The Hitchcock Experience - a Spatial Montage project

Rémi Sagot-Duvauroux[1], Nils Quaetaert[2], François Garnier[1], Rémi Ronfard[2]

[1]École des Arts Décoratifs – PSL (EnsadLab)
[2]Univ. Grenoble Alpes, Inria, CNRS
[1]Paris, France [2]Grenoble, France
remi.sagot@gmail.com, nils.quaetaert@polytechnique.edu,
francois.garnier@ensad.fr, remi.ronfard@inria.fr

Abstract

This paper describes an ongoing art project exploring the generalization of classical montage theory to the emerging technology of “room scale VR”. We take the emblematic “crop duster” scene in Alfred Hitchcock's *North by Northwest* and invite the spectator to experience it “from the inside”. This raises interesting research issues. How to inhabit this cinematic story world? How can the traditional tools of cinematography and montage be used to direct the audience in this new kind of experience? How can the rhythm of the spectator's body be matched to the rhythm of George Tomasini’s (Hitchcock’s editor) fast pace editing? And what form can this montage take and for what aesthetic and dramaturgical effects? To answer these questions, we propose the experiment of a montage that adapts in real time to the displacements of the body and the gaze of the spectator, engaging a dialectic between narrative rhythm and bodily rhythm. We propose new algorithmic tools to transport the audience into the story as originally planned by Hitchcock while at the same time respecting the behaviors of the audience to guide the experience. In doing so, we seek to create a new form of relationship between the author of a narrative experience in virtual reality and the spectator who explores and activates the experience with his own body.

Keywords

Virtual Reality, Montage, Narrative, Aesthetics, Human-Computer Interaction, Computer Graphics.

DOI

10.69564/ISEA2023-82-full-Sagot-Duvauroux-et-al-The-Hitchcock-Experience

Introduction

While the imaginary around virtual reality devices generates fears and fantasies, it seems essential to question the explicit and implicit bodily interactions that we have with digital spaces. By its algorithmic nature, the virtual environment, calculated in real time, constitutes a functional space that can be manipulated by the immersant (1) but can also control and constrain his/her movements and the sensitive experience, according to a relationship set by an author.

These explorable environments can also carry within them an emotional dimension due to their digital image nature. When dealing with a virtual reality device, our bodily habits of space cohabit with our visual habits of image. We think therefore that it is possible to invoke our culture of the cinematographic images in the spatially lived aesthetic experience.

With the cinema, we agree to sit still for the duration of a film and let ourselves be hypnotized by the gaze of a director. This privileged moment, out of time, offers a frame to rethink our relationship to images (to their time, rhythms and meanings). Depending on the quality of the montage, we leave our own space-time aside to align our perception and our thoughts with the rhythm proposed by the film.

Thanks to virtual reality devices, we are able to inhabit a cinematic representation space. How then to adapt a cinematographic dramaturgy while allowing the emergence of new phenomena related to the involvement of the body in action specific to the immersive nature of virtual reality displays? Should we be limited to explore the scene without discontinuities or should we experience a form of montage? To answer these questions, we designed a virtual reality experience (Figure 1) that immerses us inside the mythical “crop duster” scene in Alfred Hitchcock's *North By Northwest.* (2)

Figure 1. Point of view of the immersant in our experience.

In the original film, the montage by George Tomasini (Hitchcock’s editor) is essential to a good understanding of the space and the distances that separate the character (played by Cary Grant) from the other characters and vehicles that interact with him [1] . Moreover, its evolution through the sequence highly participates in suspense, tension and violence. But if the spectator is physically immersed in this space and able to move through it, the rhythm of the editing needs to be matched to the one of his/her body. This implies a new aesthetic and dramaturgical approach due to the spatial shift of the virtual reality medium. These are the issues that guide this research-creation project entitled: *The Hitchcock Experience* (3).

We propose the experiment of a montage guided by the movements of the immersant. It explores the questions of distance and identification with the character by transposing certain cinematographic shot scales into the scale of the immersed body regarding the represented space. Moreover, this experiment proposes a form of editing that adapts in real time to the displacements of the body and the gaze of the immersant, engaging a dialectic between the narrative rhythm and the one of the experienced space. We wish to explore a new relationship to the ongoing narrative, profitable to the emancipation of the immersant and harmonizing his/her bodily rhythm to the one of a fragmented inhabitable environment.

The aim of this experiment is not to propose a VR experience that competes with Hitchcock's film. By confronting this masterpiece of the cinematographic montage, we precisely try to understand, through comparative analysis, the profound transformation implied by its adaptation in virtual reality.

First, we will establish related works regarding the practice of montage in virtual reality devices. Then, we will explain our research-creation process and comment on the different steps we went through developing our spatial montage system. We will expose our experimental protocol that we set to collect data from 17 participants. Then, we will show our first relevant results and observations. Next, we will discuss and initiate theoretical explorations on the aesthetic experience of the immersants. Finally, we will conclude on our spatial montage device and the role of its editor.

