



## JOURNAL OF INTERACTIVE NARRATIVE

# The Role of Voice in Virtual Reality Interactive Narratives

**Jonathan Barbara**, Saint Martin's Institute of Higher Education, Malta,

**Mads Haahr**, School of Computer Science and Statistics, Trinity College Dublin, Ireland

---

### Abstract

Providing enough information to allow the VR player to self-identify is an important factor in their immersion into a virtual world. The sensory information being provided to the player's eyes and ears through the headset must support the suspension of disbelief and telepresence into the virtual world. Given sound's easily realized potential for diegetic ambiguity and its influence on presence as immersion, the aim of this paper is to explore the role of disembodied voice in VR interactive narratives and its effect on presence through self-identification. We start by analyzing the shift of focalization in the film *Saving Private Ryan*, the trans-diegetic narration in the game *StarCraft II*, and the address of the player in the game *Call of Duty*. We then contribute with close readings of two VR productions, *The Last Goodbye* and *The Book of Distance*, in terms of the use of second-person voice and interactivity as well as an analysis of the resulting levels of self-identification. A further contribution is a Twine interactive narrative that demonstrates some of the key concepts discussed in the paper. We conclude that future empirical work should explore the impact of second-person voice and interaction on the resultant self-identification and immersion.

**Keywords:** virtual reality, second-person, immersion, interactive narrative

---

While visual discrimination between non-diegetic user interfaces and diegetic elements of a virtual world seen through a Virtual Reality (VR) headset is aided by depth perception, the aural dimension presents a challenge for the player to assess whether a sound is diegetic or not. Building on Bernstein's [1] analysis of audio in terms of what information sounds provide the player, Ekman classifies sounds in relationship to the diegesis of their referent as 'the thing being told by the sound' [2], while empirical research found 'participants directly connecting the sounds with an object/event' [3]. Ekman presents two approaches to such an assessment: (i) whether its apparent source is itself diegetic; or (ii) whether the non-player inhabitants of the virtual world react to the sound. Both approaches are dependent on a non-guaranteed relationship between the sound and its potential source's visual representation or lack thereof. The ambiguity is compounded by the layers of information carried by audio signals: frequency, timbre, and semantics [4] and the exaggerated use of audio to compensate for visual shortcomings [5] such as the use of the 'menacing zombie drone

Journal of Interactive Narrative publications are covered by a [CC BY-NC-ND 2.5](https://creativecommons.org/licenses/by-nc-nd/2.5/) Generic License. This means that you are free to share these works as long as you give appropriate credit, do not use them for commercial purposes, and do not create derivative works. Please note that images, videos, audio files, and interactive works featured in articles are owned by their respective copyright holders. They are not included under the Creative Commons license. Accessing or using these works does not grant you any rights to them, and you cannot assume any ownership or rights to commercially use or modify these works. The owners retain all rights to their content. For more details on the copyright rules applicable to authors contributing to the Journal of Interactive Narrative, [you can access the full copyright statement here](#).

∅OPEN ACCESS

for brains' [6] in *Zombies, Run!*

These multiple levels of aurally delivered meanings offer a challenge of understanding to the listener as they acclimatize and give meaning to the virtual space. As sound bounces off hard surfaces irrespective of their smoothness, which is a requisite for image reflection, we often hear objects before they become visible, causing players to speculate and build expectations as to what is to come [6]. This behavior is re-enforced in films where, when an off-screen sound calls for attention, the shot cuts to reveal the source of the sound [8]. Should these expectations fail to manifest themselves, such as when a voice is heard by the player but no speaker is seen or no reactions to it come from diegetic characters, the player's understanding of the virtual space is challenged, undermining narrative engagement [9]. This is akin to the disruption of smooth filmic reception – what Pólya calls “perceptual reflexivity” – such as undermining the 3D illusion, which “leads to a state of quasi-awareness in the viewer” [10]. Following Ekman [2], this disembodied voice is thus perceived to be non-diegetic, which negatively influences the VR player's immersion, as they are reminded of their non-diegetic existence. Embodied speech, on the other hand, supports the player's (tele)presence, especially through the use of the second-person ‘you’ which helps create the feeling of being addressed through aesthetic-reflexive involvement [11] and thus present in the virtual world, which in turn increases immersion [12–14].

Given sound's easily realized potential for diegetic ambiguity and its influence on presence as immersion [15], the aim of this paper is to explore the role of disembodied voice in VR interactive narratives and its effect on presence through self-identification. To this end, the first section considers voice as disembodied speech by asking ‘Who speaks?’ To demonstrate the effect of the ambiguity of disembodied voice, we use three popular titles from the cinematic and video game genres that present these concepts clearly and are closely tied to the Interactive Narrative form. *Saving Private Ryan* presents a very clear example of identity shift as a result of ambiguity resulting from a misdirected connection between voice and speaker, the *StarCraft II* campaign ties its gameplay tightly to a progressive narrative using a trans-diegetic voice to guide the player, while *Call of Duty* uses diegetic voices that use the second-person to address the player providing a compelling narrative that is based on non-fiction.

In the second section, we then shift to the VR player's voice: ‘Who am I that speaks?’ and interaction: ‘Who am I that acts?’ We explore the use of second person voice and self-identification in VR by considering the three identities a narratee may take in traditional narratives: narrator, protagonist, or non-protagonist third party. These offer a lens through which a close reading of two VR interactive narratives (*The Last Goodbye* and *The Book of Distance*) explores different implementations of voice and perspective, including a consideration of the elicited self-identification to compare its effect between the different uses.

In the final section, the concepts touched upon in the first part of the study are considered in the context of these two experiences and an accompanying IDN that embodies the outcome of the discussion is described.

## 2. Background work: Who speaks in films and games?

In his essay on Narrative Discourse, Genette [16] uses Vendryes's definition of voice as used in his *Traité d'accentuation grecque* (Treaty of Greek Accentuation): "the mode of action of the verb considered for its relation to the subject" ([17] as cited in [16]). Thus, voice determines 'who speaks?' where the subject of the action includes the narrator and the narratee(s). In the context of fiction, the narrator is a fictional character invented by the author, thereby separating the act of narration from the author's writing and indeed from the author being in the narrative [16]. This frees the author from narrating in the diegetic first-person (homodiegetic narration), and gives the option to narrate in the non-diegetic third-person (heterodiegetic narration) [16], which in turn results in new options for narrating the protagonist's experience of the plot: as the protagonist (autodiegetic narrator), as a non-protagonist character diegetic to the story (homodiegetic narrator), or as a non-diegetic, unrepresented, character (heterodiegetic narrator) (see Figure 1 for an overview). The author may decide to leave the relationship between the narrator and the protagonist ambiguous and even let it remain unresolved until the end of the narrative, if at all.

