




# Instinct-Based Decision-Making in Interactive Narratives

Tobías Palma Stade<sup>(✉)</sup> 

Digital Creativity Lab, University of York, Heslington YO10 5GE, UK  
tobias.palmastade@york.ac.uk

**Abstract.** This paper examines the expressive potential of instinct-based decision-making as a method to enhance narrative immersion in interactive storytelling. One of the key challenges to propose leaned-back interactive narratives lies in the methods through which users and system exchange inputs and outputs. While explicit interfaces tend to disrupt leaned-back participation – demanding a leaned-forward type of agency – and, thus, immersion in the narrative environment, this model proposes interactions based on diegetic stimuli that avoid interfaces and encourage instinctive and immediate reactions from the user in order to navigate a narrative immersive environment. The notion of instinct-based decision-making has been observed in three stages: (1) Conceptualization through previous practices and literature review, (2) Design and production of a Cinematic Virtual Reality (CVR) interactive prototype, and (3) system-testing of the model’s key functional aspects.

**Keywords:** Interactive storytelling · Immersive environments · Instinctive decision-making · Cinematic virtual reality · Hands-off interactivity

## 1 Introduction

Post-broadcasting media – as it has been named by some media scholars [10] – is defined by the agency gained by audiences, by a much more hands-on approach to the curation of a variety of formats and contents. Yet, the consumption itself still remains mostly *leaned-back*: “TV viewing tends to be a more passive, ‘lean back’ experience, in contrast to the active involvement required of the ‘lean forward’ computer/Internet activity” [9]. In recent years, researchers and practitioners have started to attempt to merge these two approaches. Today we can see that in interactive television (SmartTVs, Netflix, Amazon Video, and many other On-demand platforms), once we select the content, the act of “watching the telly” remains mostly the same – sometimes complemented by side-consumption of complementing content through our mobile phones or other devices. While the content is accessible to be curated, it isn’t by itself interactive. For now.

Technology has reached a point in which the exploration of interactive film and television content is possible, and imagining it in “the living room of the future” [1] seems feasible, if not immediately, in the short term. The fast development of immersive and haptic technology and its application into television devices might accelerate the

leap into the next stage of interactive media in the following years. In the eventuality of this scenario, we have the possibility to explore and experiment with these technologies, in order to understand their potentially expressive means.

This paper focuses on the exploration of instinct-based interactions as a way to enhance leaned-back and hands-off interactive storytelling. The notion of interactivity has been for long related to lean-forward formats such as video games, that require explicit interaction through a device – usually a remote control of a certain kind – and always on a conscious level in which users are aware of the decisions they are making and their effects in the narrative. Instead, this project aimed to explore the possibilities of a leaned-back interaction defined by unconscious decision-making, through minimizing explicit interaction and focusing on implicit interaction and on the use of diegetic stimuli, enhanced by the characteristics of an immersive medium like Cinematic Virtual Reality (CVR).

## 2 Conceptualization

### 2.1 Immersion, Interactivity, and Agency

Generally, the definition of interactive storytelling is founded in the tension between the agency given to the user and the author’s narrative intentions [3, 8, 11, 19] or a *Free Will versus Determinism* conflict, or a *Man versus Author* conflict. As in any interactive storytelling experience, the problem lies in the paradox of interactivity: “[T]he integration of the unpredictable, bottom-up input of the user into a sequence of events that fulfills the conditions of narrativity – conditions that presuppose a top-down design” [19]. Most of the revised literature discusses interactivity as the ability to affect the storyworld from *within* – considering that most of the authors refer to interactive storytelling mainly in video games or similar media.

According to Marie-Laure Ryan, “the combination of narrativity and interactivity oscillates between two forms: the *narrative game*, in which narrative meaning is subordinated to the player’s actions, and the *playable story*, in which the player’s actions are subordinated to narrative meaning” [19]. Especially nowadays, these two forms correspond more to a spectrum than to a binary categorization. Another way to understand this would be according to the objectives of a specific interactive story, which can oscillate between performing a series of tasks and accomplishing missions, to an intrinsically contemplative reception of the story.

Michael Mateas’ more Aristotelian approach to interactivity distinguishes between *Interactive Drama* and *Interactive Storytelling*: “In interactive drama, the player assumes the role of a first-person character in a dramatic story. The player does not sit above the story, watching it as in a simulation, but is immersed in the story” [11]. On the other hand, Interactive Storytelling would not necessarily involve a *player*, but a user that, without being involved in the story – or *storyworld* – has the ability to manipulate elements of it.

Chris Crawford’s definition of interactivity refers to it as a *conversation*: “A cyclic process between two or more agents in which each agent alternately listens, thinks, and speaks” [3]. According to this, in interactive storytelling the narration must contain several stimuli across its timespan to allow at least an equal amount of reactions from the

user, where the term *reactions* summarizes Crawford's three steps of listening, thinking, and speaking, which he also clarifies are not to be taken literally, but rather as receiving the stimuli, processing it, and responding to it.

A more specific debate around interactivity is the paradox between narrative immersion and agency: "[A] distinction should be made between ludic and narrative immersion. Ludic immersion is a deep absorption in the performance of a task. Comparable to the intensity with which a mathematician concentrates on proving a theorem, or a soloist performs a concerto. (...) [N]arrative immersion is an engagement of the imagination in the construction and contemplation of a storyworld that relies on purely mental activity" [19]. TV audiences are not players and agency is not a condition for audiences to be immersed in stories. On the contrary, it is the inclusion of agency in storytelling that apparently affects narrative immersion.

This complicates the definition of agency. One perspective sees it as "the feeling of empowerment that comes from being able to take actions in the world whose effects relate to the player's intention" [11] a position that stresses that the user's agency must have effects on the storyworld [3]. Another, more dynamic vision observes that interactive narratives face the challenge of reconciling the tensions between immersion and agency, since the latter would affect the former [20]. As for immersion, Ermi and Mayra define it as "becoming physically or virtually a part of the experience itself" [4] a definition that is quite generous, as it can be applied to narrative and/or environmental immersion. They also distinguish three different types of immersion: Sensory, challenge-based, and imaginative. "Sensory immersion can be intensified with better graphics and sound; challenge-based by engaging gameplay [...] and imaginative immersion as a 'game experience in which one becomes absorbed with the stories and the world'" [7].