Related work

The problem of directing techniques suitable for building "room-scale games" and more specifically "room-scale movies" is widely overlooked in recent academic research. Contemporary room-scale VR experiences are based on the strong sensation of presence caused by forcing the viewer to remain in a fixed space and time. As a result, the use of montage is the exception rather than the rule. Hodgkinson [2] describes some techniques used by Google Spotlight Stories [(4)] for moving the immersant in space and in time during a VR experience, which bears some resemblance with traditional montage. For example, *Rain or Shine* (directed by Felix Massie) uses visual occlusion to smooth out unwanted camera motions. *Pearl* (directed by Patrick Osborne) introduces temporal ellipses which can guide the narration, while keeping the same spatial location. *Age of Sail* (directed by John Kahrs) allows the immersants to navigate between a small number of spatial locations, in effect allowing them to create their own spatial montage.

The question of montage has been more deeply explored in the related, but different, context of cinematic VR, where the immersant can only turn his/her head horizontally across 360 degrees and vertically across 180 degrees. If he/she can choose the direction of his/her gaze in the sphere image, the viewer cannot move his/her point of view to explore the spatial dimension of the scene, which is the specificity of room-scale VR. In this limited context, Jessica Brillhart[3] has proposed the notion of a “probabilistic experiential montage” where each 360 degree shot offers the possibility of multiple experiences by the immersant, and the montage is created on the assumption of the most probable experience.

Garnier[4] provides a detailed account of the various geometries involved in viewing film, either on a traditional screen or in a virtual reality headset, and emphasizes the importance of a proxemic interpretation of shot sizes in film. From another perspective, Pope [5] also recommends the use of proxemics for staging in VR following some well-established theater techniques.

Rothe et al.[6] propose an analysis of camera control in cinematic VR and also relate shot sizes and camera distances under the framework of proxemics. Rothe also examines the effect of camera height in distance and size perception in cinematic VR[7].

In our previous work entitled *Reframing VR,*[8] we propose a vocabulary of shot values suitable for room-scale VR based on the new notion of a spatial frame. In our framework, spatial montage can be defined as a temporal arrangement of spatial frames with different spatial scales, which play a similar role to the shot sizes used in traditional film theory. Because the immersant is free to move in the virtual world, “shot size” is not measured only in terms of apparent visual size, but also in motion parallax and capacity of action.

Artistic process

Regarding this state of the art, we propose a process of research-creation, based on an artistic device engaging a reflexive loop between practice of the "spatial montage" and theorization. In this part, we explain how we elaborated our artistic experiment.

Scene reconstruction

Based on a 3D scene provided by the Anima team of Inria [(5)], we virtually reconstructed the entire set and all the props, characters and vehicles present in the scene using representative keyframes from the movie. We then created a rough "layout" animation of all the character and vehicle movements, synchronized frame by frame with the original movie. We used a floor plan view diagram from Raymond Bellour’s *The Analysis of Film* [3] as a reference to approximately reconstruct the trajectories of all characters, vehicles and the plane, as shown in Figure 2, and fine-tuned them to match to the existing views from the movie. Some movements could not be seen in the original shots and had to be crafted manually to plausibly match the visible parts.

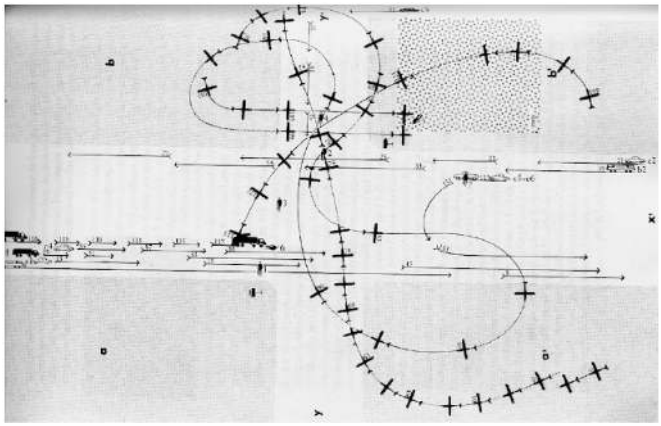

Figure 2. Floor plan views of the crop duster scene. Trajectories of all characters, vehicles and plane.

This work enabled us to build a 9 minutes and 45 seconds 3D animated scene corresponding to the action of the original film. We then imported this scene into the Unity game engine to implement the virtual reality set up.

Re-framing

In order to immerse the spectator into this reconstructed 3D scene, we had to find how to translate the frames of the movie to a spatial experience specific to virtual reality.

In room-scale VR, the perceptual boundaries of a continuous virtual reality experience are based on the subject's freedom of movement in the real environment. The technical equipment and the perimeter of the physical space tracked by the device therefore seems to correspond to a frame. In the virtual environment, it defines a scaled, spatialized and oriented zone in which the subject acquires a limited walking and acting area. We therefore decide to take this zone of potential action as our editing cell. The frame in virtual reality no longer delimits a portion of the image as in cinema, but a zone of possible actions in a space.