Such ambiguity is not restricted to the narrative of disembodied literary text, i.e., text without a visual representation of its source accompanying it. In the feature film *Saving Private Ryan* [18], the opening scene presents us with a silent old man visiting a military cemetery with his family in tow, falling to his knees in front of one of the white crosses. As the shot zooms in onto his blue eyes, we are taken back to June 1944, leading us as viewers to believe that we are visiting his memory and thereby marking him as the narrator. Throughout the film we follow the story of blue-eyed Captain John Miller leading his company on a search and rescue mission to find Private James Ryan. While the story is acted out using cinema's photographic medium of mimesis, we also feel that the story is being narrated to us by a character in the story. Finding no clues in the silence of the elderly man, we are given the impression that the thinking blue-eyed narrator is the protagonist Miller (autodiegetic narrator), as only he could have known the details of his long search for Ryan. While we are given subtle clues as to the incorrectness of this assumption throughout the film, it is finally negated in the last scene of the story with the stillness of the fatally wounded Miller's trembling hand and Ryan's blue-eyed visage cross-fading into the face of the old man at the cemetery as we return to the opening scene. How Ryan could have known the details of Miller's journey to him through France, from Omaha Beach to Ramelle, is a question left unanswered.

NARRATIVE VOICE	DIEGESIS	PERSON	CHARACTER	NARRATIVE PERSPECTIVE
homodiegetic	diegetic	1st Person	non-protagonist	external focalization
autodiegetic	diegetic	1st Person	protagonist	internal focalization
heterodiegetic	non-diegetic	1st Person	non-represented	non-focalized

Figure 1: Overview of Narrative Voice and usually related Narrative Perspective

## 2.1 Narrative Voice and Perspective

Aare [19] clearly distinguishes between the narrative voice (who speaks) and the narrative perspective (who knows), while making a slight reference to the camera's perspective (who sees). The latter manifests itself as 'ocularization' and 'auricularization' of narrative [20, 21] shifting attention onto the narratee's perspective, which we address in section 3.1.

As for narrative perspective, Genette categorizes it into three forms: (i) in Internal focalization, knowledge is bound to one character, usually the protagonist-narrator; (ii) in External focalization, knowledge about the protagonist is limited to a third-person's perspective, and (iii) in Non-focalization, the knowledge would not be limited to the perspective of a single character. This could suggest the existence of a non-diegetic implied author [16] or an impersonal voice of the narrative, linked to no one character in particular [22] and comparable to a third-person heterodiegetic narration (see Figure 2).

The main story we follow in *Saving Private Ryan* is initially understood as being narrated by the old man whose identity is based on the camera perspective ('ocularization' and 'auricularization') of a heterodiegetic monstator<sup>1</sup> accompanying the protagonist Miller throughout his journey. As we follow Miller's story, beginning from an assumed internal focalization, we discover that, like the soldiers under his command, there are many questions about the protagonist that are left unanswered, shifting us slowly to an external focalization. This is confirmed when we learn of Miller's hometown and civilian occupation simultaneously with his squad as he opens up to defuse a high-tension situation towards the end of the journey. This shift of focalization helps smoothen the transition from autodiegetic to homodiegetic narration as we identify Ryan, rather than Miller, as being the narrating old man.

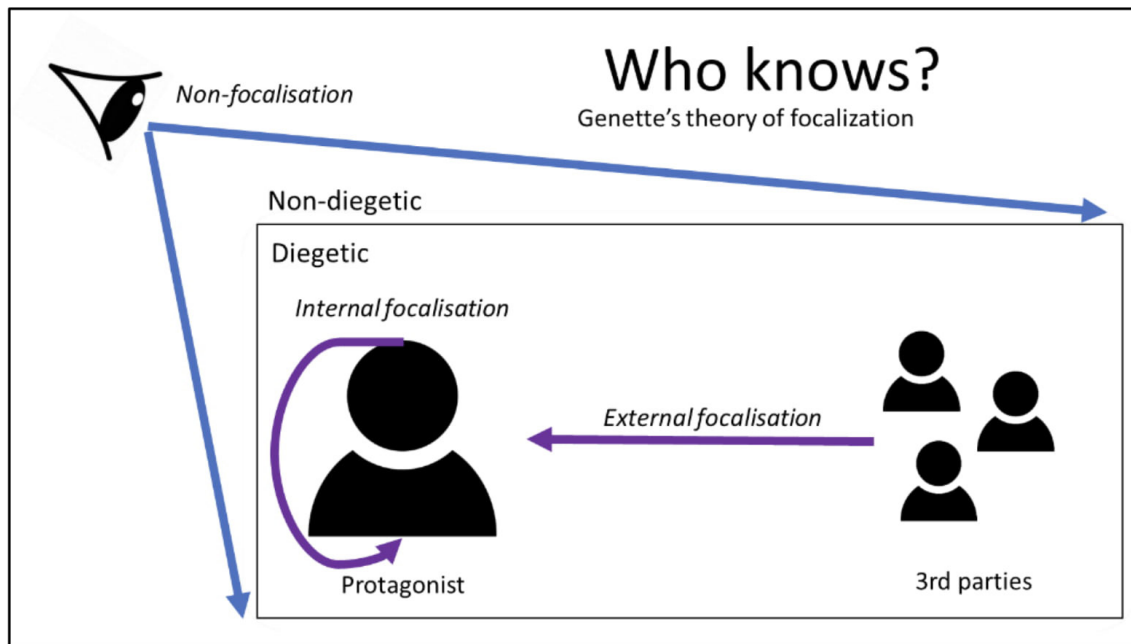


Figure 2: Genette's theory of focalization [16]

An alternative understanding is provided by Nielsen [17]. When the first-person narrator lets on more than they should know, Nielsen suggests that rather than being represented as one particular character, the narrator would represent an impersonal voice of the narrative, linked to no one character in particular. This results in a situation that is comparable to third-person heterodiegetic narration (as Ryan was non-participant in Miller's story until their meeting in Ramelle) and thus provides non-focalization (access to certain details that would have been beyond Ryan's knowledge).

