1161: MULTIMEDIA ALTERNATE REALITIES

User attention and behaviour in virtual reality art encounter

Mu Mu[1] · Murtada Dohan[1] · Alison Goodyear[2] · Gary Hill[3] · Cleyon Johns[1] · Andreas Mauthe[4]

Abstract

With the proliferation of consumer virtual reality (VR) headsets and creative tools, content creators are experimenting with new forms of interactive audience experience using immersive media. Understanding user attention and behaviours in virtual environment can greatly inform the creative processes in VR. We developed an abstract VR painting and an experimentation system to study audience art encounters through eye gaze and movement tracking. The data from a user experiment with 35 participants reveal a range of user activity patterns in art exploration. Deep learning models are used to study the connections between the behavioural data and the audience's background. The work also introduced new integrated methods to visualise user attention for content creators.

Keywords Virtual reality · VR abstract painting · User attention · Eye-tracking · Machine learning

✉ Mu Mu
  mu.mu@northampton.ac.uk

  Murtada Dohan
  murtada.dohan@northampton.ac.uk

  Alison Goodyear
  alisongoodyear@me.com

  Gary Hill
  gary.hill@cranfield.ac.uk

  Cleyon Johns
  cleyon.johns@northampton.ac.uk

  Andreas Mauthe
  mauthe@uni-koblenz.de

[1] University of Northampton, University Drive, Northampton, NN1 5PH, UK

[2] Alison Goodyear Artist, Bedford, MK40 1QE, UK

[3] Cranfield University, College Road, Cranfield, MK43 0AL, UK

[4] Universitat Koblenz-Landau, Universitätsstraße 1, 56070 Koblenz, Germany

🖄 Springer

1 Introduction

With the proliferation of consumer virtual reality (VR) headsets and readily available creative tools, it is evident that more content creators and artists are experimenting with alternate interactive audience experiences using immersive media [4, 11]. Creating content using new media brings new opportunities, which historically, is an approach long-used in the field of fine art. With VR, visual art artists can enjoy "unrestricted" space for creative work while exploiting features that are not possible in the physical world [13]. However, employing new media and alternate reality in fine art requires new understanding of how audiences interact with and experience new forms of content to inform a creative process. For the visual arts, this means establishing a better understanding of how viewers approach an artwork and navigate through a virtual environment. The research on user attention in VR encounter can lead to new developments within art practice where user activity data and computer programming can serve as a dynamic tool for innovation [6].

In order to further explore the topic of content creation in alternate reality, a group of computer science researchers, 3D game designers, and a fine art artist collaborated on a VR art project. Goodyear is experienced in using Google Tilt Brush to create experimental abstract VR paintings. An experimentation system is developed using Unity game engine in order to install any artwork and enable user navigation. With a combination of hardware and software functions, the system is equipped with multi-modal eye gaze and body movement tracking capabilities to capture user interactions with the artwork. A user experiment was conducted, where 35 participants explored an abstract VR painting according to their own preference. The participants' eye-gaze, body movements, and voice comments were recorded while they freely moved within a 3 by 4 metre physical space. The majority of participants approached the virtual works with confidence, to become immersed within the artwork. Many participants claimed that the new form of media encouraged an interest in not only these abstract VR paintings, but also abstract art in general. The quantitative experimental data show distinctive activity patterns which reflect different preferences in moving locations, body positions, and viewing angles. In order to model and predict user preferences, deep-learning techniques are employed to study the connections between users behavioural data and the audience background/previous experience. To enhance the feedback loop from user experiments to content creators, new ways to visualise user attention in virtual environments are also piloted. The results can be particularly useful for artists who are looking to adapt their art practices based on audience experiences and for the creative industry to develop new applications. It is felt that a tracking and data analysis engine, paired with game development engines have the possibility to become a new tool to augment art creation process, echoing how AI assists engineering and computer science designs.

Main contributions of this paper include:

– A system that supports systematic studies of interactive VR artwork.
– An experiment with 35 participants to study of how people interact with an abstract VR painting using eye-gaze and body movement tracking.
– New integrated methods for visualising user attention in 3D virtual environments for VR content creators to evaluate how the content has been perceived.
– The use of deep learning models to analyse multi-dimensional user behavioural data to support future development of personalised encounters.