This space area constrains the movements that the immersant can perform. By changing the relative sizes of the virtual and the physical worlds, we can define "wide shots" with miniature non-player characters where the room appears to be 100 meters wide; "long shots" with dolly-size characters, where the room appears to be 10 meters wide; "medium shots" with three-quarter-size virtual characters; "point of view shots" with real-size virtual characters; and even "close shots" with larger-than-life characters where virtual movements are limited to a meter or less.

We started by transposing the shot grammar proposed by Hitchcock in our own "immersant's scale" classification. By testing a variety of scales of the immersant and his/her physical tracking space relative to the scale of the virtual environment, we experimentally found different size ratios that seem to reproduce the distancing effects of the cinematic frame. Moreover, the closer we get to the character's scale, the more we come to inhabit his space and the more we feel we belong to the diegesis. Finally, we analyzed that each scale brings a difference of amplitude of our bodily action in the virtual environment (a more or less large space is reachable) implying different agencies in the reception and therefore a new rhythm to find for the montage of these scale variations in space and time.

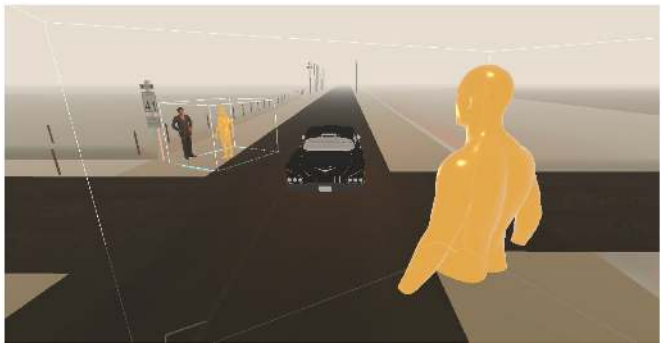

Figure 3. Two positioned and oriented scaled areas. The immersant's scale is represented by the yellow avatar.

We then adapted the position and orientation of the scaled area (Figure 3). In order to reduce the motion sickness issues, we imposed that the zones had to always be parallel to the ground. In other words, they could not be tilted in their x or z axis. Moreover, at the beginning of each "spatial shot", the y axis orientation of the imersant's gaze is imposed no matter what he/she was looking before the cut.

In a first step, we placed our oriented scaled areas regarding the camera positions of Hitchcoch's montage. Then, we decided to take some liberties and to re-cut the sequence regarding the medium shift of our device.

Re-cutting

The form of the cut that we propose consists in a change of scale and in the setting of the position and orientation of the immersant's point of view (Figure 4a). Each cut imposes an initial point of view ("initial gaze frame") to the immersant who is then free to choose his/her gaze movement until the next cut (Figure 4b). So the "final gaze frame" depends on the immersant.

In order to adapt the editing, we started by slowing it down. First of all, we have to take into account the necessarily longer processing time of stereoscopic images with a large field of view. Considering spatial breaks comes up with specific cognitive constraints related to the perception of virtual environments that implies spatial reasoning [10]. A too short time between two cuts can make the image unintelligible.

That's why we allowed a time for exploration, choice and habitability of the virtual environment. According to the scale and to the action which takes place, we evaluated the necessity of the montage. By moving his/her body and viewpoint, the immersant performs a form of editing that must be taken into account. For example, at the beginning of the sequence, Hitchcock uses shots of about three seconds alternating between a third-person view of the character and his subjective view of the

landscape he is looking at. Reproducing this montage while the immersant is free to move around does not work well. Instead, the rotation of his/her gaze must be taken into account to produce the same effect.

Figure 4a. An immersant experiencing a cut from the plane to a low-angle view of the character.

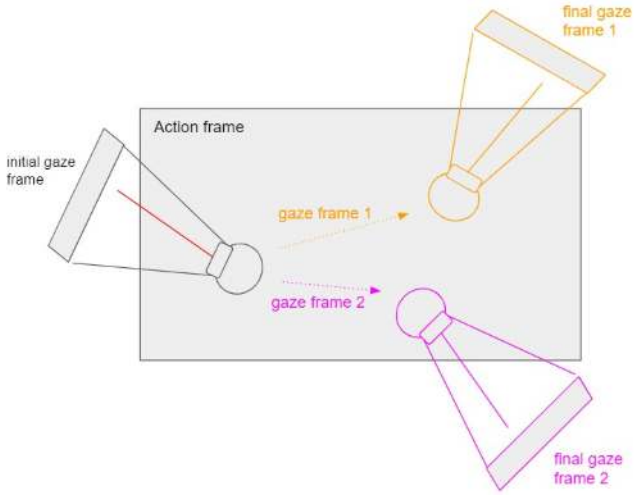

Figure 4b. Starting from an initial gaze frame imposed by a cut, different gaze movements can be actualized by the immersant until the next cut.

During this step, we studied empirically which cuts proposed by Hitchcock were replaced by the moving point of view of the immersant naturally positioning itself as intended by the film's editing. We thus observed that the cuts linked to the spatial understanding of the space had to be revised. In addition, imposing a point of view on the empty road systematically incited the subject to turn his/her head, which lost the desired dramatic result.