The old man's silent disposition in the opening scene is continued throughout the recounting of the 1944 D-Day landings and subsequent events: the film uses no voiceover that could justify the above understanding in terms of diegesis rather than the mimetic form of cinema. However, had there been a voice-over, the use of first-person or third-person voice could have disclosed the elderly man's identity early on and possibly diminished the power of the final reveal.

## 2.1 Narrative Voice and Perspective

Narration has been an important counterpart to mimesis in theatre [23] when, in prologues and epilogues, the presence of narrators on stage present a prolepsis (what is yet to be mimetically shown) and, in some non-Western drama, such as Japanese plays, the importance of the narrator is as much as that of the main characters in terms of lines and time on stage. Indeed, in *Saving Private Ryan*, the silent opening

scene at the military cemetery serves as a prologue for the presentation of events happening in 1944, while the closing scene rounds off with an epilogue, showing old Ryan's grief and seeking acknowledgement from his wife as to whether he had earned the sacrifice of Miller and his squad for his rescue.

In games, prologues find their equivalent in tutorial missions. Due to their interactive nature, games most often use tutorials to guide the player-narratees in their participation of the narration through their gameplay. Such players are addressed using voices whose diegesis depends on the game's genre. Non-diegetic narration serves to direct the mimetic gameplay, giving motivation and justification for the player's in-game actions, similar to 'generative narrators,' whose narration instantiates action [23].

An example of this is the in-game tutorial for Starcraft II [24], where an android with a female voice welcomes the player, who is addressed as a 'recruit', to the 'Dominion Future Commander Training Simulation Module' in order to learn how to use the interface and issue orders to the available forces (see Figure 3). The visuals present a controlled environment with gated progress, and visual indicators that are diegetic to the game are overlaid by non-diegetic arrows and highlights to indicate relevant sections and directions. The android, represented by a framed animated profile video feed forming part of the HUD, vocally guides the player through the easy challenges in the tutorial by making reference to non-diegetic aspects of the game such as the screen, keyboard and mouse. Any out of place actions executed during the tutorial, like attacking a unit out of sequence, are handled by red non-diegetic error text messages displayed underneath the android's profile without interrupting its scripted speech. Completing actions quickly does however terminate the explanation using a fade out.

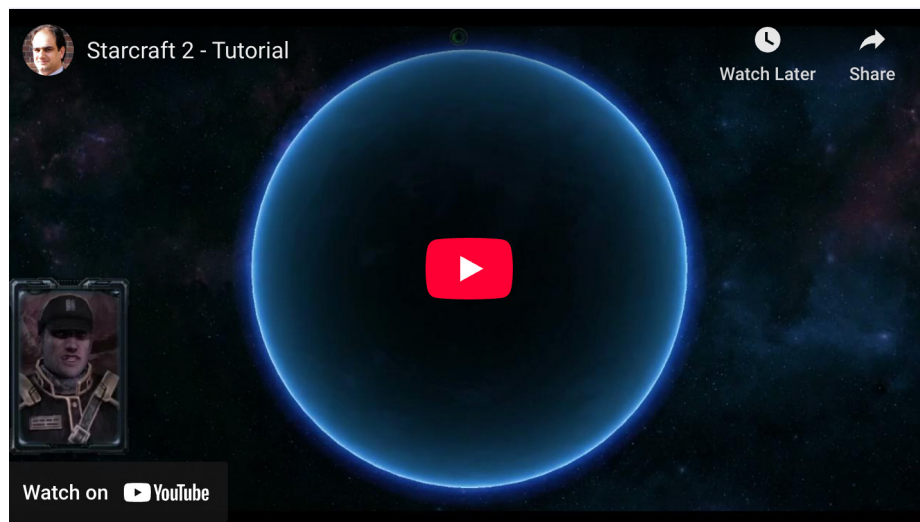


Figure 3: Starcraft II Tutorial interface showing (at 1:46 in the video) red text error messages on speaking android's profile and (at 5:00) red text error messages accompanied by potentially disembodied female voice of android.

Once the tutorial's objectives have been reached, an incoming communication takes over, replacing the android's feed with another from a male human officer of the Dominion who enlists the player into an upcoming mission. Once the mission has been accepted and the setting changes to the planet on which the mission is set, the voice becomes more natural and fitting for the scenario, but occasionally, reference is still made to the player's screen as more features of the UI are introduced. What is interesting, however, is that when the player attempts an illegal action, such as training units when sufficient resources are unavailable, the female voice of an android – such as the one encountered in the initial tutorial – reports back the erroneous action in addition to the red error text messages. An android's animated profile, representing the selected Command Centre, does not lip sync with the voice, thereby putting the provenance of the voice into question and suggesting a disembodied voice of another off-screen android. The movement of the game's voices between diegetic and non-diegetic lead to a trans-diegetic experience that, together with the HUD overlay onto the 3D game setting, fits well within the omnipresent interaction mode offered by the Real-Time Strategy (RTS) genre of this game.

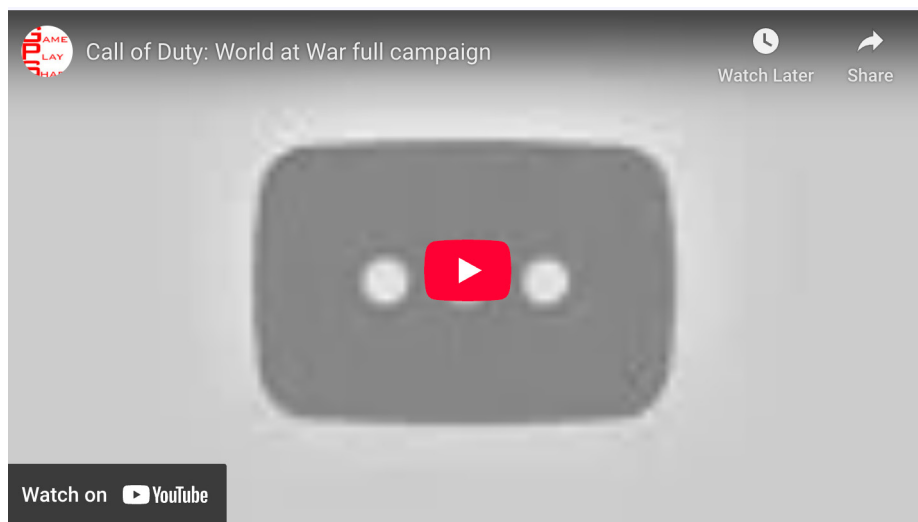


Figure 4: In *Call of Duty: World at War*, the player is verbally and mimetically addressed as they are (at 2:12) given a helmet and gun and (at 2:16) ordered to grab a rifle. Instructions on how to carry out the latter order are given non-diegetically

