://developer.oculus.com/resources/audio-intro-mixing/, 2017. Accessed: 2021-1-12.

[40] I. Kelly, M. Gorzel, and A. Güngörmüsler, "Efficient externalized audio reverberation with smooth transitioning," *Technical Disclosure Commons*, 2017.

[41] Google LLC, "Developer guide for resonance audio for unity." https://resonance-audio.github.io/resonance-audio/develop/unity/developer-guide, 2018. Accessed: 2021-12-7.

[42] M. Grimshaw, "Sound and immersion in the First-Person shooter," in *Proceedings of the 3rd Audio Mostly Conference*, 2008.

[43] T. D. Ullman, E. Spelke, P. Battaglia, and J. B. Tenenbaum, "Mind games: Game engines as an architecture for intuitive physics," *Trends Cogn. Sci.*, vol. 21, pp. 649–665, Sept. 2017.

[44] W. W. Gaver, "How do we hear in the world? explorations in ecological acoustics," *Ecol. Psychol.*, vol. 5, pp. 285–313, Dec. 1993.

[45] M. Droumeva, "An acoustic communication framework for game sound," in *Game Sound Technology and Player Interaction* (M. Grimshaw, ed.), pp. 131–152, IGI Global, 2010.

[46] C. F. Underriner, "The Sound-Poetry of the instability of reality: The audio reality effect and mimesis," *Organised Sound*, vol. 22, pp. 20–31, Apr. 2017.

[47] T. A. Garner, "Cinematic sound design for players," in *Foundations in Sound Design for Interactive Media* (M. Filimowicz, ed.), pp. 1–23, Routledge, June 2019.

[48] J. Mütterlein, "The three pillars of virtual reality? investigating the roles of immersion, presence, and interactivity," in *Proceedings of the 51st Hawaii International Conference on System Sciences* (T. X. Bui, ed.), Jan. 2018.

[49] D. Cotting, "Storytelling vs. story enabling: Crafting experiences in the new medium of virtual reality." https://medium.com/@DanCotting/storytelling-vs-story-enabling-crafting-experiences-in-the-new-medium-of-virtual-reality-8b982359d3b8, Nov. 2016. Accessed: 2022-3-8.

[50] J. Lietzén, J. Miettinen, M. Kylliäinen, and S. Pajunen, "Impact force excitation generated by an ISO tapping machine on wooden floors," *Appl. Acoust.*, vol. 175, p. 107821, Apr. 2021.

[51] M. Yee-King and I. Dall'Avanzi, "Procedural audio in video games," in *Encyclopedia of Computer Graphics and Games* (N. Lee, ed.), pp. 1–5, Cham: Springer International Publishing, 2018.

[52] AtomicJoe, "Reverb always at full volume at any distance · issue #35 · resonance-audio/resonance-audio-unity-sdk." https://github.com/resonance-audio/resonance-audio-unity-sdk/issues/35, May 2018. Accessed: 2021-11-30.