The remainder of this paper is organised as follows. Section 2 discusses the background and related work in VR art and behavioural tracking in VR. Section 3 introduces the authors'

VR artwork, experimentation system, and the user experiment. Data analysis and modelling are discussed in Sections 5.1 and 5. Section 6 explores new methods for visualising user attention, whilst Section 7 includes discussions. Section 8 concludes the paper.

2 Related work

2.1 Virtual reality art and exhibitions

There is an increasing adoption of alternate reality platforms by content creators and visual artists worldwide. Blortasia is an abstract art world in the sky where viewers fly freely through a surreal maze of evolving sculptures [27]. Authors believe the exploration through art and nature reduces stress, anxiety and inflammation, and has positive effects on attitude, behaviour, and well-being. Hayes, et al. created a virtual replication of an actual art museum with features such as gaze-based main menu interaction, hotspot interaction, and zooming/movement in a 360 degree space. Authors suggested that allowing viewers to look around as they please and focus their attention on the interaction happening between the artwork and the room is something that can't be replicated easily [16]. In [2], Battisti, et al. presented a framework for a virtual museum based on the use of HTC VIVE. The system allows moving in the virtual space via controllers as well as walking. A subjective experiment showed that VR, when used in a cultural heritage scenario, requires that the system should be designed and implemented by relying on multi-disciplinary competences such as arts and computer science.

Many classical and iconic paintings have been reimagined and recreated in VR which offer audience a completely new perspective of the original artworks. James R. Eads recreated *The Great Wave of Kanagawa* by Katsushika Hokusai using Google Tilt Brush where audience can experience the power of sea waves while virtually sitting in the boat [7]. Salvador Dalí's 1935 painting *Archaeological Reminiscence of Millet's 'Angelus'* was recreated in VR as part of *Dreams of Dalí* exhibition [30]. Visitors can "immerse in the world of the Surrealist master like never before, venturing into the towers, peering from them to distant lands and discovering surprises around every corner". Johannes Vermeer's *Girl with the Pearl Earring* was brought to life by Eric Lynx Lin who animated the painting with eye blinking and subtle mouth moving [24].

Hammady et al. developed MuseumEye for the Egyptian museum to create an alternative tour guide system using Mixed Reality (MR) based on location and other environmental information [14]. In a smart museum experiment, Hashemi et al. developed behavioural user models based on a deep neural multilayer perceptron (MLP) for location-aware point of interest (POI) recommendations [15]. The model takes into account user information interaction behaviours in the physical space. A location-based VR museum was developed to enable 100 users to explore VR exhibitions simultaneously in a shared physical space with tracking sensors [26]. Zhou conducted a systematic, theoretical approach to study the issues involved in VR Art Museum and inform art museum professionals in their decision making process [45]. The discussions revolved around accessibility, enrichment, interaction, presence, and content. Parker et al, reported on a project designed to address some concerns in adopting VR at Anise Gallery in London with focused on multi-sensory. The research explored how the inclusion of VR might alter the practice of people watching and whether the incorporation of VR might produce a qualitatively different experience of the art museum as a shared social space [33]. In [37] Raz observed how VR is increasingly adopted by diverse artists and attains growing recognition at film festivals and argues that

VR is endowed with immersive affordance, which qualitatively differ from those of any other art media.

In February 2020 we hosted *Paint Park* [9] an installation of immersive virtual and physical paintings which blurs the lines and boundaries of traditional mark-making through the immersive qualities of virtual reality abstract painting. Visitors were invited to put on VR headsets and explore the VR abstract painting *Topsy Turvy*. The painting, which was created using Google Tilt Brush includes two parts. *Topsy* contains conventional style brush strokes such as Oil Paint and Wet Paint, while *Turvey* includes special effect brush strokes such as Neon Pulse, Chromatic Wave and Electricity.