Diegetic voice is often situated inside the virtual environment via spatialization effects and the Barthesian concept of 'voice grain' to assist in matching the aural dimension with the visual representation of the source. Such diegetic voice needs to be loaded with emotional semantics in order to help deliver the narrative and cohere it with the gameplay [5]. The tutorial for *Call of Duty: World at War* [25], for example, starts with the player as a prisoner of war in a Japanese camp during WW2 who gets saved by a platoon of US soldiers. The rescue team address the player as Miller, hand him a helmet and a gun and then instruct him to take a rifle

see Figure 4). The soldiers' voices are raspy whispers, full of expletives and hatred toward the depicted enemy, which helps encourage the player's actions. The reference to the player's keyboard is made solely through a non-diegetic message with a different font than the ones used for the subtitles to the soldiers' diegetic voices. These subtitles, prefixed with the name of the soldier uttering them, help the player keep up to date with the mission objectives and his comrades' actions without having to look away from possible oncoming danger. The disembodied voices help sustain the characters' existence when outside the player's field of view. A study describes the voice of the player's mentor as 'familiar' and draws the player into the 'intense mindset' of the soldiers [26]. The minimal UI and the diegetic voices loaded with emotional semantics relevant to the narrative in *Call of Duty: World at War* are very fitting for the avatar-based interaction mode offered by First Person Shooter (FPS) games

### 3. "Who am I that speaks?" in Virtual Reality

VR marries theatrical mimetic performance with gameplay as, rather than pushing keys or buttons, the player virtually embodies the action-triggering performance. The question "Who speaks?" is pertinent to the VR player who, sensorially immersed in the virtual world, will seek out the owner of the voice being heard and attempt to identify their diegetic nature. This will inform the perspective, and thus the focalization, of this voice. Non-focalization gives access to information that would have been beyond one's knowledge. This may confuse the player and lessen realism, such as the unexplained knowledge of Captain Miller's journey to Ramelle by Private Ryan. Non-diegesis, or even trans-diegesis such as that of the *StarCraft II* tutorials, would lessen the VR player's immersion, because they are reminded of their physical self's role as an external audience. Diegetic voices presenting an external focalization, on the other hand, would represent a realistic character's perspective and thus heighten immersion [27], such as the soldiers in *Call of Duty: World at War*

Once the characters behind the voices around the player and their nature are identified and, especially, when addressed directly by these characters, the next question for the player is self-identification: 'who am I that speaks?' or rather, due to the affordable strong literal interaction [28] of VR, 'who am I that acts?'

#### 3.1 Second Person Voice and Self-Identification in VR

When the audience is addressed in the second person 'you' in a narrative, their presence in the virtual space is acknowledged and reaffirmed [29], whether they have no visible impact upon the narrative or its space, or they are given a role as co-creators[30], giving them a share in the responsibility of decision-making [31]. In digital interactive experiences such as first-person perspective games and interactive documentaries (i-docs), this responsibility demands knowledge of who they are relative to the storyworld and the characters within it, aided by the 'ocularization' and 'auricularization' chosen for the camera perspective [21]. Following



Following Bell and Ensslin [29], answering the question ‘who am I that acts?’ in VR demands sensitivity to media-specific affordances, specifically having the player entering the virtual world through the aural and visual senses of a virtual character diegetic to the storyworld and demanding an interactive role in the virtual space. So, who could this character be?

In traditional narratives (e.g., novels or films) the narratee may identify as: (1) the narrator, who may or may not be the protagonist of the story; (2) the protagonist of a story, told by a third-party narrator; or (3) an onlooker to a protagonist’s story, narrated by the protagonist or a third party [32]. These identities are now considered in a VR context:

The first identity has the player identify as the narrator by assigning a voice to the player’s avatar that mismatches the player’s own voice but this breaks immersion [33]. Thus using the player’s own voice for commands or reading out choices may be considered instead, as used successfully in games like *In Verbis Virtus* [34] and *Phasmophobia* [35]. Using the player’s own voice would allow us to address the original question, ‘who am I that speaks?’ and would also help further immerse the player through ‘ludonarrative consonance’ by matching voice with action [5]. Higher presence can be achieved if the voice is augmented with real-time echoes to model the spatialization effects of the virtual space being inhabited. Such effects suggest that the source of the sound is diegetically situated in the virtual space, and since the source of that sound is the player him/herself, then this supports the experience of telepresence [36].

The second identity has the player identify as the protagonist [37], which is akin to the use of second-person voice in *Choose Your Own Adventure* (CYOA) books, where players have a say in the narration of the story through the interaction afforded by the narrative device. The reader is addressed with a “you” as the main protagonist within the narrative (intradiegetic narration) but also as the decision-maker in charge of the non-diegetic interaction (extradiegetic narration), causing the narratee to shift alternately between the two [32]. Interactive Fiction, offline text adventure games, Multiuser Dungeons (MUDs), and their object-oriented variations (MOOs), as digital adaptations of CYOA books, also employ second-person voice as they address the player in the role of a character – the same player who writes textual commands to guide the narrative in their role as co-author [33]. In the case of the digital game *The Stanley Parable* [38], the player’s actions cause Stanley to behave differently from the narrator’s description, as the role of the player from narratee to protagonist slowly reveals itself to the narrator.

In the context of VR interactive narratives, however, having the player inhabit the avatar’s visual and aural senses may minimize the separation between intradiegetic and extradiegetic narration by embedding their interaction into their diegetic agency and may thus augment their feeling of presence in the interactive narrative’s storyworld. This may cause tension between the authorial control of the narrative and the agency provided to the player, known as ‘ludonarrative dissonance’ [39]. For the third identity of the player as an onlooker, Larsen [40] suggests a solution in the use of second-person point-of-view that allows viewers to

participate as sidekicks (diegetic observers) to the protagonist without the ability to modify the narrative structure. This solution lets the audience tackle side quests alongside the main narrative, while still allowing the main narrative to progress resolutely, irrespective of player action. Larsen's approach also suggests a preference for non-protagonist roles of the VR player in order to support 'ludonarrative consonance.' Thus, the character may be a bystander who perceives the protagonist as Other, as they suffer the narrative (narratee != protagonist). Whether the audience are passive observers or active sidekicks will depend on the agency afforded to the player [40].