In order to have an effect, the shifts in viewpoint must be justified by a shift in distance or the desire to create a rhythmic change effect.

Then, we were confronted with the fact that the cuts can create a conflict between the immersant's point of view and the one imposed on him/her. These imposed discontinuities can be frustrating and/or seen as a punishment. Moreover, if it takes place during a head rotation, the cut can be disorienting. The two rhythms enter into a struggle. If the immersant lets go and stops moving, he submits to the rhythm of the editing accepting its tempo. For a better experience, we therefore study how to let the immersant create his/her part of the tempo.

Re-cutting with cues

We developed tools with the Unity game engine to enable a montage that adapts to the behavior of the immersant. It reacts to a system of "cues" produced by the human-computer system during the experience.

These cues can be represented by boolean variables computed in real time according to biometric data (gaze and body movements), geometric data (arrangement and orientation of objects in the virtual space) and temporal data (event system).

We use the following preliminary cues:

–

“Does not move”: the immersant is not moving faster than a certain value.

–

"Is looking at": the immersant is watching a certain character or vehicle.

–

"Object in frame": A certain character or vehicle enters or leaves the immersant’s field of view.

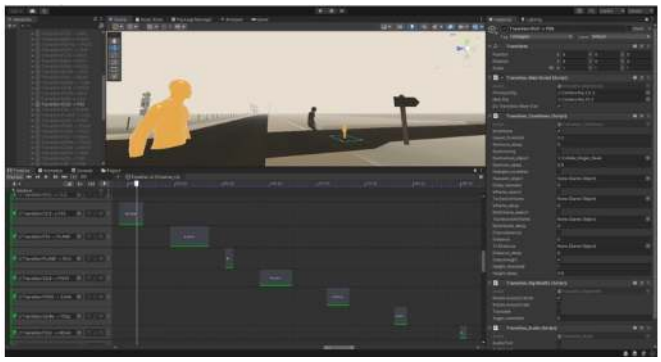

Figure 5. Interface in Unity. The editor can choose the transition cue (right) and set the transition objects on the timeline (bottom).

We define a type of computing object capable of making the transition between two previously defined spatial frames when the right conditions are verified (Figure 5, right). By setting these objects on a timeline (Figure 5, bottom), we can choose to activate and deactivate them at the desired time. This enables us to define a time range in which a transition can take place. Instead of having a cut defined at a precise moment, we have a delocalized transition that may occur within a certain duration. We call it: "cut porosity". The moment of the cut is then determined in real time according to the received cues. These tools allow an editor to assemble the experience through the timeline interface. The main difference with traditional editing is that the editor works on the duration of the transitions' possibilities rather than that of the rushes. This system also enables the creation of branches: it is possible to define several transitions that can take place depending on the cues encountered. This feature makes it possible to create alternative setups depending on the immersant's behavior.

Using this new system of cues and dynamic timeline, we created a new montage that adapts to the rhythm of the immersant. First of all, we mainly used the "Does not move" condition to make sure that the subject has stabilized his/her gaze before the cut (final gaze frame) to avoid disorienting him/her. We also often used the "Is looking at" condition to trigger the cut when the viewer's gaze is stabilized on the character. This allowed us to create on-axis cuts with scale changes to reduce our distance to the action. By using this condition, we were also able to control changes in focalization (from sympathy to empathy) depending on our interest in the character or not. Finally, the "Object in frame" condition was used when watching the plane. When the latter left the field of view, there was nothing to see but the sky, so it was useless to stay in that position.

To illustrate the importance of the cues system combined with the changes in scale, let's take the example of the first plane attack. In the film, this moment is edited with a shot/reverse shot between the crop duster and the character observing it until he understands at the last moment that he is the target. By alternating shots showing the plane getting closer and closer to the camera and shots showing the character in tighter and tighter frames, Hitchcock creates an attraction between the plane and the character until the two meet when Cary Grant jumps to the ground.

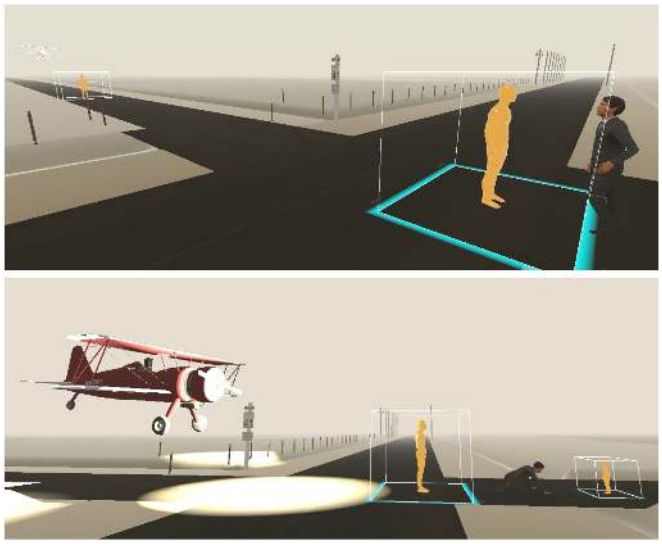

Figure 6. The immersant (represented by the yellow avatar) moves from one scale to another according to the cues set by the editor.