Another dimension of the *agency vs immersion* problem is what Janet Murray approaches with the concept of *threshold objects*. These are devices that "take us across a symbolical and sometimes literal passageway" [13]. Murray and Ryan both criticize the real sense of immersion that these devices provide since, as a physical material object, they are also the material evidence and reminder that the storyworld is an illusion. They argue that interactivity which relies too much on the threshold device fails to provide an effective sense of immersion, since it would be the same object what pulls the user back from the story. While Murray refers specifically to physical devices – especially prosthetic devices, that are in any way attached to our bodies, such as controls or Head Mounted Displays (HMD) – on-screen non-diegetic interfaces have similar characteristics, in terms of, on the one hand, facilitating and allowing interactivity as they serve as a toolbox to affect the digital storyworld, since their usually non-diegetic and inorganic presence is a reminder of the nature of the immersion, and could even suppose a visual barrier as a sort of emerging fourth wall. Moreover, the counterintuitive usage of these devices and interfaces suppose a process of learning, adaptation, and acquisition of skills to properly perform the agency on the interactive narration. And yet, these same devices and interfaces are the ones that allow interactivity in the first place, which is the base for the paradox presented threshold objects.

This is one of the issues that leaned-back and hands-off interaction schemes aim to attack: How to provide interactivity with little-to-none friction in terms of immersion, aiming to reduce the explicit, non-diegetic, and inorganic, presence of threshold objects

and interfaces. The current technological developments point to opening possibilities to this kind of interactivity, with technologies like eye-tracking, or motion-sensing input devices like haptics-based technology, that could allow interactivity with little effort from users and little friction in terms of crossing a physical threshold.

The challenge, in these hypothetical scenarios, would be to create strong narrative immersions based on story elements and not relying upon the consciousness of the devices. In this sense, amusement parks are usually referred as the best examples of spatial immersion with very restrained interactivity, but that doesn't necessarily affect the immersion [8, 13, 19]. This is based on the principle of *environmental storytelling*, which "creates the preconditions for an immersive narrative experience in at least one of four ways: spatial stories can evoke pre-existing narrative associations; they can provide a staging ground where narrative events are enacted; they may embed narrative information within their *mise-en-scène*, or they provide resources for emergent narratives" [8]. Similarly, *spatial stories* are defined as "stories which respond to alternative aesthetic principles, privileging spatial exploration over plot development. Spatial stories are held together by broadly defined goals and conflicts and pushed forward by the character's movement across the map", where "the organization of the plot becomes a matter of designing the geography of imaginary worlds, so that obstacles thwart and affordances facilitate the protagonist's forward movement towards resolution" [8]. A practical approach to this proposed type of narrative is the application of Spatial Story Density (SSD) as the measurement of story elements that are arranged in the space of a scene simultaneously [23]. This means that spatial stories require a balance in the information contained in the spaces, in the same way, linear storytelling is founded on the principle of tension-and-release; the balance and dynamism of events taking place at a fast pace against moments of tension being built upon slower pace and moments of more reflexive contemplation. In other words, "with high temporal story density, narratives are fast paced, with high spatial story density, many narratives are happening simultaneously. Mental effort increases when temporal or spatial story density is high" [5].

## 2.2 Guidance

One of the key principles to consider in the engagement and interaction in immersive media is guidance, which refers to the methods used to establish a dialogical relation between system and user. Immersive environments present a particular challenge in terms of guidance because the surrounding space tends to be more stimulating and distracting than screen-based media, where attention can be more easily directed by the author. The dialogue between system and user should be subtle enough so the user doesn't feel that they are losing their agency. Guidance should feel like the system is suggesting decisions to the user, never imposing them.

Previous studies have explored the possibilities to provide guidance from within the narrative world in cinematic VR, [14, 22] in other words, encouraging diegetic guidance. For this, we can discriminate between explicit and implicit cues: "Implicit cues are likely, but no necessarily contingent upon bottom-up salience, whereas explicit cues are likely to cause voluntary, top-down shifts in attention" [14]. By this definition, most non-diegetic guidance corresponds to explicit cues, in the sense that attention is being explicitly called by the narrator.

Immersive spaces, by allowing the user more direct access to the diegetic storyworld, also allow more direct access to diegetic cues. Speicher et al. [22] specifically focused their research on how much attention do users pay to different types of stimuli, both visual and acoustic, diegetic and non-diegetic. One of their key findings is that users welcome guidance, as it facilitates participation, while Passmore et al. note that users tend to feel that navigation and exploration are facilitated by guidance: “The ability to look around is linked to the users’ certainty about what they should be looking at, and their attention to, and concentration on, story” [15]. From then on, the question is what types of guidance are preferred. Gödde et al. identified five types of diegetic attentional cues in their study on cinematic conventions on CVR, which include: Gazes, Motion, Sound, Context (narrative elements like anticipation or suspense), and Perspective (the disposition of elements in the space by proximity, size, etc.) [5]. To these, Brillhart has added timing and rhythm [2]. She developed an editing technique she calls Probabilistic Experiential Editing, which aims to facilitate fluidity between cuts through the synchronic presence of relevant dramatic elements in the virtual space. This understanding of timing is considerably more subtle than what is broadly advised by other authors, “to give around at least 20 seconds per shot for the viewer to orientate to the new scene, unlike traditional TV, where shots can be much shorter” [15]. Brillhart’s notion of timing and editing precisely aims to avoid the need for orientation through hopefully instant recognition of what she calls Point of Interest, or simply put, “places in a scene where a viewer is likely to be looking” [2]. Timing lies on a swift and organic cut cued by the visual elements and their role in their surrounding environment.

The conceptualization of these notions aimed to provide a theoretical, yet practical, framework to facilitate a type of interaction that relies on the user reacting to the diegetic elements strategically organized in an immersive environment through the application of SSD. While the proposed storyworld responds mainly to a top-down system which facilitates a series of diegetic stimuli, the prerendered nature of CVR makes the author concedes the narration to however the user reacts to these stimuli and whichever decisions they make from such reactions, ideally resulting in the self-editing of a unique and unrepeatable version of the story proposed by the author.