## 3.2 Case Studies

We now consider the use of voice in a close reading of two VR documentaries. Both are situated during World War II, but our interest is in their use of voice, rather than their setting or subject matter. Both case studies project the viewer as a non-protagonist non-narrating character, but they use second-person voice differently and provoke different levels of self-identification through the provided agency..

### 3.2.1 Case Study #1: The Last Goodbye (2017)

Pinchas Gutter is an 89-year-old Polish Jew who survived the concentration camps of Nazi Germany. A series of interviews were held with Gutter in order to capture his memories and give evidence of the Holocaust atrocities perpetuated against the Jewish communities in Europe. There were four recorded narrations of Gutter about his experiences: two video interviews held during the 1990s, a collection of video-recorded replies used for an interactive interview [41] and a voice-over for the first ever Holocaust VR film entitled *The Last Goodbye* [42], in which the player accompanies Gutter on a visit to the Majdanek concentration camp through a 360 camera. The visit is narrated by Gutter: as a guide explaining the function of the place and his memories of it and as a voice-over in the opening and closing scenes, as well as during the visit to the crematorium, which brings him back memories that are too emotional for him to revisit.

While he never addresses the player with an absolute "you," he often acknowledges the player's virtual presence by breaking the fourth wall and looking at the camera, rather than at the objects he is describing. The effect is to help the player focus on the subject. The experience offers no interaction with the objects except for room-scale movement and observation, which sometimes leads to the 'Swayze effect' [33]. The illusion of presence is further broken by the penultimate scene when the camera drone's shadow, cast on the grass as we accompany Gutter, is revealed by the sun. The single use of a derivative of "you" is left until the final scene. We are beside Gutter, who is seated on a bench. A young boy, possibly Gutter's nephew, comes along on a scooter and talks with him inaudibly. The boy goes away on his scooter and Gutter, in a voiceover tells us that he expects better times for all of us in the future: "not in my lifetime, but maybe in yours..."

Our unacknowledged, intangible presence as Gutter looks away, the very recent presence of the young boy, and the suggestion that we are listening to his thoughts through the use of voice-over rather than being addressed directly, makes us wonder whose future Gutter was referring to: the implied player's or the young boy's.

### 3.2.2 Case Study #2: The Book of Distance (2020)

Randall Okita is the nephew of Yonezo Okita, a Japanese immigrant to pre-WW2 Canada who left behind his family in Hiroshima to seek a new life amongst other Japanese immigrants. Randall serves as the embodied narrator who takes the VR player along an imagined journey of his grandfather Yonezo in the VR short documentary entitled *The Book of Distance* [43]. In the experience, Yonezo is presented through original photographs, digitized and made virtually available for picking up for close-up inspection, during which Randall's voice explains who is who in the photos. A game of horseshoes – which serves as a connection between Yonezo and his nephew as evidenced by one of the photos – is presented to the player as an ice-breaker into the interactive narrative, empowering the player with agency and bringing out Randall's character. Playing on the lack of ambient light, Randall often moves off into the shadows, shifting the player's attention to the unfolding story that his off-screen voice narrates. Members of Yonezo's family, such as Randall's father, are presented through their Japanese cartoon representation as well as the playback of recorded vocal interactions. Yonezo's family in Hiroshima, including his younger sister, are presented through animated colored silhouettes but are not given a voice, reflecting the narrator's lack of familiarity with them.

Throughout the first part of the experience, the player acts as a sidekick to Randall's grandfather, helping him pack his clothes and photographic camera into his luggage for his voyage to Canada, responding to his family and future wife's waving whilst on the ship, having his passport stamped at the Canadian customs, taking photos of his housebuilding activities, giving a hand in clearing up the land, building the house, sowing the strawberries and serving them at the dinner table. All this ends abruptly as World War II starts and Yonezo's family are taken away by Canadian military to an internment camp for the Japanese. All their possessions are taken away: the camera, the house, the strawberry business, and likewise the player's narrative agency is greatly reduced.

Yonezo eventually returned to freedom, but it was hard earned and he never spoke about it to his nephew. Randall's lack of familiarity with this part of the story is reflected in the narrative's shift of focus away from the grandfather onto the father, and in the player's lack of agency, distancing them from the storyworld. The photographic camera is back in post-war Yonezo's hands as he takes photos of his growing children, photos that the player had seen at the beginning of the story. These are now brought back to the player's scrutiny with the addition of more recent, colorful photos of the grown characters. The story ends with the player

looking through photos of Randall's father's childhood while, in the background, Randall discusses them with his father, seeking to further understand the last years of his grandfather's life in Canada.



Figure 4: The book of distance VR experience.

## 4. Discussion

The two case studies provide contrasting examples of the use of second-person voice and its effect on self-identification from the viewer's behalf. With respect to the Holocaust VR film *The Last Goodbye*, Zalewska [42] reports that the subject of the VR film is the player's experience of the Majdanek camp rather than that of Gutter's. This is provided through an autodiegetic narration providing an internal focalization of Gutter's traumatic experience as he imparts his painful experience of the concentration camp and his turmoil in remembering nothing about his twin sister except her golden braid. The 3rd person ocularization and auricularization, which is later revealed to be that of an impersonal camera, offers no requirement or incentive for the player to identify themselves with Gutter or with any other relevant character. The player is offered no meaningful agency except linear narrative progression. Thus while it is clear that Gutter is the one 'who speaks', there is no self-identification for the player.

*The Book of Distance* tackles the VR player's self-identity very differently. The player is immediately given agency through a book presented in front of them with instructions to turn the page – reminiscent of CYOA books. From the beginning, the player is addressed with the second-person voice: "To You, the Time Traveller" and, as soon as Randall (the narrator) makes an entrance, he addresses the player directly, teaching them how to throw the horseshoe. This second-person address of the player continues throughout the journey of discovery with the narrator's voice (who speaks?) embodied by Randall's character explicitly represented in the scene or implied to be hiding in the shadows around the player.

The Book of Distance makes use of mechanics and structures frequently used in games and traditional films, even though the experience itself is neither. The empowerment of the player's agency and its subsequent reduction is a common technique used to challenge the player by limiting his/her skills. In The Book of Distance, however, the effect is not a feeling of increased challenge but a sense of loss of freedom, of identity. It starts with a prologue where modern-day Randall presents his grandfather, whose story we explore together throughout the rest of the experience, until we reunite with his father for the story's epilogue. We are then told exactly how Randall knows of his grandfather's life: through the photos, his father's testimony, and the letters received from the Canadian government, to name a few. The experience uses touched-up heterodiegetic narration, attempting an internal focalization as the nephew tries his best to understand his grandfather's experience of his life as a Japanese immigrant in war-time Canada.

