Influence of Narrative Elements on User Behaviour in Photorealistic Social VR
Anonymous Author(s)

Figure 1: a) Living room where the starting position of each user is indicated. In details, 1 and 2 are HMD users while users in position 3 and 4 are desktop users. There are also two interactive objects (i.e., the light switch on the left and phone finder on the right) and the main virtual character, detective Sarge. b) A floor map of the virtual house is reported with the user heatmap of main locations visited over time.

ABSTRACT
Social Virtual Reality (VR) applications are becoming the next big revolution in the field of remote communication. Social VR provides the possibility for participants to explore and interact with a virtual environments and objects, feelings of a full sense of immersion, and being together. Understanding how user behaviour is influenced by the shared virtual space and its elements becomes the key to design and optimize novel immersive experiences that take into account the interaction between users and virtual objects. This paper presents a behavioural analysis of user navigation trajectories in a 6 degrees of freedom, social VR movie. We analysed 48 user trajectories from a photorealistic telepresence experiment, in which subjects experience watching a crime movie together in VR. We investigate how users are affected by salient agents (i.e., virtual characters) and by the narrative elements of the VR movie (i.e., dialogues versus interactive part). We complete our assessment by conducting a statistical analysis on the collected data. Results indicate that user behaviour is affected by different narrative and interactive elements. We present our observations, and we draw conclusions on future paths for social VR experiences.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ACM Reference Format:

1 INTRODUCTION
Immersive Virtual Reality (VR) applications are going through a rapid evolution of technology, getting integrated in daily-life devices such as smartphones and laptops. Therefore, it is possible to imagine that in the near future video calls will give users a completely different experience than now: people will be able to chat, walk together, interact with virtual objects and watch events such as concerts or movies together in a common virtual environment [15]. This is what social VR applications are promising to enable, becoming the next big revolution in the field of communications. VR applications are already going beyond the passive media consumption of traditional video allowing for a higher degree of immersiveness and interaction by placing the users at the centre of the action; a further step forward is attempted by social VR. This emerging remote communication tool is indeed aiming at
overstepping current remote communications through 2D screens, enabling instead virtual co-presence of more users within the same virtual environment and allowing body interactions similarly to face-to-face communication [7, 13, 14]. The new challenge is therefore to increase the realism in the virtual experience and interaction. Therefore, the key aspect that needs to be fully understood in order to advance in this technology is the user, the main director of the virtual experience.

Emerging Social VR platform, such as Facebook Horizon\footnote{https://www.oculus.com/facebook-horizon/} and Mozilla Hubs\footnote{https://hubs.mozilla.com}, have rapidly grown in popularity. In most of these applications, participants take part of 3D virtual spaces through a computer-generated and customised avatar, and they can interact with each other. Physical displacement and proxemic interactions in virtual environments have been analysed to investigate which social cues are the most influencing and therefore are needed to ensure presence and immersion [12, 26]. Moreover, many works have investigated the advantages to have a more realistic self-representation versus a 3D avatar providing a higher degree of immersion and presence in the VR experience [8, 11]. Therefore recently, a more natural self-representation of participants has been introduced, thanks to real-time acquisitions and reconstructions of point clouds by depth cameras [9]. Not to be neglected is the technological aspect of these new applications. Social interactions and photorealistic representations came with both computational and bandwidth overhead for transmission and rendering [20]. Understanding user behaviour represents the first step towards the building of real-time and realistic VR systems at large that can optimize delivery based on user-centred adaptation [23, 25].

In this paper we focus on a better understanding of how people interact with a virtual environment and other users within it. In details, we present a first attempt of behavioural analysis in a 6-Degrees-of-Freedom (DoF) social and interactive VR movie. We based our analysis on navigation trajectories collected in a novel type of interactivity crime movie: four users, either equipped by a head-mounted display (HMD) or desktop computer and a controller, were watching together a VR crime solving movie [6]. Figure 1 (a) show a snapshot of the living room in which the story mainly takes place, with its main virtual character. The challenges that we aim to address in our analysis are twofold. As first we need to consider both the new physical settings and locomotion functionalities given to users in a 6-DoF system. The user now not only can select the portion to be displayed by rotating the head as in a 3-DoF system but can also move inside the virtual environment changing the distance and perspective with the displayed content. The second aspect to consider are the new social and interactivity features of the application that brought an added level of dynamics. For instance, during the experience participants are asked to make simple tasks, such as to look for and to press a bottom; their action influences directly the narration of the story since the time to solve the task is not fixed. Therefore, we compare user behaviour in terms of spatial displacements in the virtual environment and viewing direction with respect to movie characters and other participant within the whole experience. In particular, we show how much narrative elements of the movie, such as virtual characters movements or request of interactions, influence user behaviour.