The recent COVID-19 outbreak has greatly impacted the art and cultural sector in many parts of the world. Many galleries and museums are closed indefinitely while exhibitions and auctions are postponed or cancelled. This triggered a new wave to explore alternative digital spaces with online and VR exhibitions [42]. Now physical spaces are no longer the priority, the cultural sector is rushing to adapt events, exhibitions and experiences for an entirely digital-first audience. In America, Art Institute of Chicago and Smithsonian are among institutions that have embraced VR and taken on new significance as the lock-down deepens [12].

2.2 Eye gaze and behaviour tracking

One of the first use cases for eye gaze tracking in VR was Dynamic Foveated Rendering (DFR), which allowed VR applications to prioritise resource usage on the foveal region where a user is looking [44]. This helped reducing the volume of heavy rendering for complex scenes [32], enhancing the image quality or improving the frame rate [44]. Chen et al. used infrared sensors and cameras to detect and robustly reconstruct 3D facial expressions and eye gaze in real time to enhance bidirectional immersive VR communication. The research was based on the argument that users should be able to interact with the virtual world using facial expressions and eye gaze, in addition to traditional means of interaction [5].

In [34] Pfeil, et al. studied humans eye-head coordination in VR compared to physical reality. The research showed that users statistically move their heads more often when viewing stimuli in VR. Nonverbal cues, especially eye gaze, plays an important role in our daily communication as an indicator of interest and as a method to convey information to another party [21]. Kevin et al. presented a simulation of human eye gaze in VR to improve immersion of interaction between user and virtual agent. A gaze aware virtual agent was capable of reacting towards the player's gaze to simulate real human-to-human communication in VR environment [21]. Marwecki, et al. presented Mise-Unseen [28], a software system that applied scene changes covertly inside the user's field of view. Gaze tracking has also been used to create models of user attention, intention, and spatial memory to determine the best strategies for making changes in scenes in VR applications [28].

Embedded eye trackers were used to enable better teacher-guided VR applications since eye tracking could provide insights into student's activities and behaviour patterns [36]. The work presented several techniques to visualise eye-gaze data of the students to help a teacher gauge student attention level. Pfeuffer et al. investigated body motion as behavioural biometrics for virtual reality to identify a user in the context of authentication or to adapt the VR environment to users' preferences. Authors carried out a user study where participants perform controlled VR tasks including pointing, grabbing, walking, and typing while the system monitoring their head, hand, and eye motion data. Classification methods were used to associate behaviour data to users [35]. Furthermore, avatars are commonly used to

represent attendees in social VR applications. Body tracking has been used to animate the motions of the avatar based on the body movements of human controllers [3]. In [25] full visuomotor synchrony is achieved using wearable trackers to study implicit gender bias and embodiment in VR. The gender-based eye movement differences in indoor picture viewing was studied using machine learning classification in [38]. Authors discovered that females have a more extensive search whereas males have more local viewing.

3 VR art experimentation system and user experiment

The main purpose of the research described in this paper is: 1) to develop a VR system that offers the capabilities to gather and model user interaction data for VR visual artists and HCI designers, 2) to study how audiences interact with an abstract VR painting. This includes how they use free walk to move to different parts of the artwork or to change their viewing perspective, how they split their attention between thousands of brushstrokes, whether they walk into the artwork to have a more immersive experience, whether they attempt to touch the virtual painting with their hands, and how they describe their experiences. 3) to visualise user interactions in the VR painting as a means to provide user experience feedback to content creators. 4) to model eye-gaze and movement data to enable a better understanding of human attention and personal experiences in VR. Although many findings from our work will be specific to one VR painting, the developed system can accommodate other VR paintings for purpose of HCI research and user experience improvement.

The immersion techniques used for visitors to explore the artwork include viewing through head-mounted displays, changing of view port by head movements and changing of location by free walking within a physical space. The exploration mechanisms avoid any additional cognitive load of operating hand controllers, so participants can fully immerse in the virtual environment.