Self-identification suffers, however, as the actual role of the VR player is indeterminate. Observation of Randall's exploration of his grandfather's story through his photo album evolves into active participation as he starts recounting his grandfather's story. As players, we become active sidekicks to the grandfather, helping him paint, write letters, prepare the luggage for his voyage to Canada, and making him aware of his family waving him off. But as the experience progresses, it is not clear whose sidekicks we are: Randall's or his grandfather's. As we help the latter build his house, sow his fields, serve the strawberries on the table, the question begs to be asked: who am I that acts? A friend of the family perhaps? This becomes even less clear when, as Yonezo is separated from his family and taken away to a field, it is the player who gets to raise the lever that traps the Japanese farmer inside the internment camp. Are we now sidekicks of the Canadian government? It appears to be impossible to reconcile these three different roles: chronologically separate sidekicks to the nephew and to the grandpa and morally separate sidekicks to the grandpa and to the Canadian military, and it raises the question: Whose side is the sidekick on? As a result, the player identifies with none of these diegetic roles, but instead the experience constructs a non-diegetic player role whose responsibility is to push the narrative forward. By not sticking to a specific persona, the virtual character that is embodied becomes transient across time and actions, such that the question, 'who am I that acts,' does not resolve to a specific characterization, leaving the player with themselves as enactors of the experience. This was a design choice with the player only identifying themselves as 'a part of the story' [43].

Thus, neither of the two experiences manages to successfully assist in self-identification. The Last Goodbye, through its limited agency and non-address of the player only serves to inform of the terrible loss of Pinchas Gutter at the hands of the Nazi regime. The Book of Distance makes it a point to continuously address the player and to provide agency that serves ludonarrative consonance as the gameplay contributes to the progression of the narrative while the empowerment and reduction of agency progresses with the narrative. However, as the experience follows a prologue/interlude/epilogue structure pertinent to theatre and film, the diegetic identity of the player transcends time, and the fruitful agency of the player's character places them on the same diegetic space as the younger grandfather and his family to which the nephew

seems to have no access, except as an observer. The already fragile link between the two transchronological personas is further disconnected with the entrapment action that results in Yonezo Okita's confinement in the internment camp, the character's body language showing as much disbelief at the player's traitorous act as his introvert self could afford.

It seems that involving the player to push on the narrative ever onwards was given greater priority than having them identify with a specific character within the grandfather's story and its exploration.

## 5. Accompanying IDN

This article is accompanied by a novel contribution in the form of a short interactive narrative resulting from the analysis carried out above. It attempts to demonstrate the use of second person voice, perspective, gameplay and self-identification using a prototypical Twine IDN that mimics VR interaction. The reader is invited to become an interactor and attempt a few replays before referring to the below explanations. There are four different possible endings.

**All five persons in the Principal's office  
had their eyes on Tom.**

**"His behaviour has changed lately",  
muttered Ms Spinnerwick.**

**"I wonder what's gotten into him..." her  
voice trailed off unconvincingly.**

**Look at Ms Spinnerwick.**

**Look at the principal.**

This short IDN uses second-person voice to address the interactor when a new character comes into the room and seeks further information about what is going on. This helps situate the interactor into the scene, pins their perspective to an occupant of the room, and gives them a say into the narrative's outcome. The gameplay is provided by shifting their attention to one of the characters present or by choosing what to say, each of which determines narrative progression. This eventually facilitates self-identification as the interactor falls into a specific role: the bully or the friend.

The setting is the principal's office in a school where three students are present while a teacher, Ms. Spinnerwick, is complaining about one of these students' behavior, identified as Tom, in order to hide the bullying behavior of one of her favorite students who is also present. After she has made her claim, a sixth character enters the scene, a Mr. Brimstone, who we later learn is an Education Officer (EO), who demands to know what is going on. After the Principal explains, the EO turns to address the interactor (who at this point is unaware of their identity: they could be the bully or the third student) and asks if there is anything else they want to add.

Besides clicking on NPC utterances for narrative progression in the initial stages of the IDN, the agency provided to the interactor are twofold: looking at specific characters and choosing what to say. Looking around is an interaction highly afforded by Virtual Reality headsets while speaking is an agency that this paper has emphasized as a potential interactive mechanism for self-identification. Thus, this Twine IDN can be considered as a prototype for a VR interactive narrative.

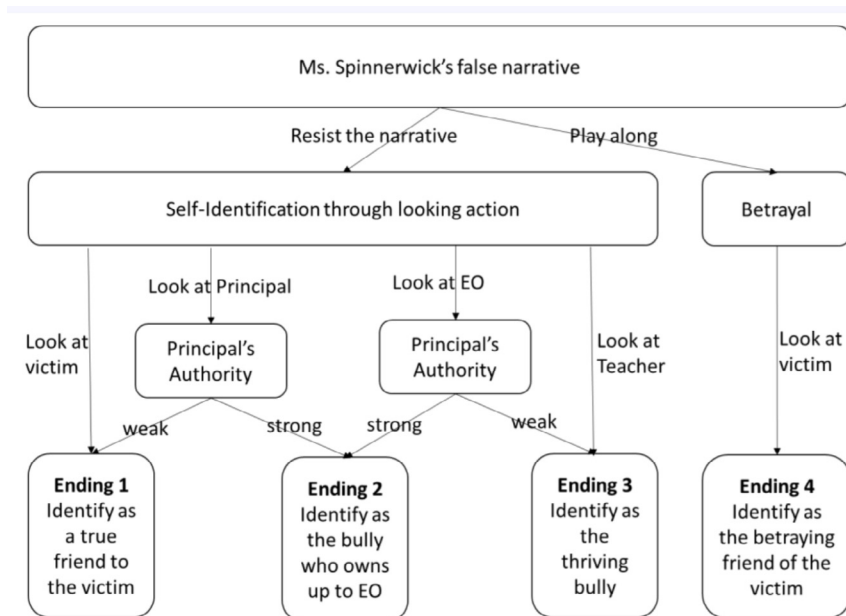


Figure 5 IDN's flowchart



Particularly, the identification process is carried out by the looking mechanic. This was preferred over the use of vocalized statements as the latter would have been too obvious. Rather, at the moment of identification, the interactor is given the choice to look at four characters: Ms. Spinnerwick, Tom, the Principal, or the EO (see Figure 5). If the interactor looks at Ms. Spinnerwick, then they are looking for support, because they identify as the bully. If the interactor chooses to look at Tom, it's because they seek his counsel on what to say, as they identify with his friend. If the interactor looks at the Principal, they are judging his strength of character. If weak, then here's the chance to bring justice to Tom, as they identify as his friend. If the principal has a strong character, this inspires the interactor to identify as the bully and own up. If the interactor's gaze falls onto the EO, ignoring the other occupants of the room, they identify with the bully, confirming Ms. Spinnerwick's statement on Tom's bad behavior and blaming it on the Principal's weak character, or owning up on his bullying behavior if the Principal was found to be of strong character.