2 RELATED WORK

Depending on the locomotion functionalities and the type of content representation, navigation in immersive media environments can range from 3-DoF to 6-DoF. In the first case, users are placed at the center of a spherical content, and by changing their viewing angle, they are able to select the portion of the content they want to visualize at any given moment. In case of 6-DoF, users are additionally able to change their position in the 3D environment, which is now populated with volumetric objects that can be observed from any position and viewing angle. The 6-DoF scenario promises a much more natural interaction and exploration of the 3D scene than its 6-DoF counterpart, thus increasing the feeling of presence and immersiveness [2] and opening the gate to VR innovations. Incorporating information about user behaviour in user-centric systems is of paramount importance to optimize transmission and rendering [18, 25], highlighting the need to understand users’ behaviour in 3- and 6-DoF VR settings. In the first scenario, users trajectories has been intensely analysed by different tools: angular velocity, frequency of fixation, and mean exploration angles, saliency maps [4, 21] but also through clustering approach and information theory metrics [16, 17, 19]. In the past, preliminary study on locomotion and display technology were presented for CAVE environments [3, 24] and more recently 6-DoF trajectories have been analysed in terms of angular velocity [1, 22]. Despite the great potentials, the study of user behaviour has been overlooked focusing only on a general characterisation and none specific tool or procedure has been proposed for 6-DoF trajectories.

3 A SOCIAL VR MURDER MYSTERY MOVIE

We contextualize our experiment in a social VR setting, in which 4 users to experience an immersive movie, occasionally being called to participate and offering their input. The scenario allows for 6-DoF navigation within a photorealistic 3D environment, which is populated by three virtual characters. In the following, we describe the movie timeline, the setup used during experiments and finally, we show some general performance of the system.

3.1 Movie plot

The interactive and immersive VR movie used in the experiments is about a murder mystery investigation [6]. The victim is Elena Armova, who lived in a luxury apartment in central London. A total of 4 participants are selected to aid the investigations lead by detective Sarge Hoffsteler and his assistant, Rachel Tyrell. The investigation is split into 3 chapters, as depicted in Figure 2. In chapter 1, the 4 users are placed in the virtual living room of the victim Ms. Armova, adopting an initial fixed positions indicated in Figure 1 (a). Participants mainly listen to the victim interrogation, which is possible thank to a futuristic machine based on artificial intelligence. There are also two moments in which users are asked to interactively interact with objects in the scene:

- user 1 is asked to switch on the light,
- user 2 has to pick up a phone finder controller and press the bottom.
Therefore, we can distinguish two moments with a task for participants and three narrative moments where mainly virtual characters are talking and walking around the scene, respectively named task and narrative in Figure 2. At the end of chapter 1, users are split into two groups: user 1 and 3 are conducted in the virtual kitchen with detective Sarge, whereas user 2 and 4 are led to the virtual bedroom with Rachel and Ms. Armova. In both the room, participants are listening to narrative dialogues of virtual characters. At the end of chapter 2, the users are brought back together in the virtual living room for the final chapter, where detective Sarge describes how the murder has been solved.

3.2 Experimental Setup

A low-latency volumetric video delivery pipeline, based on point cloud representation, was used to place each participant into the virtual scene [9]. Each user was captured using 3 Kinect Azure devices, placed in a circle around them, 120° apart from each other. This allowed each participant to be captured from multiple angles, ensuring a photorealistic representation while they interacted with the scene.

Two devices were used to visualize the social VR experience: users 1 and 2 were equipped with Oculus Rift HMDs, complete with controllers, whereas users 3 and 4 could watch the scene through 50-inch monitors, and could navigate using gaming joysticks. For HMD users, teleportation was enabled in key locations of the scene, as physical locomotion was restricted due to the acquisition setup, whereas for desktop users, movement was enabled through the gaming joystick. Due to the configuration of the controllers, only HMD users were able to engage with the interactive elements in the scene.

A total of 48 participants was recruited for the experiment, resulting in 12 social VR sessions. The participants were between 21 and 56 years old (μ = 34.9, σ = 10.3). The gender distribution was balanced (23 males, 25 females). All of them were fluent in English, and had no motor or visual impairment.