Figure 1 depicts the system tailored to support the research. The artwork is created in Google Tilt Brush and ported to Unity, a game engine, using an in-house conversion tool. Then an art exploration application was created in the format of a VR game allowing user
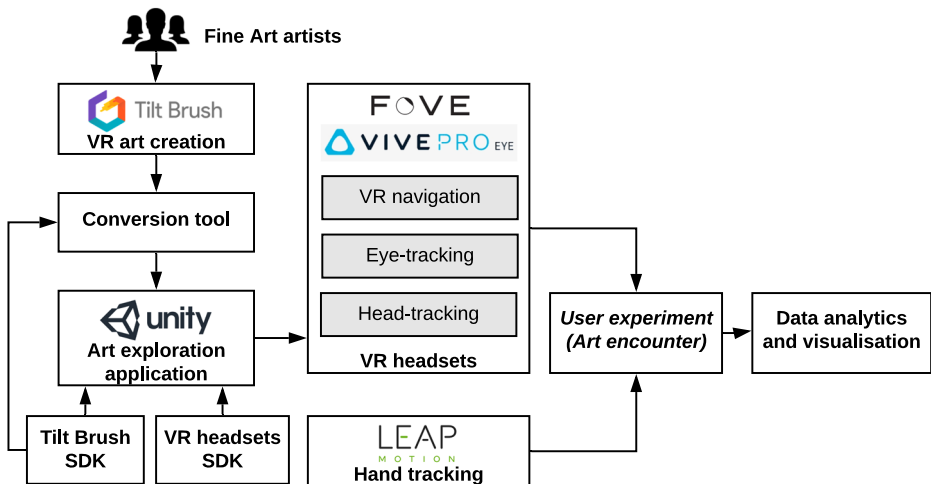

Fig. 1 Experimentation system diagram

navigation and activity tracking. VIVE Pro Eye [19] was chosen as the reference VR headset for the user experiment while the system also supports FOVE 0 [8]. VIVE Pro Eye has built-in eye-tracking capabilities (Tobi-based) and head position tracking using base stations. An additional real-time application was developed to extract, filter and store eye gaze and behavioural data in a database. Hand tracking was also incorporated by attaching a Leap Motion onto the VR headset though the analysis of hand movement data is outside the scope of this paper.
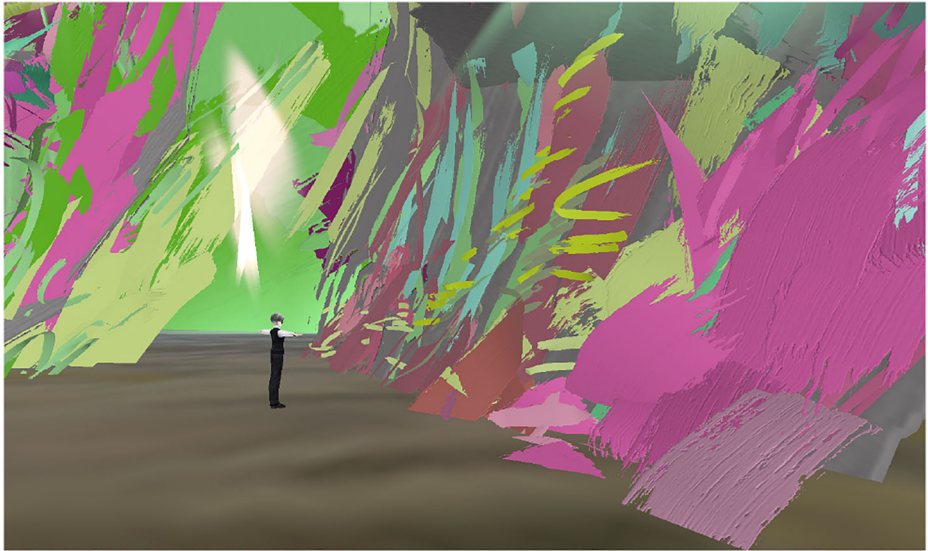
3.1 Abstract VR painting

The abstract VR painting *Caverna Coelo* (Fig. 2) used for this research was created by author and artist Goodyear who was developing skills in the new medium as an abstract VR painter. This work draws from Goodyears more-established physical painting process, which is principally a reflection on the painting process i.e. painting about the process of painting (that is the making and the thinking behind it) [10]. To achieve this Goodyear used documentation of the painting process within the painting and in most cases this involved using their paper-based paint palettes. This process was continued in Google Tilt Brush, where an image was imported into the paint palette followed by extrapolation from it. As seen in this work, Goodyear experimented with using physical palette imagery as a type of floor for the artworks and also as a backdrop. These works are described as "phygital", where the physical and digital were combined. This term, around since the turn of the century, was originally found in marketing to describe how customers in banking would interact with online platforms. Now it has been appropriated into many different fields, including the art world.