If we apply this same editing in VR, we must first take into account that the immersant might perform the field against field by him/herself. It was therefore necessary to have the "Look at plane" condition to switch to the character or the "Look at character" one to switch to the plane. It is indeed useless to show the plane if the immersant is already looking at it. It just creates an unpleasant jumpcut. Moreover, the condition "Does not move" avoids that the cut takes place when the immersant is turning his/her head and thus being punished or disoriented.

But, as the cues are limited in time, there is a moment when even if they are not reached, it is necessary to move to the next point of view. This is where the scale change becomes necessary to create a shift in distance (close-up effect) to the character, even though the immersant is looking at him before and after the cut. By downscaling from the scale of Cary Grant to the one of a mouse (Figure 6), proper editing effects are created when the character falls, even if the "Look at" or "Does not move" conditions are not satisfied.

Finally, it was necessary to work on the time given to the activation ranges of the conditions. The more the range extends in time, the less the editing can impose its rhythm. In this case, the cut may arrive at an ideal moment for the immersant but be misplaced and meaningless in terms of the narrative. We therefore had a tendency to shorten these areas to establish the rhythm of the montage, imposing at certain moments pace accelerations where the immersed body has almost no time to move between several cuts. When the character falls on the floor, it is important to have the cut at this precise moment. This was also particularly the case when the first truck passes, or during the final climax when the tanker truck arrives, hits the character and the plane crashes. It was therefore a matter of finding a balance between the editor's pleasure in imposing a rhythm that he/she liked and the frustration of leaving some freedom of rhythm to the immersant. We then had to evaluate this balance.

Experimental protocol

In order to conduct a qualitative analysis of this montage, we set up an experimental protocol that we tested on 17 participants. 12 of them were novices to virtual reality.

The immersion in virtual reality is made with the HTC Vive Pro Eye system. A Vive tracker is attached to the participant's plexus in order to record his/her body position and orientation data. In order to use and record the eye tracking data, a calibration of the participant's eyes was necessary. The participant was asked to move within a square of 2.80 m on each side. These boundaries were reproduced in the virtual environment. The participant is filmed. A video of his/her point of view is captured. The eye tracking data and the position and orientation of his/her head and chest are recorded.

The experimental protocol consists in living successively two versions of the montage of 9 minutes and 45 seconds. In the first one, the immersed person always remains at the same scale as the character and we impose to him/her the rhythm of the editing planned by Hitchcock and George Tomasini. Between two editing points, the immersant is able to move in space, but at each cut, he/she is positioned and oriented like the film shot. Whatever we do, the cuts thus bring us back into the authors' point of view at the moment chosen by them. In other words, if I don't move, I see the film as it was cut in the original film. This “Hitchcock” version (we called H) contains 133 cuts and the average time between two cuts is 4.4 seconds. The second experience consists of living the version of the montage edited by us with the changes of scale and the system of cues reacting to the immersant’s behavior. This “re-cutted” version (we called RC) includes 56 cuts and the average time between two cuts is 10 seconds.

The order of the two experiences is arbitrary and varies for each person. After having experienced the first one, the participant is invited to answer a questionnaire relating to the bodily and cognitive appreciations and feelings, the sensation of presence, the relationship to the narration and the editing. Then, the participant experiences the second version. The same questionnaire is therefore to be completed. Some questions, specific to this second questionnaire, concern a comparison between the two experiences.

Moreover, an oral interview of about twenty minutes is conducted allowing the participant to justify his/her answers to the questionnaires, to criticize the experience and to go into more detail about his/her sensations and impressions.

An authorization for the recording, use and distribution of images, videos, sounds and biometric data is signed by each participant.

Finally, we reconstructed in Unity the movements of the participants, enabling us to replay the experience in the 3D space. We can thus replay the experience either on a screen or immersed in virtual reality with a headset.

The protocol of the experiment is presented in this video: https://vimeo.com/779289940.

Questionnaires’ results

In this part, we present the first observations made from the answers to the questionnaires. These have helped to guide our reflection and to draw some conclusions about the effects experienced through the montage in virtual reality. For each question, the subject could choose a qualitative answer ranging from-2 to 2 (Figure 7).

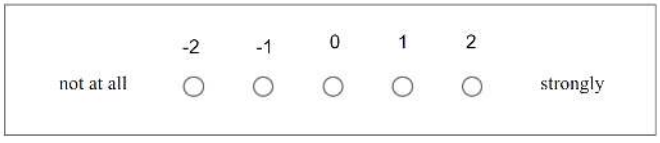

Figure 7. Format of the answers to the questionnaire. From “not at all” to “strongly”.

In terms of general feeling, our "re-cut" (RC) version of the montage was more appreciated than the direct transposition of Hitchcock's montage (H) in VR. To the question "I was absorbed by the experience", RC gets a better score (0.94) than H (0.53). To the question "I didn't feel confused or comfortable", H is less well lived (0.22) than RC (0.63) when experienced first. To the question "I had the sensation of being there in the virtual environment", the score is better for RC (0.56) than H (-0.13) when it is experienced in second. This indicates that the phenomenon of presence persists for RC even after the startling effect of the VR technological experience has faded. Finally, on the question "I liked this experience better than the previous one", RC scores higher (0.5) than H (0.0).