### 3 Methodology

The study followed a research-by-practice approach, with a strong interdisciplinary perspective that merged techniques from diverse expressive disciplines, mainly filmmaking and computer design, but also including theatre, screenwriting, and sound design. Its development consisted of three stages: Writing and Design, Production of a prototype (which in many ways resulted in an extension of the design), and a System-proof testing.

In the first stage, the processes of writing and design informed each other simultaneously in order to meet specific needs from one and the other. The second stage, the production of the prototype, prolonged the design of the model through the exploration of diverse expressive techniques in CVR. Despite its cinematic technical qualities, CVR presents a fertile ground for challenges and experimentation from multi and interdisciplinary approaches. There have been diverse experiences that have explored editing and montage in cinematic VR [2], adaptation of filmmaking practices into immersive

narratives [12, 15], the properties of guidance [5, 14, 22], and that observe the feasibility of theatrical techniques in immersive spaces [16, 17].

This particular study added a somewhat more thorough and dramatic use of acoustic stimuli in immersive environments, exploring the potential of sounds as diegetic stimuli to provide guidance. This appears to be a novel approach to VR, since “[t]he technical literature exploring VR’s potential is somewhat ocularcentric, focusing on two dominant visual techniques: computer-generated imagery (e.g. 3D environments) and 360° videography (a.k.a. “cinematic VR”). Yet VR also affords virtual soundscapes via ambisonic” [6]. Just like films make a dramatic use of sound, this study aimed to adapt some of the expressive acoustic techniques used in filmmaking into immersive narrative environments.

The final stage of system-proof consisted on a very basic demonstration of the prototype to a total of seven people, and focused on the functionality of the expressive concepts applied during the two previous stages. The sample involved in the evaluation, this was composed of: volunteers (4) – from here on referred to as participants – who got to watch the prototype and participate in a semi-guided focus group, experts (2) from academia and the industry, who watched to prototype and responded to a semi-structured interview designed for experts in the field, and crew members (2) who answered a survey at the end of the production process, and another (1), the sound designer, who watched the prototype and participated of a semi-structured interview.

It is worth noting that this consultations took place during the Covid-19 pandemic, particularly during the second lockdown, which affected the original planification for the system test. The consultations that took place consisted in simplified versions of this planification, in order to obtain sufficient data to evaluate the prototype and the key concepts behind its design.

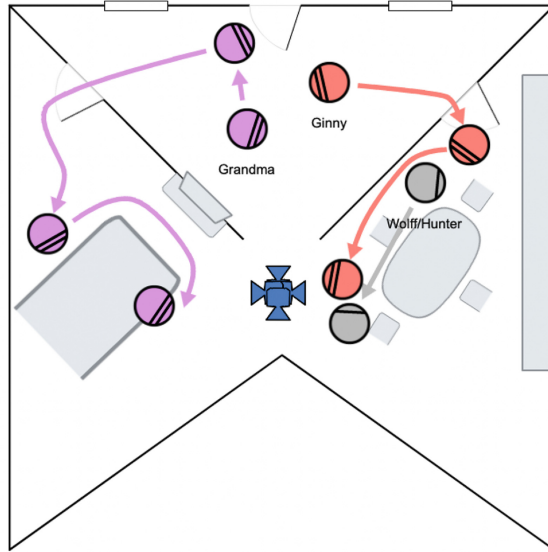
## 4 Design and Production

### 4.1 Scriptwriting

The design and the scriptwriting processes informed each other simultaneously, taking into account that the conception of the story had to meet the requirements for an immersive and interactive model based on diegetic stimuli and instinct-based interactions. This stage aimed to apply the notion of Spatial Story Density, focusing on a) how actions and characters are distributed across the surrounding virtual environment, and b) how to use soundscape as an articulation between micronarratives and as a unifying element of the overall drama.

The idea for the technology is simpler than the architecture of the story. It consists of a virtual reality space divided into relatively similar angles, set from the perspective of a first-person viewer in the centre of the space. The surrounding sphere is then divided into as many acoustic environments as wished – in this case, three. In Fig. 1 we can see an example of a virtual space divided into four acoustic areas, so that the viewer, placed in the middle, has the ability to distinguish what’s happening in each of these areas independently.

Applying this conception of the virtual space, the script consists of two parallel plots –one with its own main character: Ginny and the Grandmother– which we can see in the



**Fig. 1.** Diagram of a virtual space divided in acoustic sub-spaces, where different micronarratives take place.

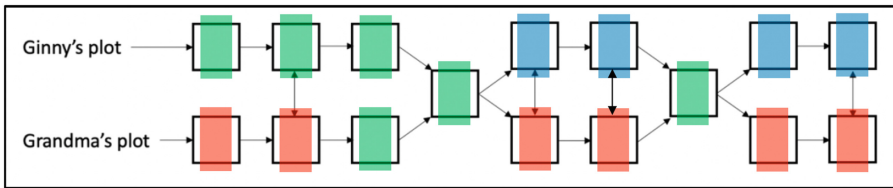
form of interlinked outlines in Fig. 2. Each plot consists of nine bullet-points that balance each other respectively, while they also mark the turning points and the moments when the characters meet and depart. Both plots are interdependent, influencing each other and forming part of one overall story. There are events in Ginny's plot that affect the Grandmother's and vice versa. There are also external events – embodied in secondary characters – that affect both protagonists, hopefully in a proportionate manner.

| INTELINKED OUTLINE                              |   |
|---|---|
| Grandma and Wolff are cuddling                  | Ginny and Hunter are discussing on the phone    |
| Wolff leaves                                    | Hunter arrives by surprise                      |
| Grandma runs into Ginny while looking for Wolff | Ginny runs into Grandma while hiding Hunter     |
| Grandma tries to hide her affair                | Ginny she tries to hide Hunter                  |
| Grandma hears Hunter – she believes he's Wolff  | Ginny kicks Hunter out                          |
| She flirts with Hunter                          | She finds Hunter's present                      |
| Ginny finds them                                | Ginny finds Hunter being courted by her Grandma |
| Hunter/Wolff leaves                             | Hunter/Wolff leaves                             |
| Grandma starts to believe she's delusional      | Ginny realizes she's losing her boyfriend       |
| She decides to deny it                          | She decides to take him back                    |

**Fig. 2.** Interlinked Outline, we can see the two plots' bullet-points and how they keep the synchronicity and balance of the dramatic development of both plots.