 Caverna Coelo uses both physical and digital painting elements and contains areas of concentrated painting marks and hidden pockets of almost isolated space, which Goodyear describes as painting caves. When navigating deep into the work through the layers within the work, these spaces are revealed, often unexpectedly, to the viewer. This approach, as the following sections will go on to discuss, shows promising signs of encouraging a virtual exploration.

3.2 Exploration of an artwork using a game engine based application

The artwork was ported into the Unity game engine to enable customised user navigation and behavioural tracking. While Tilt Brush allows VR creations to be exported directly in various formats such as FBX, all brushstrokes of the same type such as *Wet Paint* are grouped as one mesh object (even when the brushes are not connected) to make the handling of artwork more efficient. This means the system would only be able to tell whether a certain type of brushstrokes was looked at. As the authors were interested in user attention at individual brushstroke level, a conversion tool has been developed based on Tilt Brush Unity Toolkit [43] to retain brushstrokes as independent mesh objects using a unique identifier with the naming scheme: *brushtype_startingcolour_artworkref_seq*. For instance, an oil paint brushstroke from the VR painting *peacock* with a starting colour 4278853398 (RGBA) and a unique ID of 5251 would be identified as *OilPaint_4278853398_peacock_5251*. The conversion tool can export the artwork as a single FBX file with brushstrokes separated internally or as separate FBX files while each holds one brushstroke. The tool also outputs the metadata of all brushstrokes as a JSON file for data analysis purposes.

 The artwork was then correctly scaled in Unity by the artist and the elevation of the camera was set to mimic the artist's view when the painting was created. Various environmental

(a) View 1

(b) View 2

Fig. 2 Abstract VR painting (with an added human-size virtual character to indicate scale)

configurations such as lighting were made to match how the painting was felt in Tilt Brush. Due to the scale and complexity of the artwork, occlusion culling was use to maximise the scene rendering performance and maintain a high frame-rate for the improved viewing experience. For this experiment, participants were able to explore the artwork by walking freely within a set physical area. Therefore the navigation was enabled by mapping any physical

movements (walking, crouching, leaning, head turning, etc.) to camera positions and movements in Unity like a first-person computer game. The settings also allowed participants to walk through brushstrokes.

3.3 Eye gaze and behaviour tacking

The VIVE Pro Eye is an advanced VR headset equipped with an embedded eye tracking sub-system. It is built to meet the requirements of the most perceptive commercial clients and academic researchers since it provides research-grade data. VIVE Pro Eye comes with a compatible SDK that supports various game engines as well as data stream for research purposes.

Leap motion was added to the headset in order to extend the tracking features. The SDKs of both devices work synchronously within our Unity viewing application. The idea behind this was to introduce a unique tracker approach that synchronises eye, head, body, and hand movements at the same time. Leap motion was used to provide data of hand movements and to simulate these movements in the game thus increasing content reliability from a player's perspective. The system also tracked eye blinking events to validate or bridge eye gaze activities. This was particularly important to identify long gaze events, which can be wrongly segmented due to eye blinking.

VIVE Pro Eye SDK provides two types of eye tracking data, local and world space. Local space data provides eye orientation relative to the headset and is not associated with head orientation. World space data was used to take head orientation into consideration when computing eye orientation. World space data provided the raw data (eye_x, eye_y, eye_z) as a combination of both left and right eye coordinates on a unit sphere. Any tracking data marked as "invalid" by the SDK is discarded.

The data captured via the VR headset included both head position and eye direction, as it provided coordinates of both head position (head_x, head_y, head_z) and eye orientation (eye_x, eye_y, eye_z). Head coordinates are correlated to real world measurements. Meaning that, head_x, head_z coordinates refer to width and depth of the inside walk zone whereas, head_y refers to the height of the head. While, eye origin refers to a point in the world space where the eye was looking at and its coordinates help to generate the direction vector in the game that indicated the final position of what users have seen. In addition, Unity provided the player position along with head rotation that represented the direction of the head in the field using four dimensions (rotation_x, rotation_y, rotation_z, rotation_w) named quaternion. Player position (player_x, player_y, player_z) represented the location of the participants within the virtual environment. Finally, a timestamp was generated for each data frame to enable sequencing and validation of the raw data.