A part of the questionnaire also focused on the phenomenological relationship related to the mediation. To the question "Did you feel more immersed in images or spaces?", the subjects feel more in a space during the first experience (H or RC). This seems logical given that when they put on the headset for the first time, they strongly engage their spatial perception system. On the other hand, during the second experience, the immersants feel more immersed in images with H. To the question "I had the impression of sharing the same space as the character", H obtains a better average score on the two experiments (0.76) than RC (0.35) which seems logical given that in H, the immersant keeps the scale of the character. We should note that this impression is more marked for H if it is experienced before (1.0) than after (0.5) RC. This result is interesting to put in relation with the fact that the immersant seems to feel more in images in this last case.

Concerning proprioception, to the question "I felt like I didn't have a body", RC obtains a null score (0.0) whereas the one of H is negative (-0.47). This could be explained by the fact that in H, immersants don't have time between two cuts to conscientize their bodies in space. Moreover, the score increases significantly in the second experiment. This confirms the tendency of immersants to move from immersion in a space to images when they get used to the fragmented grammar of the experiences.

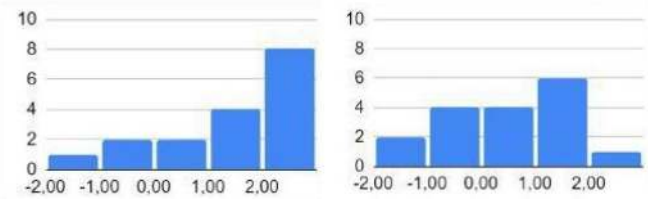

Figure 8. Answers to the question: "I was disoriented by the cuts". Left: H, average 0.94. Right: RC, average 0.0.

Regarding the relationship with the "cuts", RC seems to be better experienced. To the question "I was disoriented by the cuts", the difference is very clear between H (0.94) and RC (0.0) as shown in Figure 8. It is also interesting to note that the score becomes even negative when RC is experienced in second. We get the same tendency for the question "I was frustrated by the cuts" (0.35 for H and -0.47 for RC). Finally, on the question "I feel like I missed some important things in the story", RC gets a better score (-1.59) than H (-0.94).

Discussion

These experiments enabled us to refine an analysis of the aesthetic experience of the immersant. Based on the answers to the questionnaires and the interviews conducted with each participant, we were able to extract certain theoretical explorations.

The rhythm of the immersant: a dialectic between looking and inhabiting

When immersed in a virtual environment, the immersants must quickly understand their role. Is it an environment where they can interact explicitly? Are their presences taken into account by the narrative? Should they move or stay still, explore or contemplate, inhabit or watch? According to the answers to these issues, we observe that the immersants engage their perceiving body according to different dynamics.

The rhythm of the immersant is first influenced by his/her own characteristics. According to his/her emotional state, expectations, habits of virtual reality systems, he/she will engage his/her body differently. While some remain still, naturally observing the mediated spaces, others will be active in exploring the limits of the simulation in a more videogame-like approach. We can thus make the hypothesis that there are typologies of immersant to detect in order to adapt a rhythm of montage.

The contract that is created between the immersant and the artwork depends on the expectation of the mediation experienced. When dealing with the device, we observed that some participants believed that they must adapt to make the system work and get the most out of it, while others expected the system to adapt to what they wanted to do. In other words, the contract depended on the responsibility the immersants give themselves in the quality of their artwork reception experience. On this point, the answers to the questionnaires seem to show that the participants live

better the experience (RC) when they have the time to position their gaze and to control a minimum their reception.

In our device, the immersant is standing and knows that he/she can move at any moment, which creates a particular relation to the perceived image. We don't observe in the same way when we know that we can look at something else at any moment. The tension of Hitchcock's sequence, created by the rhythm between shots of the character and of what he sees, is changed into a tension between deciding to look at or not to. But if the editing systematically orients our gaze toward what we need to see to follow the story, the exploration becomes meaningless. In this way, the rhythm of the immersant depends on the staging. Does he/she have something to see? If yes, then why should he/she move? A working hypothesis is to decentralize the point of view regarding the salient narrative elements, inciting the immersant to seek a new point of view. The rhythm of the immersant could thus be influenced by the time he/she takes to switch from the exploration of the environment to the stabilization of his/her gaze on an image.

Then, according to his/her scale in relation to the representation space, the immersant needs more or less to engage important movements. A point of view of a giant densifies his/her distancing towards the elements of the scene. He/she does not need to move much to approach any element allowing him/her to embrace globally a large space to situate a scene. This "dollhouse view" engages a heterocentric referential of spatial perception, inducing less body interaction. Being so different from his/her size, the environment is on the one hand not inhabitable, but seems to increase the emotional distance to the narration.