The script itself was reformatted into this two-columns layout, to facilitate the choreography of the actions between the two plots and to provide the reader a good sense of simultaneity. Screenwriting required the conception of both space and time manipulation: To give users the ability to explore space, while the author choreographs time.

The outline shows the story divided in a series of units, or micronarratives [3]. These take place in one virtual world but are relatively disarticulated from each other until each micronarrative develops into the next. If the user were to choose to follow only one character through the whole story, there might be a risk to have an isolated plot and an incomplete story. Micronarratives must be articulated and linked to each other through their position in a dramatic space susceptible to being explored. Therefore, users must have some kind of consciousness of a larger world beyond each micronarrative, so that they would feel tempted to explore it. The soundscape provides that sense of a world, first through the construction of an acoustic environment, but mainly through the use of dramatic sound cues that dominate the development of the story over the interdependent micronarratives.



**Fig. 3.** Diagram of the narrative structure of *The Hunter & the Wolff*. Each square corresponds to a scene or micronarrative. The unidirectional arrows signal the linear temporal direction of the story, while the bidirectional arrows mark the points where sounds cues can be heard in all the immersive space.

Figure 3 shows how micronarratives articulate to each other linearly in time and in space, forming a narrative structure. While both plots unfold linearly, the simultaneity is illustrated by the parallel development. The merging micronarratives mark the points in the story when the plots merge both in time and space (which also divide the structure into thirds, following the three acts structure). The bidirectional arrows also mark the moments when sound cues can be heard in all of the immersive environment, linking parallel micronarratives together. While the medium allows the user to switch between spaces and plots at any time, and not necessarily in the points marked in the diagram, this illustration shows how narrative structures incorporated notions of interactivity and spatial storytelling.

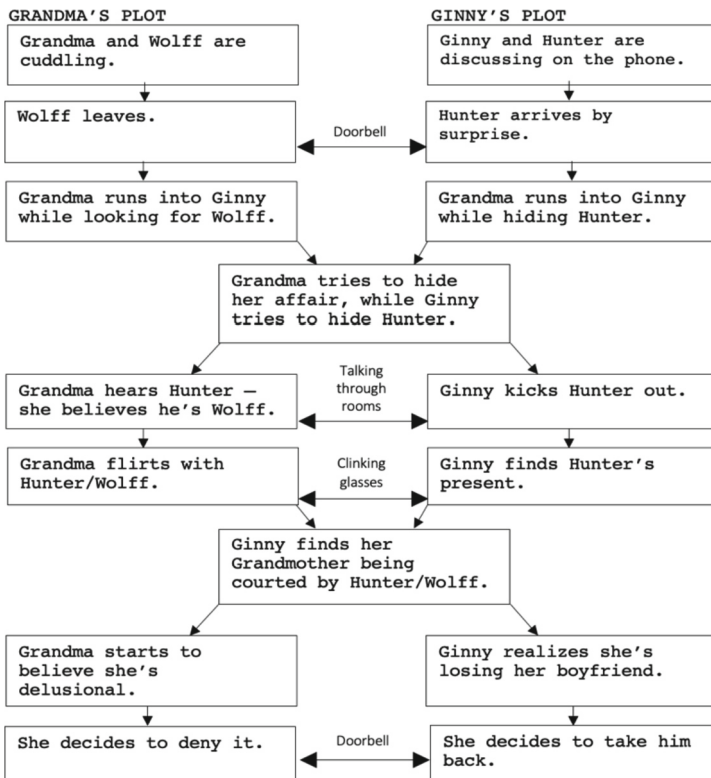
In more practical terms, the overall, unifying soundscape is reflected in the script through *cue sounds*, which are very specific sounds related to the organic environment and that have direct dramatic influence over the characters. While each micronarrative is distinguished by its own soundtrack – which allows users to isolate the micronarrative from any other external disturbance and, therefore, focus on it – these cue sounds form part of the larger structure of the story and are omnipresent; users will hear them despite the micronarrative they’re following at a given moment.



|   |   |
|---|---|
| <p>a dark purple night gown. She combs her hair, joyful.</p> <p><b>THE DOORBELL RINGS.</b></p> <p>GRANDMA<br/>(singing)<br/>Mirror, mirror in the wall,</p> | <p>Don't say that!<br/>I'll compensate you, you know...<br/>Hunter, behave!... Hunter?</p> <p>The phone call ends abruptly, but the <b>DOORBELL RINGS</b> immediately. Ginny goes back...</p> <p>INT. LIVING ROOM</p> |
|---|---|

**Fig. 4.** Cue sounds. In this excerpt we can see how an overall sound is present in both micronarratives, having dramatic influence in the two plots (Color figure online).

In Fig. 4 we see an example of this principle. Stressed in yellow and caps we can read the cue “DOORBELL RINGS” relatively in the same line of each plot (Ginny’s on the left, Grandma’s on the right). This is a very short excerpt but long enough to notice that the actions going on are completely different for each character, yet the sound cue generates a reaction. These sounds have three key functionalities in the medium: dramatic, environmental, and interactivity-driver. They are dramatic because they generate reactions in the characters and keep the plots moving; environmental because they belong to the diegetic, organic world and hold the micronarratives together into one dramatic space; and they are the organic interactivity-drivers that are expected to stimulate users to *turn around* and *look the other way*. They are one of the main stimuli –apart from characters’ movement and other visual elements of the blocking and *mise-en-scène*– that are supposed to generate organic, instinctive reactions on users.



**Fig. 5.** Detailed diagram of the narrative structure of *The Hunter & the Wolf*. This merges Fig. 2 and Fig. 3, illustrating how the micronarratives are organized through the parallel timelines how these are interconnected with cue sounds.