Thus, through their looking and their speaking, the narrative may conclude in one of four endings. In two of the endings, #2 and #3, the interactor identifies with the bully while in endings #1 and #4 the interactor identifies with Tom's friend. Whether the bullying behavior is uncovered depends on two factors: the Principal's strength of character and the faithfulness of Tom's friend. If the Principal is weak, then there will be impunity and the bully (revealed to be Peter) will thrive (endings #3 and #4), unless the friend is strongly faithful to the bullied boy (ending #1). A strong Principal will drive the bully to own up (ending #2) unless the friend betrays the bullied boy (ending #4). While the latter is determined by the interactor's choice of words, the Principal's weakness is determined by whether the interactor chooses to look at the Principal in the opening sequence, putting pressure on him and finding him weak. This is reflected in additional descriptions along the narrative as well as available narrative choices.

## 6. Conclusion

Providing enough information to allow the VR player to self-identify is an important factor in their immersion into a virtual world. The alternate sensory information being provided to the player's eyes and ears must present a virtual world that is supported by its visuals and sounds in order to support the suspension of disbelief and telepresence into the virtual world.

Narrative theory on fiction provides answers to the 'who speaks?' question by identifying possible relationships between the narrator, the protagonist and the narratee on a spectrum having them as three separate characters at one end, and all in one character at the other. Another aspect of the spoken text is the content itself: what is being said. How that knowledge is known by the speaker reflects on the Genettian perspective of focalisation. The narrator-protagonist is expected to show internal focalisation while a non-protagonist diegetic narrator is expected to demonstrate superficial knowledge, thus external focalisation.



Unrestricted knowledge of the storyworld reflects non-focalisation and projects a non-diegetic narrator or an impersonal voice.

The use of narrators in mimetic forms such as film and theatre are commonly used in epilogues and prologues and are similarly used in gameplay tutorials in games. The diegetic nature of the narrator depends on the game's genre with the diegetic use in first-person view games being most relevant to virtual reality. Once the owners of the voices around them have been identified, VR players will next question their relationship to these characters: 'who am I that speaks/acts?'

Having the sole function of a narrator is superfluous to VR unless the player's own voice is used to move the experience forward, such as in *In Verbis Virtus* [34]. Being the protagonist implies a certain amount of agency that may conflict with the narrative direction of the experience, favoring the role of the bystander, or in Larsen's words, 'virtual sidekick' [40], tackling side-quests along the game spine [44] that is the main narrative. Moreover, having the player acknowledged and addressed by the diegetic characters and favoring the diegetic narrator through touched-up heterodiegetic narrations with internal focalization improves immersion through telepresence.

In light of the literature, two case studies were analyzed in terms of the use of secondperson voice and interactivity. The scripted VR film, *The Last Goodbye*, in which the player passively accompanies a Holocaust survivor on a visit to the concentration camp where he lost his family, does not address the player resulting in reduced telepresence. On the other hand, the 'virtual pilgrimage' *The Book of Distance* addresses and empowers the VR player to be the narrator's grandpa's virtual sidekick as he emigrates from Japan to Canada to start a family and a business, but loses all when he ends up in an internment camp during the Second World War. As the diegetic narrator attempts to understand his grandpa's experience, a touched-up heterodiegetic narration with internal focalization is seen [19] becoming more pronounced as the protagonist's social belittling is reflected in the reduced agency of the player, making his loss felt in parallel by the player. What diminishes the players' immersion is the non-identification of whose sidekick one is playing: the narrator's, the protagonist, or the antagonist?

Thus, the use of diegetic narrators and having the VR player act as a sidekick with agency magnitude that reflects the protagonist's power over his narrative suggest greater levels of immersion for the player once they have identified the owners of the voices around them and their character's relationship with them. A short IDN prototype attempts to present an example where the interactor may identify with the antagonist or with the protagonist's sidekick, through actions that mimic VR interaction such as looking and speaking.

Future work ought to improve upon the work by Vosmeer et al. [33] to explore any causal relationship between the use of second-person voice to address the VR player and the level of identification perceived. This study explored off-the-shelf VR experiences and presents arguments that favor such a relationship.

## Note

This article is based on an earlier publication ([https://doi.org/10.1007/978-3-030-92300-6\\_1](https://doi.org/10.1007/978-3-030-92300-6_1)) published as part of the Proceedings of ICIDS 2021 by Springer. This extended version includes a more extensive literature review, explanatory figures, as well as an accompanying IDN that demonstrates the concepts presented in the article..

## References

- Bernstein, D.: Creating an Interactive Audio Environment. *Gamasutra*. 1, 2013 (1997).
- Ekman, I.: Meaningful noise: Understanding sound effects in computer games. *Proc Digit. Arts Cult.* 17, (2005).
- Bala, P., Masu, R., Nisi, V., Nunes, N.: “ When the Elephant Trumps” A Comparative Study on Spatial Audio for Orientation in 360o Videos. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. pp. 1–13 (2019).
- Wallmark, Z., Kendall, R.A.: Describing sound: The cognitive linguistics of timbre. *Oxf. Handb. Timbre Adv. Online Publ.* N. Y. NY Oxf. Univ. Press. 14, (2018). <https://doi.org/10.1093/oxfordhb/9780190637224.013>.
- Ward, M.: Voice, Videogames, and the Technologies of Immersion. 15 (2010).
- Gardner, K.: 5 Braaiinnsss! *Technol. Gothic Lit. Cult. Technogothics*. 71 (2015).
- Guerilla Bandit: *Zombies, Run!*, (2012).
- Barker, J.M.: *The tactile eye: Touch and the cinematic experience*. Univ of California Press (2009).
- Busselle, R., Bilandzic, H.: Measuring Narrative Engagement. *Media Psychol.* 12, 321–347 (2009). <https://doi.org/10.1080/15213260903287259>.
- Pólya, T.: Omitting Depth Cues: The aesthetics of perceptual reflexivity. *Narrat. Spectatorship Mov. Images*. 246 (2009).
- Mildorf, J.: Reconsidering second-person narration and involvement. *Lang. Lit.* 25, 145– 158 (2016).
- McMahan, A.: Immersion, engagement, and presence: A method for analyzing 3-D video games. In: *The video game theory reader*. pp. 89–108. Routledge (2013).
- Brown, E., Cairns, P.: A grounded investigation of game immersion. In: *CHI’04 extended abstracts on 14. Human factors in computing systems*. pp. 1297–1300. ACM, Vienna, Austria (2004).
- Lombard, M., Ditton, T.: At the heart of it all: The concept of presence. *J. Comput.-Mediat. Commun.* 3, JCMC321 (1997).
- Nordahl, R., Nilsson, N.C.: The Sound of Being There: Presence and Interactive Audio in Immersive Virtual Reality. In: *The Oxford Handbook of Interactive Audio*. Oxford University Press, United Kingdom (2014). <https://doi.org/10.1093/oxfordhb/9780199797226.013.013>.
- Genette, G.: *Narrative discourse: An essay in method*. Cornell University Press (1983).