Moreover, we notice that if the immersant is the same size as the character, he/she will naturally have a more exploratory attitude. Moreover, his/her presence in the representation space will be better induced and the behavioral responses (bending down when the plane passes or leaving the middle of the road when the truck arrives) will be stronger.

Finally, if the immersant is very small in proportion to the environment, the translations of the body have much less amplitude to explore the scene. On the other hand, since the whole world is larger, head rotations have much more impact on changing what the immersant sees in its field of view. In other words, the more we reduce the size of our virtual body relative to the environment, the more we are confined to a localized area but the more we have to turn our head to follow the narrative's salient elements.

Each scale brings a different relationship to our potential actions and to the more or less tangible nature of the representation space. They offer specific qualities of perception and engagement [4]. These relationships evolve over time and will be shaken up by the discontinuities of the montage.

The rhythm of the editing: between image rhythm and space rhythm

The rhythm of the editing gives a limited temporality to the action and perception characteristics of the immersant's scales. The spatial reference frame does not persist, which leaves only a limited time to put in relation the physical body and its virtual scale. According to the delay between each cut, the immersant is more or less conscious of the presence of his/her body in the representation space. If the montage is slow, then he/she can get used to the scale of his/her body, go from perceiving an image to inhabiting the space, exploring it, building a point of view and having the time to re-stabilize the gaze on an image.

If on the contrary the montage is fast, the image does not have time to become a space. A frenetic editing may not leave time for a behavioral response. We saw in the results that it also may affect the awareness of the perceiving body. We make the hypothesis that these images (not yet spaces) are printed differently for the perceptive system and modify the inner construction of the space. This could be confirmed by the fact that participants felt more immersed in images when they made H (133 cuts) after RC (57 cuts). The rhythm of the montage has thus a role in the transformation of the space into image or of the image into a space, and thus in the nature of the mediation.

Depending on the scale and initial point of view (initial gaze frame) imposed by the cut, the editing can also impose a movement by decentering the viewer in relation to the action. The proposed point of view would then be perceived more as a position in space than as an image. This spectatorial decentering should invite the immersant to move in a certain way towards the search for framing an image. Through editing and staging, the immersant can thus adopt a certain rhythm linked to the recentering of his point of view on the action. This has the effect of changing the nature of the cut, which here would be more of a collision than a continuity.

Matching the rhythms

Discontinuities in perceptual experience are not natural in the framework of mediation in virtual reality. There is a difficulty in repositioning ourselves in a sudden new virtual space [11] and a need to provide the immersants certain visual cues so that they can orient themselves [12]. We saw in the questionnaires' answers that in H, cuts were often frustrating, disorienting and meaningless. In interviews, participants often complain about the beginning of H when there are cuts every three seconds between the character and what he sees. We hypothesize that in order for them to be accepted, and especially to acquire a meaning, the cuts must be part of the acceptance of a cinematographic language of the experience. It would therefore be necessary that before and after the cut, the immersant shall be focused on an image rather than in the process of exploring. By detecting the moments when the immersant is looking for a point of view, we can avoid making cuts at such moments. When this exploration phase ends and the gaze stabilizes on a final gaze frame, the montage can then propose a change in response to the stabilized image.

Then, in order to match the rhythms, it also seems important to pay particular attention to the first contacts between the immersant and the montage. A montage too fast at the beginning can discourage the subject from moving. It seems important to allow time for exploration at the start of the experience and to gradually change the pace as the language becomes better established. The complexity lies in the fact that we must both intuitively invoke our culture of film editing while at the same time liberating our bodies in space. If we realize that it is the role of the editing to show us points of view, then our role is no longer to go and find them by ourselves. That is why it is important for the editor to take into account this progressive learning of the immersant to set up the cues and their durations. By creating this relation between cues and immersant, the role of the editor is here fundamental in the harmony of the rhythms and in the resulting aesthetic impressions.

Narrative effects

While the conducted experiments have raised new phenomenological questions, they have also opened up new directions of reflection in the theorization of narrative effects linked to montage in virtual reality. Firstly, thanks to the system of cues, when the cut is made between two stabilized points of view, we can observe the persistence of a "Koulechov Effect"6 in virtual reality. In other words, the creation of a narrative meaning extending beyond the two separate points of view. We might also explore the difference of this Koulechov Effect if the immersant passes from his/her point of view to the one of the montage or if he/she chains two points of view of the montage (if he/she doesn't move between two cuts).

A second track to explore is the multiplicity of points of view that the montage proposes or prevents. As the "framemaker" of his/her own experience, the immersant can decide to look at a character or at what the character is looking at. This engages different effects of identification (with the character or with his/herself) and testifies to his/her empathetic or sympathetic relationship towards the character. These choices vary the level of presence in the environment and the distance to the narrative. By imposing changes in scale and point of view, editing can in turn constrain some of these effects in time. Moreover, depending on these changes in relation to his/her attitude, the immersant may or may not feel that he/she is missing events.