Figure 5 illustrates how the two plots are organized simultaneously and shows, on the one hand, the moments where both plots merge spatially and, on the other, the sound cues that link parallel micronarratives. It also provides a visualization of how the two plots are intended to balance and complement each other.

## 4.2 Design

In this model, the virtual space is divided into three, which also has consistency with the environment: From left to right, the Grandmother's room, a Hall, and a Kitchen, which also provides a more instinctive and organic way to produce a mental map. In the script, the two main characters are relatively assigned one of the spaces, where most of their plots take place: Grandma's plot takes place mostly in her room, and Ginny's in the kitchen, while the hall serves as a transition area where characters meet each other and transit.

Spatial-narrative thinking greatly differs from traditional filmmaking; "it turned out that planning and shooting a film in 360° does not work the way it does in normal movies" [5]. Of the challenges for filming in CVR, the impossibility to frame is probably the easiest to spot. Of course, we lose the close-ups and, therefore, a very important tool to guide the spectator's attention to specific and sometimes small details. The impossibility to place and move the camera in relation to the characters forces the necessity of the opposite approach: To move characters and objects in relation to the camera. This choreography ought to be thought as a way of generating visual stimuli that could guide and suggest possible decisions to the user, [5, 17] and can benefit greatly from disciplines that traditionally concentrate in the use of space as a narrative device, like theatre [16]. Indeed, if in films the camera is the main narrator, in CVR this role seems to fall on the space.

Another important aspect derived from the absence of framing in CVR, is the loss of out-of-frame or off-screen as a narrative and expressive resource. A common way to approach this in film is through sound, a principle that founds the proposal for an acoustic-guidance in CVR. 360° acoustic environments, such as binaural or virtual soundscapes, provide the possibility to feel surrounded by virtual acoustic stimuli. This, backed with a visual representation of such stimuli, provides the possibility to navigate this environment. Immersive acoustics provide a spatial architecture that uses all three vectors, while traditional cinematic acoustics tend to rely on a horizontal spectrum, despite the number of channels. In this sense, immersive acoustics could potentially provide very complex tools to explore virtual spaces and to develop spatial storytelling, whether cinematic or not, but especially CVR, considering its restraints derived from being based on prerendered assets.

This principle is complemented by the limited human visual field of view, which is hindered even further in VR. This natural condition – enhanced by the technology – offers the possibility of designing and managing stimuli outside the user's range of view. The user simply can't see the whole virtual environment at the same time, and yet, they should be able to hear everything going on around them and, in doing so, creating a mental architecture of the acoustic space. In other words, the user is able to hear what they don't see. This condition is a fertile field for the adaption of off-screen narrative in film language into an out-of-sight narrative in spatial storytelling: CVR provides

the possibility to counterpoint visual and acoustic guidance, offering simultaneous and diverse stimuli to the user.

In CVR the camera becomes an embodiment of the user's presence, differently from film, where the camera is the narrator and a mediator between the spectator and the story, leaving them outside the storyworld. CVR requires the understanding that such mediation vanishes, facilitating the provision of agency to the user.

The user's narrative presence or embodiment has to be addressed as early as working in the script, since it affects the blocking and the relation actors, props, and scenery in the virtual environment have with the camera. We see that "in VR, framing and camera position correlate with each other even more than in normal film" [5], referring to the fact that camera position might have more drastic consequences for the user's immersion than those of the film camera to the film spectator, where the camera operates as a mediator instead of as a threshold [13].

The absence of mediation puts the user in a determining position from the start of the narrative, because the creation-of-sense of the world relies more on them than in a narrator that is now absent and has transferred their abilities to the user. The notion Initial Viewing Direction (IVD) is then crucial in immersive narratives, since they might determine the user's mental architecture of the narrative space for the rest of the story or, at the very least, the scene. Previous research shows that IVD is "accepted as the 'correct' viewing direction for both sitting and standing viewers, and the action is usually anticipated to begin there. Hence, the attention usually goes back to the IVD after the orientation phase, except when an attentional cue leads to a potential Point Of Interest (POI) somewhere else" [5]. In this proposal, which is based on the simultaneity of two plots with two main characters, the IVD might determine the perception that users have of the story. Considering there are two starting micronarratives taking place simultaneously at the very beginning of the piece, the IVD is assigned randomly by the system to provide equal chances of starting with one or the other character.

Once transported to the virtual environment, the user starts to make sensorial sense of the space. Immersive technologies so far have had a tendency to recreate naturalistic environments, particularly in the use of sound. On the contrary, this sound-based interactivity device challenged the naturalistic approach. To provide a more dramatic and interactive use of acoustic stimuli, sound mixes in each of the micronarratives are purposely manipulated to induce dramatic a reaction in the user, guiding them for narrative purposes. In other words, not only there are three independent acoustic spaces but also they are intervened to make them dramatically inter-dependent. The dramatic use of sound has a long history in screen-based media, and it makes sense that the acoustic aspect would play a significant role in immersive narratives, especially if we consider that "anecdotal reports indicate that sound has a highly significant impact on presence, and one study showed that spatialized sound was associated with higher reported presence than either no sound or non-spatialized sound" [21].

The conception of a dramatic and non-naturalistic approach to an acoustic landscape necessarily affects the design of the spaces, since sounds is one of the main perceptual accesses to the space. Considering that the story consists of two plots that take place in three spaces, one of the most relevant corrections was rethinking empty spaces. Throughout the overall story, at least one of the spaces always remains empty, either because

one of the characters is absent or two of them are interacting in one of the rooms. This supposes many moments of wasted, undramatic spaces and the risk of having the user missing out part of the action(s) and, thus, being disoriented. Indeed, *Fear Of Missing Out* (FOMO, “a pervasive apprehension that others might be having rewarding experiences from which one is absent” [18]), which is usually understood as a risk and a liability of immersive drama, has been taken into consideration as an opportunity to enhance a dramatic engagement in the guidance device, a notion that can be summarized in the intention of users experiencing FOMO as a consequence of their own decisions – being forced to choose between two or more paths – and not because of the lack of guiding elements in the dramatic space.