During the sessions, the position and rotation of the camera objects associated with each users were recorded at 30 Hz. Moreover, timestamped data was logged in a machine readable format, after we determined that this log collection had no discernible influence on system performance. The logs also contained enough information to allow combining the data from the individual machines after the session with the correct timestamp information (with an uncertainty of network roundtrip time, which was good enough for the problem at hand). Logged information included number of decoded points, transmission delay, and framerate. From the logged data it was possible to compute, for each rendering device, the elapsed time between the encoding of a given frame of the point cloud representation and its rendering on the virtual scene. This information effectively provided the latency between encoding and rendering, for each device under use. The data was obtained at the granularity of one second, to avoid disrupting the performance of the system.

The observed latency in all sessions was generally lower than 1 second, and was remarkably smaller for self-representation. In particular, mean latency for self-representation was 0.1147 seconds (25 percentile: 0.079; median: 0.105; 75 percentile: 0.136), whereas for the rest of the cases, it amounted to 0.5526 seconds (25 percentile: 0.408; median: 0.538; 75 percentile: 0.689). On average, the point count for each frame was 86342 points (25 percentile: 82412; median: 90268; 75 percentile: 95727). Finally, the observed framerate for all the representations was on average 9.08 frames per second (25 percentile: 7.5; median: 8.9; 75 percentile: 10.6).

4 BEHAVIOURAL ANALYSIS

A general overview of the exploration behaviour of users is given in Figure 1(b), which shows a heatmap of the most visited locations in the virtual house, obtained by aggregating all the position data collected in the experiment. As described in Section 3.1, large part of the movie takes place in the living room (Chapter 1 and 3), which is
Whereas movement is more spread in the living room, in the other
widder distributions, indicating larger variance in the way users
was then computed with respect to the total number of frames. It
users tend to further explore the environment around them: even
tive and interesting to notice the different user’s behaviour between
opposite direction (looking towards direction of character location (\(\theta\)) or in in the
user’s viewing direction with the respect to other participants. As in the previous analysis,
more spatially focused, indicating that participants were more static
in these spaces.

Figure 3 displays the boxplot comparison between percentage of
motion exhibited by each user. For every frame, we considered the
user to be “in motion” if either their relative position with respect
to the previous frame changed more than 0.05 cm, or if any of
their rotation angles varied by more than 0.01 rad (0.573\(^{\circ}\)). Both
measures were taken into account to cover both spatial exploration
behavior, as well as changes in viewing angles. The percentage
was then computed with respect to the total number of frames. It
can be observed that desktop users (user 3 and 4) exhibit a larger
percentage of motion over the course of the movie, with respect to
HMD users. Motion was present in the first and last chapter with
wider distributions, indicating larger variance in the way users
behaved, whereas in the second chapter, users generally showed
smaller variance in percentage of motion.

In order to better understand how users behaviour changed over
time as the movie progressed, we analyse their position and orien-
tation with respect to the virtual characters (avatars), namely Sarge,
Rachel and Armova. To give an idea about avatar displacements
and actions in the movie, Figure 2 shows distance covered by each
carer over time. We can notice that during task parts, avatars
are statically waiting for users to make the action. The characters of
Armova and Rachel are leaving the scene of chapter 1 before
Sarge, and indeed around the 5-minute mark, there are two peaks
in their displacement. It is worth noting that Armoa is in the scene
only in chapter 1 and 2. Figure 4 shows the distribution of relative
position and orientation of the users with respect to the avatars,
separately per each chapter of the movie. In details, the first line
of each subplot depicts the distance between each user and each
avatar. The second line of subplots shows instead the angle between
user’s viewing direction and the vector which connect user and
avatar at any given time. This angle indicates therefore if user is
looking towards direction of character location (\(\theta \rightarrow 0\)) or in in the
opposite direction (\(\theta \rightarrow \pi\)). In chapter 1 (Figure 4 (a)), it is therefore
interesting to notice the different user’s behaviour between narra-
tive and task parts of the story. During the three narrative parts,
users tend to further explore the environment around them: even
if on average, users are looking in the direction of the three virtual
characters, their variance indicates a non-uniform behaviour over
time, suggesting that participants where also looking around.