- Vendryes, J.: *Traité d'accentuation grecque*. C. Klincksieck (1904).
- Saving Private Ryan. Dreamworks Distribution (1998).
- Aare, C.: A Narratological Approach to Literary Journalism: How an Interplay between Voice and Point of View create empathy with the Other. 34 (2016).
- Bellardi, M.: The cinematic mode in fiction. *Front. Narrat. Stud.* 4, s24–s47 (2018).
- Schlickers, S.: Focalization, ocularization and auricularization in film and literature. In: *Point of view, perspective, and focalization*. pp. 243–258. de Gruyter (2009).
- Nielsen, H.S.: The impersonal voice in first-person narrative fiction. *Narrative*. 12, 133–150 (2004).
- Richardson, B.: Point of View in Drama: Diegetic Monologue, Unreliable Narrators, and the Author's Voice on Stage. *Comp. Drama*. 22, 193–214 (1988). <https://doi.org/10.1353/cdr.1988.0017>.
- Blizzard Entertainment: *StarCraft II: Wings of Liberty*, <http://eu.blizzard.com/engb/games/sc2/>, (2010).
- Treyarch: *Call of Duty: World at War*, (2008).
- Ramsay, D.: Brutal games: "Call of duty" and the cultural narrative of World War II. *Cine. J.* 94–113 (2015).
- Barreda-Ángeles, M., Aleix-Guillaume, S., Pereda-Baños, A.: An "Empathy Machine" or a "Just-for-the-Fun-of-It" Machine? Effects of Immersion in Nonfiction 360-Video Stories on Empathy and Enjoyment. *Cyberpsychology Behav. Soc. Netw.* 23, 683–688 (2020).
- Ryan, M.-L.: Immersion vs. interactivity: Virtual reality and literary theory. *SubStance*. 28, 110–137 (1999).
- Bell, A., Ensslin, A.: "I know what it was. You know what it was": Second-Person Narration in Hypertext Fiction. *Narrative*. 19, 311–329 (2011).
- Walmsley, B.: Co-creating theatre: authentic engagement or inter-legitimation? *Cult. Trends*. 22, 108–118 (2013).
- Ciancio, G.: Active spectatorship, changes and novelties in the performing arts sector. In: Bonet, L. and Négrier, E. (eds.) *Breaking the Fourth Wall: Proactive Audiences in the Performing Arts*. pp. 90–96 (2018).
- DelConte, M.: Why you can't speak: Second-person narration, voice, and a new model for understanding narrative. *Style*. 37, 204–219 (2003).
- Vosmeer, M., Roth, C., Koenitz, H.: Who Are You? Voice-Over Perspective in Surround Video. In: Nunes, N., Oakley, I., and Nisi, V. (eds.) *Interactive Storytelling*. pp. 221–232. Springer International Publishing, Cham (2017). [https://doi.org/10.1007/978-3-319-71027-3\\_18](https://doi.org/10.1007/978-3-319-71027-3_18).
- Ferrari, M.: In Verbis Virtus. Presented at the (2015). [https://doi.org/10.1007/978-3-319-05326-4\\_18](https://doi.org/10.1007/978-3-319-05326-4_18).
- Kinetic Games: Phasmophobia. Presented at the (2020).
- Johansson, M.: VR For Your Ears: Dynamic 3D audio is key to the immersive experience by Mathias johansson· illustration by Eddie guy. *IEEE Spectr.* 56, 24–29 (2019).
- Fludernik, M., Fiction, S.P.: Narrative YOU as Addressee and/or Protagonist". *Arb. Aus Angl. Am.* 217–247 (2005).
- Wreden, D., Pugh, W.: *Stanley Parable*, (2011).

Hocking, C.: Ludonarrative Dissonance in Bioshock, [http://clicknothing.typepad.com/click\\_nothing/2007/10/ludonarrative-d.html](http://clicknothing.typepad.com/click_nothing/2007/10/ludonarrative-d.html), last accessed 2016/01/24.

Larsen, M.: Virtual sidekick: Second-person POV in narrative VR. *J. Screenwriting*, 9, 73– 83 (2018). [https://doi.org/10.1386/josc.9.1.73\\_1](https://doi.org/10.1386/josc.9.1.73_1).

Traum, D., Jones, A., Hays, K., Maio, H., Alexander, O., Artstein, R., Debevec, P., Gainer, A., Georgila, K., Haase, K., others: New Dimensions in Testimony: Digitally preserving a Holocaust survivor's interactive storytelling. In: *International Conference on Interactive Digital Storytelling*. pp. 269–281. Springer (2015).

Zalewska, M.: *The Last Goodbye* (2017): Virtualizing Witness Testimonies of the Holocaust. 8 (2017).

Oppenheim, D., Okita, R.L.: The Book of Distance: Personal Storytelling in VR. In: *ACM SIGGRAPH 2020 Immersive Pavilion*. pp. 1–2. ACM, Virtual Event USA (2020). <https://doi.org/10.1145/3388536.3407896>.

Bateman, C.: *Game Writing: Narrative Skills for Videogames* (Charles River Media Game Development (Paperback)). (2006).

An Academic Publication of the Association for Research in Digital Interactive Narratives <https://journal.ardin.online>