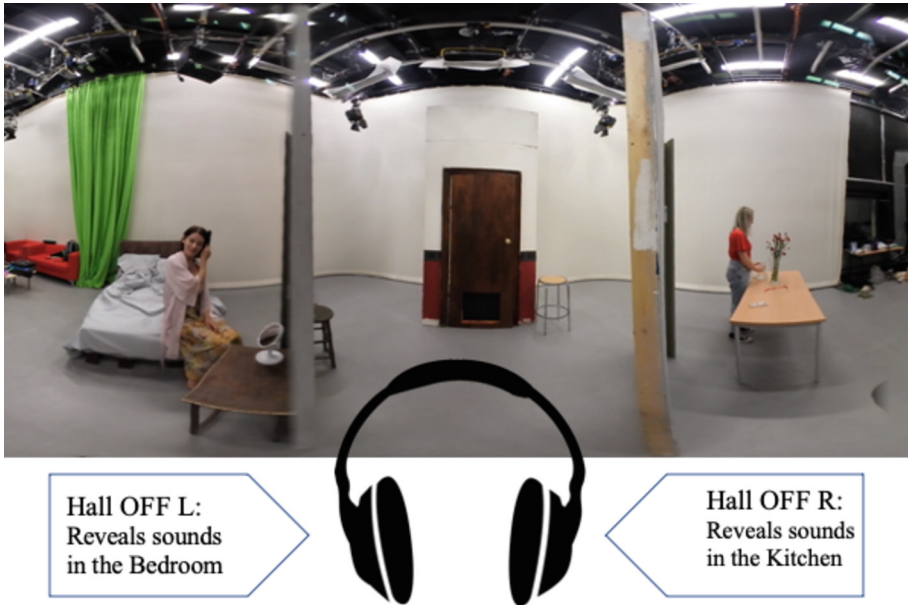
This has been tackled with the notions of ON and OFF mixes. ON mixes correspond to the acoustic presence of elements and events that take place in the room where the gaze is centred at a specific moment – and any sound relevant to the actions according to the script – while OFF mixes correspond to the acoustic presence of sounds in the next room, primarily dialogues and other relevant sounds. In a way, this mechanic is related to the notion of *on-screen* and *off-screen* diegetic sounds frequently used in traditional filmmaking. Once again, this conception of the acoustic space doesn’t respond at all to a naturalistic approach; ON and OFF mixes conveniently alternate in terms of functionality, requiring the OFF mix only when a room is empty, strongly suggesting the user, through sound, to move to the next room.

It is crucial to understand that OFF mixes are determined by the needs of each specific space. Each space is assigned ON and OFF mixes that are played continuously and simultaneously, but are mixed so they sound alternatively: They are always playing, but they turn quiet depending on the actions taking place in each room. For instance, when looking at the Bedroom when this is empty, then the Bedroom OFF mix would turn on, which corresponds to the sounds that come from the Hall, the room next door, guiding the user to assume that characters and actions are taking place in that room, as it can be seen in the bottom image of Fig. 6. Considering this prototype is still based on stereo mixing, this means that Bedroom OFF’s sounds are concentrated on the right speaker – and also mixed according to spatial specifications, like lower volume and lower reverberation. This is why the Hall has two OFF mixes; one according to the room on each side. To illustrate in the top image of Fig. 6, when there are no actions taking place in the Hall, but there are in the other two rooms, if the user places their gaze in the Hall they will be able to hear the Hall OFF L mix on the left speaker (what is happening in the Bedroom) and the Hall OFF R mix on the right speaker (what is happening in Kitchen). In Fig. 7, we can see an illustration of how, while the Hall is empty, different actions are taking place in the other two rooms, which we could still hear through the OFF mixes.



**Fig. 6.** Each of the three spaces contains a number of mixes (either 2 or 3). The figure shows how each mix “activates” depending on the actions taking place on each space.

If the acoustic dimension of this project has taken a particular prominence, it is because it is the main tool to provide guidance and facilitate hands-off reactions, the means through which the interactive environment and the user can establish a two-sided relation. In this specific type of interactivity guidance is a particularly delicate issue, since it has to be provided without breaking the illusion of freedom of agency nor the suspension-of-disbelief. Interactive narratives that rely on explicit and extra-diegetic guidance usually offer an interesting exploration experience but run the risk of attracting attention to the tool itself, undermining the immersion in the story. In this sense, extra-diegetic guidance, although sometimes fascinating by itself, can become an expression of the problem with threshold objects, which facilitates interaction while at the same time keeping the user from actually immersing [13].



**Fig. 7.** Organization of sound deliverables: When a space is empty and no actions take place, OFF mixes come into play and allow users to become aware of what is happening in the spaces next door.

In terms of visual guidance, the organization of this type of stimuli was approached through blocking techniques based in proxemics [16, 17]: “For example, high status characters tend to be given more physical space around them, tallying with findings that show that the person with the most speaking rights has more space around them” and “rather than manipulating space in a fixed frame, such as close-ups and wide shots, actors can manipulate spatial relationships between one another in a way that is familiar in theatre and in everyday life” [16]. Positioning a character in a certain way can suggest they are to be followed, while still leaving an open option for other decisions.

Guidance through movement is mainly used as a device to track characters as they move from one room to other. For instance, when one character leaves one room to go to another, leaving another character alone in the first room. This movement supposes a decision the user must make between staying with one character or leave with the other. All these instances are specific points in the script that were conceived to propose a decision, and were brought into the blocking with this into consideration, but also in which most of the characters’ own movements are driven by external factors and stimuli.

The main diegetic guidance methods used in this prototype were sound and movement. Through blocking, these elements were meant to motivate and potentiate each other – under the blocking principle that actions should always be motivated. For instance, sound cues and “off-sight” sounds (the immersive narrative equivalent to off-screen sounds) drive characters to move from one place to other. This is also the kind of stimuli that aims to suggest decision-making and to get the user’s attention.

## 5 Results

The consultation process that served as a system-test consisted on observing how participants reacted to the prototype and focused on three main aspects: a) Perception and use of acoustic stimuli, b) perception and use of visual stimuli, and c) decision-making processes. It worth noting that participants weren't given a very detailed description of the prototype beforehand; they were only told to face forward until the end of an initial clock count as a way to have a referential starting point, and from then on they had the freedom to navigate as they wished.