We have seen that according to the scale, the immersant more or less inhabits the virtual environment. We hypothesize that with the size of Cary Grant, the immersant engages more behavioral responses to events (at the character's scale, it seems normal to feel more embodied and therefore in danger). On the contrary, in larger scales, the immersants feel more like a "framemaker-body" of their own experience. By framing the representative space, they don't change the story but the way they perceive it and thus their focus and aesthetic impressions of the story. The immersant would then be a "director-body", seeking to transform space into image. In order to be an "editor-body", he/she would have to make a conscious choice between different images in time. Perhaps the "editor-body" emerges when the control of spatio-temporality becomes conscious.

Finally, when setting up the cue system, we systematically asked ourselves the following question: is the interaction implicit or explicit? If the immersant realizes that his/her gaze actions have an influence on the editing rhythm, then the narrative experience changes. The montage may thus become embodied and open up new sensitive connections with the ongoing narrative.

Conclusion

In this project, we have explored how montage theory needs to be adapted to the new context of room-scale VR. We found that proprioceptive immersion in a virtual story world can indeed provide the necessary "cues" to

motivate shifts of points of view and scales that effectively guide the audience through a narrative experience. These meaningful discontinuities create a rhythm in the experience that the audience can follow and enjoy if they are correctly synchronized with his/her internal bodily rhythm.

Depending on the montage rhythm and its adaptability to the attitude of the immersant, we observe a dialectic between the exploratory nature of the space and the narrative nature of the image. Our experiment offered the opportunity to analyze how the dynamics of the montage may suggest to the immersant an exploratory rather than a spectator attitude, or vice versa. It constitutes a user study of how our spatial habits can harmonize with our visual habits of cinematographic images in order to make sensible the oscillation between constraints and liberties.

We made the hypothesis that the montage must also adapt to the rhythm imposed by the immersant. Thus, we explored how the montage can adapt in real time to the behavior of the immersant through implicit interactions (focus of the gaze, objects present or not in the field of vision, speed of displacement and rotation of the head), while respecting a meaningful narrative structure. The artistic and narrative impressions of the space intervals result then from a compromise between the proposals of the immersant and of the montage set by an editor and actualized by the software.

In future work, we would like to explore further this new form of relationship between the immersant (present and acting) and the author (absent) by introducing a richer repertoire of cues and actions, and creating even more variations in the resulting montage.

---

(1) *Person immersed in a virtual environment. Crop duster scene timecode: from 1:06:17 to 1:15:55.*

(2) *North by Northwest*, Alfred Hitchcock, Metro-Goldwyn-Mayer, 1959. https://www.imdb.com/title/tt0053125/

(3) https://remisagduv.com/the-hitchcock-experience-a-spatial-mont age-project/

(4) https://atap.google.com/spotlight-stories/

(5) https://team.inria.fr/anima/

References

1 Alfred Hitchcock, François Truffaut, Helen Scott, *Hitchcock/Truffaut,* Paris, Gallimard, 1993.

2 Gray Hodgkinson, “Cut, don’t cut-moving the viewer through a story in animated VR,” in *7th International Conference on Illustration and Animation*, 2019.

3 Scott Macaulay, “Look Into the Cut: Jessica Brillhart on Editing VR,” *Filmmaker Magazine*, October 28, 2015.

4 François Garnier, “Artistic Practices in Digital Space: An Art of the Geometries of Movement? (2021),” in *Space-Time Geometries for Motion and Perception in the Brain and the Arts*, ed. Tamar Flash and Alain Berthoz, Cham: Springer International Publishing, 2021, 237–50.

5 Vanessa Pope et al., “The geometry of storytelling: Theatrical use of space for 360-degree videos and virtual reality,” in *CHI Conference on Human Factors in Computing Systems*, 2017.

6 Sylvia Roth, et al., “Where to place the camera,” in *ACM Symposium on Virtual Reality Software and Technology*, 2019.

7 Sylvia Roth et al., “The impact of camera height in cinematic virtual reality,” in *ACM Symposium on Virtual Reality Software and Technology*, 2018.

8 Rémi Sagot-Duvauroux, François Garnier and Rémi Ronfard, “(Re-)Framing Virtual Reality,” in *WICED 2022 - 10th Workshop on Intelligent Cinematography and Editing*, April, 2022.

9 Raymond Bellour, *The Analysis of Film*, Indiana University Press, 2000.

10 Ogechi Nnadi, Ute Fischer, Michael Boyce, and Michael Nitsche, “Effect of Dynamic Camera Control on Spatial Reasoning in 3D Spaces,” in *ACM SIGGRAPH Symposium on Video Games - Sandbox Symposium*, Los Angeles, California, August, 2008, 9–10.

11 Michel-Ange Amorine, Ben Trumbore and Pema L. Chogyen, “Cognitive Repositioning inside a Desktop VE: The Constraints Introduced by First-versus Third-Person Imagery and Mental Representation Richness,” *Presence: Teleoperators and Virtual Environments* 9 (2), 2000, 165– 186.

12 Karl E. Steiner, Lavanya Voruganti, “A Comparison of Guidance Cues in Desktop Virtual Environments,” *Virtual Reality* 7 (3-4), 2004, 140–147.