In the two task parts, instead, the distribution of both spatial distance
and angle values is narrower. In particular, for the first task, a wider
distribution in terms of viewing angle, which skews further away
from the avatar, can be observed for user 1, which is the one tasked
with pressing the button to turn on the light. This difference might
be due to the mismatching of difficulty between the two task. In-
deed, participants took on average around 13.33 seconds to switch
on the light, while picking up the phone finder controller and press
the button required around 25.58 seconds. To validate our intu-
tion, we perform a non-parametric Mann-Whitney statistical test
between the distance and angle to avatar recorded during the nar-
rative parts of chapter 1, and the task parts. To avoid bias induced
by the large number of samples, we performed random sampling
on the data, selecting N = 200 samples across the distance vector,
and repeating the procedure across 200 sampling runs. We used
Fisher’s method [5, 10] to combine the probabilities, obtaining that
the type of task has a significant effect on the distance (\(\chi^2 = 3202.5,
p < .001\)) and on the viewing angle (\(\chi^2 = 4771.8, p < .001\)). In
fact, distance to avatar appears to be statistically different between
all sub-parts of chapter 1, indicating varying behavior in terms of
spatial movements between as the time progressed. In terms of
viewing angle, however, no discernible effect is observed on differ-
tent narrative parts with respect to the viewing angle (narr.1 - narr.2:
\(\chi^2 = 407.5, p = 0.387;\) narr.1 - narr.3: \(\chi^2 = 407.5, p = 0.388;\) narr.2
- narr.3: \(\chi^2 = 421.5, p = 0.221\)), whereas the two tasks exhibited
significantly different viewing angle distributions (\(\chi^2 = 1270.6,\n\ p < .001\)). In chapter 2 (Figure 4 (b)) participants are moved in
different rooms, kitchen and bedroom, both of them smaller com-
pare to the initial living room. This different ambient dimension
affects indeed user’s behaviour: participants in general are much
more static compare to the previous chapter. As last observation,
we notice that users in chapter 3 (Figure 4 (c)) behave similar to
narrative moments of chapter 1: there are exploration movements
both in terms of spatial and angle values.

Finally, we also analyse the user’s position and viewing direction
with the respect to other participants. As in the previous analysis,
Figure 5 depicts the spatial distance and angular difference between
each couple of users per each chapter of the movie. While in the
previous comparison between user and avatars, the behaviour of
the latter was known and stayed constant for each experiment,
in this case both users under exam have varying positions and
viewing directions over time. However, some general behaviour can
be extracted also under these conditions confirming the previous
Influence of Narrative Elements on User Behaviour in Photorealistic Social VR

Figure 4: Spatial distance (first line of each subplot) and angle (second line of each subplot) between users and avatars per each chapter of the storytelling.

findings. For instance, during task parts in chapter 1 (Figure 5 (a)), the distribution of angle values is quite narrow for most of the couples, indicating that they moved their attention from the avatars to another participant. On the contrary, during narrative parts, these distributions are wider highlighting that users were not fixating on each other, rather exploring the scene and the virtual characters within. Statistical tests show that the effect of the task is significant on the viewing angle ($\chi^2 = 2465.6, p < .001$), and differences among narrative and task sub-parts of chapter 1 are always significant ($p < .001$), with the exception of the first and the last part of chapter 1 ($\chi^2 = 415.9, p = 0.282$). In terms of spatial displacements, the distance between users remains low in all the chapters of the movie. The effect of task versus narrative is significant ($\chi^2 = 1270.4, p < .001$), and distance always differs significantly between sub-parts ($p < .001$), with the exception of the first two narrative parts ($\chi^2 = 418.8, p = 0.249$).

In summary, the behavioural analysis we conducted of users in a social VR movie leads to the following observations:

- **Observation 1:** during narrative moments of the story, participants are more inclined to explore the virtual environment with general attention to virtual characters;
- **Observation 2:** the request of interactions with the content by a specific user (e.g., to press a button) leads to reduced movement, while the attention is more focused on the task or on other participants;
- **Observation 3:** the size of the virtual environment in which is located the experience also affect the user’s behaviour. In particular, large rooms seem to be more conductive of exploratory behavior, whereas in smaller rooms, less variation in position or viewing angle is observed.

5 CONCLUSION

In this paper, we analysed the user’s behaviour during a photorealistic telepresence experiment developed on a volumetric social VR system. We mainly investigated the influence of narrative elements of the story, such as dialogues or interactive task, on participants’ movements. Our results show indeed that the motion during the VR experience was affected by the storytelling. More static and focused behaviour happened when a task (either to switch on the light or press a button in a controller) was requested to be done by a specific participant. On the contrary, exploration movements were more frequent when virtual characters were talking in the scene. These observations are key factors to be further investigated, in order to develop immersive and interactive applications. Further work is needed to understand the effect of storytelling elements and cues on user behaviour, in order to design social VR experiences that can effectively be optimized around the users.

REFERENCES

Figure 5: Spatial distance (first line of each subplot) and angle (second line of each subplot) between couples of users per each chapter of the storytelling.