Regarding the a) perception and use of acoustic stimuli, participants found that sound was a very important element within the narrative, that it helped them to place themselves within the immersive space, to re-orientate whenever they felt lost, to follow actions, and to make decisions. The system-test intended to see if the interdependent soundscapes would hinder the orientation process at the beginning of the film in any way. On the contrary, not only it didn't undermine this process; participants found it helpful to orientate themselves and to facilitate the immersion in the story.

In terms of b) perceptions and use of visual stimuli, initial observations show that movement caught the participants' attention, maybe even more than it was originally desired. In most of the cases, the secondary character was identified as the main one because of having a higher frequency in movement and his more frequent transits between the three rooms. This has several interesting implications. First, it undermines the original narrative intention of making the main characters the protagonists of their own interdependent plots, but that tended to move less and remain in one or two spaces precisely to keep the plots separated. It would become necessary to explore possible strategies to make the suggestion of these leading roles more emphatic, and develop blocking and directing strategies that would drive more attention to them. We know already that sound has proved to be very effective in this sense, so monologues, soliloquies, leit-motivs, extending the dialogues, or other dramatically relevant sounds could help. We also know that movement works in a very similar fashion, so making the characters move within the space they are already in could help to keep the attention focused on them. Finally, a more accurate and intended use of proxemics could be interesting to see how this affects the perception of the characters. Proximity has shown to be somewhat complicated, since some of the participants declared to feel discomfort when they felt that characters were too close to them, even when they weren't in an intimate space [16, 17]. The reasons for this were specially mentioned: invaded intimacy – in one or both directions – and losing field of vision, declaring that being too close to one character narrows the scope of vision.

For the most relevant aspect of this proposal, c) the decision-making process, participants of the consultation made a good part of their navigational decisions in a rather instinctive manner. On the one hand, the fact that they had to make decisions on the fly, without time to stop to think about them, generated a sense of urgency that made them feel more engaged not only with the environment but also with the story. Primarily, users indeed found themselves not stopping to consciously think about their decisions (instead, conscious decisions had to be made on the fly) but rather following the stimuli, in this case, mainly the acoustic ones, although not exclusively, as they were also driven by the characters. On the other hand, some of the participants described their decision-making

process as more-or-less conscious, since they did realise they were making decisions but not necessarily *thinking* about them.

Participants weren't given a specific tutorial on the characteristics of the device before the test, which required of them to learn to use the system by themselves, which seemed to have taken a few seconds. The fact that participants didn't have any problem adapting themselves to the navigation method shows that the interactive device relies on instinctive reactions to stimuli more than on establishing a set of instructions for users to be followed. However, it is worth considering that these initial orientation moments could have been used for the sake of the story, and that this time could have been saved with a short set of instructions.

## 6 Conclusions

The overall design of an interactive story of this characteristics needs to consider the role and agency of the viewer, as well as the use and prioritization of different stimuli, both visual and acoustic. Participants confirmed what previous studies had found, that characters and elements that move tend to catch the viewers' attention, while more interestingly for this research, they also showed that acoustic stimuli not only helped to place themselves in the virtual environment, but also motivated them to explore the space, proving to be a very relevant tool for guidance. The adaptation of "off-screen" sounds into "off-sight" ones seems to have been proved successful in at least two dimensions: a) provides a dramatic spatial orientation, letting the user know that actions are taking place and where, and b) provides stimuli that motivate users to make decisions and navigate the dramatic space.

In terms of visual guidance, the use of theatrical blocking techniques proved to be very relevant not only to get the user's attention, but also to integrate the camera into the dramatic space, understanding, in the set, the camera is also a sort of and embodiment of the user. Even in the most passive treatment of the user, which is to assume them as a "fourth wall" type of spectator in which they don't integrate the dramatic space, the use of proxemics affects the hierarchy of visual stimulations that are intended to drive the user's decision-making. For instance, participants showed to be more inclined to follow characters that move, against those who are more anchored to one space. This is particularly relevant in a design like this, where following a character can drive you away from one action into another. This could suppose a grammar in which the relevance of characters and their actions might involve considering their movements and how they are presented in relation to the camera and other elements in the *mise-en-scène*.

Characters' movements are also relevant considering the importance of the Initial Viewing Direction (IVD), which tends to determine the user's initial choices. If in an IVD there's a character moving away, chances are that the user will probably follow this character. On the other hand, when the characters are still, users tend to make a quick exploration of the surrounding space, in the presumed safety that the character won't move yet. This also shows a sense of urgency that is quite determining in the decision-making process; when the story seems to be moving a little faster, users tend to make faster decisions, also motivated by fear-of-missing-out, in the conscience of the simultaneous actions.



It seems that the main aim of the CVR narrator would be to balance the stimuli directed to the user, in order to provide them with the moments and spaces to make narrative decisions. Despite being the elephant in the room, the incorporation of the user into the narrative is still the key challenge of immersive and interactive media, since creators can't longer hide behind a fourth wall and are being pushed to face the audience. Nevertheless, the adaptation and complementation of proved techniques from other expressive media are showing to be a fruitful enterprise.

## 7 Future Work

Precisely because the initial design for an instinct-based type of interactivity have had overall positive results, it is worth persisting in the exploration and experimentation of these methods for narrative purposes. The application of some of the key concepts related to diegetic stimuli to drive decision-making, hands-off interactivity, and the dramatic use of soundscaping, might be worth of further work not only in VR but in other immersive media as well. Considering the conditions proposed in this project, it doesn't seem impossible to apply them in MR or AR, as long as we count with suitable technology. For instance, one of the main conditions for the functionality of instinct-based decision-making is the absence of explicit interfaces that could generate a pull-back from diegetic immersion. With appropriate haptic devices, it seems feasible to imagine MR capable of providing such conditions, in which digitally generated diegetic elements blend with real environments. Similarly, dramatic soundscapes could be developed into a purely acoustic dimension, relying on binaural sound mixing or similar acoustic immersive means, as a way to provide a soundscape to a specific real location – in a museum, for instance – or simply as an independent sound-based storytelling medium. The main challenge for this would be, as it was for this project, to find the suitable technology that would gather inputs from the user(s) in order to generate the necessary dialogue for interactivity to exist. Either way, the conceptualization of these notions and a successful proof of concept, open the possibility to further explore them, and to test their resilience in other media.