3.4 User experiment

The user experiment was organised in the Learning Hub, the largest academic building at the University of Northampton's Waterside Campus. The building was designed and used based on the typology of space rather than any notion of academic faculty structures. During term time thousands of students and staff from all subject areas visit the Learning Hub every weekday for lectures, seminars, social gatherings, IT support, library services and refreshments. This makes Learning Hub an ideal location to recruit participants with different backgrounds and experiences. An open communal space near the main entrance of the Learning Hub was reserved between 9am and 5pm for the experiment. Figures 3 and 4
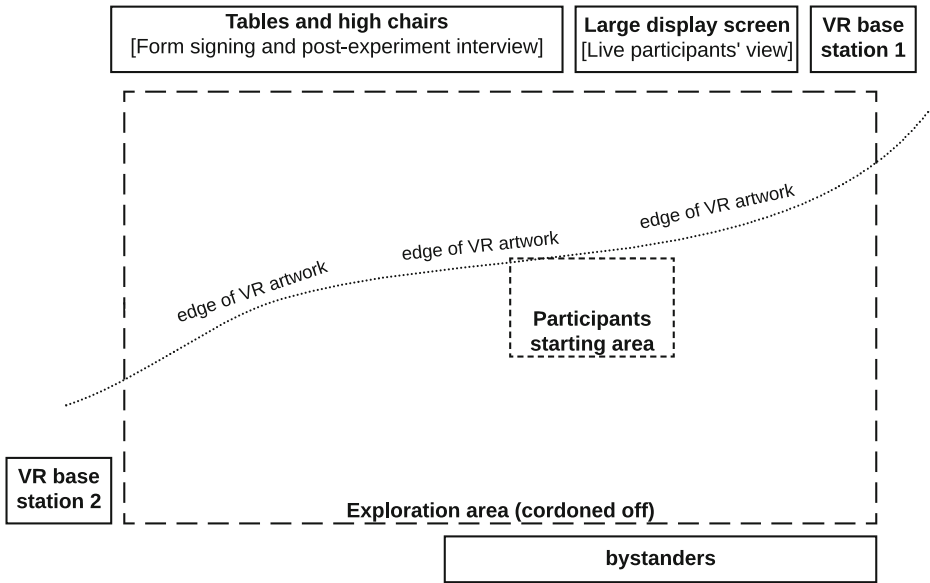
Fig. 3 Floor map

show the arrangement of the floor space (photo taken during the experiment). After an hour of setup and testing, the experiment started at 10am and ran continuously by a team of researchers and volunteers until 4pm.

Fig. 4 Participants exploring the abstract VR painting

The floor area was partially cordoned off (approximately 3 by 4 meters) to enable free exploration within this space. Each participant entered the area from the right and sat by the table to fill out research consent and user information forms. Participants were asked to answer questions on their gender, age group, game experience, knowledge of abstract art, etc. The participants were then assisted to put on a VR headset in the starting area, which is positioned near the edge of the artwork in VR. Once ready, the participants decided how to approach the artwork. An assistant helped with the cables running from the headset and ensured that the participants did not travel beyond the physical exploration area or walk into any obstacles. Any verbal comments were also recorded by the microphone on the headset.

As the participant encountered the painting their view was displayed on a large screen. This attracted more volunteer participants and allowed the assistants to communicate with the participants about what they were seeing. A short post-viewing interview was also conducted by the artist of the VR painting in order to study how the participants felt about the VR painting encounter, how it compared to physical paintings, their opinions of interactions in VR, and the most memorable elements.