On the other hand, due to the characteristics of the sample (and the disruptions generated by the global pandemic) it is important to acknowledge that these results do not intend to describe the behaviour of the audience, but to provide qualitative feedback on specific creative goals proposed for this prototype. In this sense, further research should also aim to obtain more conclusive evidence in regard of audience's behaviour towards these methods and this particular type of interactivity.

## References

1. BBC R&D: StoryFormer: Building the Next Generation of Storytelling, 2 January 2019. <https://www.bbc.co.uk/rd/blog/2018-10-object-based-media-tools-storyformer>. Accessed 28 Feb 2022
2. Filmmaker Magazine: Look into the Cut: Jessica Brillhart on Editing VR. Online resource. <https://filmmakermagazine.com/96090-look-into-the-cut/#.XVQAX5NKgWo>. Accessed 28 Feb 2022

3. Crawford, C.: *On Interactive Storytelling*. New Riders, California (2005)
4. Ermi, L., Mäyrä, F.: Fundamental Components of the Gameplay Experience: Analysing Immersion. *Digarec Series*, vol. 06, pp. 88–115 (2011)
5. Gödde, M., Gabler, F., Siegmund, D., Braun, A.: Cinematic narration in VR – rethinking film conventions for 360 degrees. In: Chen, J.Y.C., Fragomeni, G. (eds.) *VAMR 2018. LNCS*, vol. 10910, pp. 184–201. Springer, Cham (2018). [https://doi.org/10.1007/978-3-319-91584-5\\_15](https://doi.org/10.1007/978-3-319-91584-5_15)
6. Green, D., et al.: Using design fiction to explore the ethics of VR “in the wild”. In: *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video, TVX 2019*, Manchester, pp. 293–299 (2019). <https://doi.org/10.1145/3317697.3323346>
7. Haahr, M.: Creating location-based augmented-reality games for cultural heritage. In: Alcañiz, M., et al. (eds.) *JCSG 2017. LNCS*, vol. 10622, pp. 313–318. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-70111-0\\_29](https://doi.org/10.1007/978-3-319-70111-0_29)
8. Jenkins, H.: Game design as narrative architecture. *Computer* **44** (2002)
9. Katz, H.: *The Media Handbook: A Complete guide to Advertising Media Selection, Planning, Research, and Buying*, 2nd edn. Taylor & Francis; Lawrence Erlbaum Associates, Publishers. Mahwah, New Jersey & London (2008)
10. Lotz, A.D.: *The Television will be Revolutionized*. New York University Press (2007)
11. Mateas, M.: A neo-aristotelian theory of interactive drama. In: *Working Notes of the AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment*. AAAI Press (2000)
12. Mateer, J.: Directing for cinematic virtual reality: how the traditional film director’s craft applies to immersive environments and notions of presence. *J. Media Pract.* **18**(1), 14–25 (2017). <https://doi.org/10.1080/14682753.2017.1305838>
13. Murray, J.: Did it make you cry? Creating dramatic agency in immersive environments. In: Subsol, G. (ed.) *ICVS 2005. LNCS*, vol. 3805, pp. 83–94. Springer, Heidelberg (2005). [https://doi.org/10.1007/11590361\\_10](https://doi.org/10.1007/11590361_10)
14. Nielsen, L.T., et al.: Missing the point: an exploration on how to guide users’ attention during cinematic virtual reality. In: *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video, TVX 2019*, Manchester, pp. 229–232 (2019). <https://doi.org/10.1145/2993369.2993405>
15. Passmore, P., Glancy, M., Philpot, A., Fields, B.: 360 cinematic literacy: a case study. In: *International Broadcasting Convention 2017*, 14–18 September, Amsterdam (2017)
16. Pope, V.C., Dawest, R., Schweiger, F., Sheikh, A.: The geometry of storytelling: theatrical use of space for 360-degree videos and virtual reality. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI 2017*, May 2017, pp. 4468–4478 (2017). <https://doi.org/10.1145/3025453.3025581>
17. Probst, P., Rothe, S., Hussmann, H.: Camera distances and shot sizes in cinematic virtual reality. In: *ACM International Conference on Interactive Media Experiences, IMX 2021*, 21–23 June 2021, Virtual Event, NY, USA, pp. 178–186. ACM, New York, NY, USA (2021). <https://doi.org/10.1145/3452918.3458804>
18. Przybylski, A.K., Murayama, K., DeHaan, C.R., Gladwell, V.: Motivational, emotional, and behavioral correlates of fear of missing out. *Comput. Hum. Behav.* **29**(4), 1841–1848 (2013). [https://doi.org/10.1016/j.chb.2013.02.014\(2013\)](https://doi.org/10.1016/j.chb.2013.02.014(2013))
19. Ryan, M.L.: From narrative games to playable stories: towards a poetics of interactive narrative. *Storyworlds J. Narrative Stud.* **1**(2009), 43–59 (2009). University of Nebraska Press
20. Ryan, M.L.: The Interactive onion: layers of user participation in digital interactive texts. In: *New Narratives: Stories and Storytelling in the Digital Age*, pp. 43–59. University of Nebraska Press (2011)

21. Sanchez-Vives, M., Slater, M.: From Presence to consciousness through virtual reality. *Nat. Rev. Neurosci.* **6**, 332–339 (2005)
22. Speicher, M., Rosenberg, C., Degraen, D., Daiber, F., Krüger, A.: Exploring visual guiding in 360-degree videos. In: Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video, TVX 2019, Manchester, pp. 1–12 (2019). <https://doi.org/10.1145/3317697.3323350>
23. Unseld, S.: 5 Lessons Learned While Making Lost. Oculus Story Studio (2015). <https://www.oculus.com/story-studio/blog/5-lessons-learned-while-making-lost/>. Accessed 4 Feb 2020