Overall, the experiment attracted 35 participants, 20 female and 15 male (Fig. 5). The user information shows that the vast majority of the participants are aged between 16 and 25. More than half of the participants stated that they do not play or rarely play computer games (MD - Many times every day, OD - Once a day, OW - Once a week, RL - Rarely, NA - Not at all). Regarding their experience with VR, 15 had not tried VR before, while 18 had some experience. Only 2 participants claimed to be very experienced with VR. Similarly, only 3 participants, who studied Fine Art, had extensive knowledge of abstract painting while 18 participants were familiar with this form of artwork (Fig. 5).
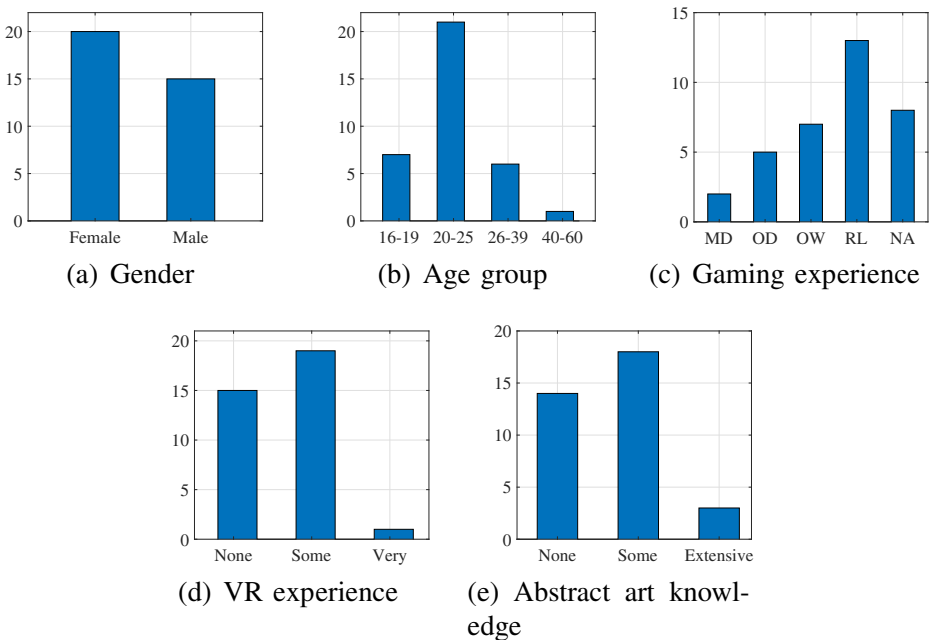

(a) Gender (b) Age group (c) Gaming experience

(d) VR experience (e) Abstract art knowledge

Fig. 5 Participant background

The participants' VR viewing session durations were also captured as shown in Fig. 6. Female participants spent 264.2 seconds on average actively exploring the VR artwork, which is slightly shorter than the 276.9 seconds average viewing time of male participants. The durations are significantly longer than the amount of time visitors spend looking at great works of art in some museums (27.2 seconds at Metropolitan Museum of Art [39], and 21.70 to 82.31 seconds at Baltimore Museum of Art [17]). Female viewing durations are more scattered with a standard deviation of 107.7 seconds while the same measurement for male viewing time is 53.3 seconds. The shortest and longest viewing durations are 75.7 seconds and 548.5 seconds respectively. User activity data are preprocessed prior to data modelling to minimise bias.

4 Data exploration

4.1 Walk

The location of each participant was gathered by using the VR headset (head position in the physical space) and the Unity game engine (camera position in the virtual space). The location data from the two systems (head_x, head_y, head_z and player_x, player_y, player_z) are highly correlated (as shown in Fig. 7). Since the eye orientation data also came from the VR headset, the location data from the same source (head_x, head_y, head_z) was chosen for data analysis and modelling in order to minimise asynchrony. Figure 7 also shows a very low correlation between the eye orientation and the head location as well as a low correlation between experiment time and user activity data. This indicates a low risk of collinearity in the selected inputs.

Based on two coordinates *head_x* and *head_z*, Fig. 8 plots how 20 participants moved their locations during the experiment. Orange lines show the traces of movement and the blue lines contour the edge of the artwork at the floor level. The figure clearly demonstrates the willingness of the participants in becoming familiar with the artwork and moving to different locations to explore the artwork from different perspectives. The data also reveals
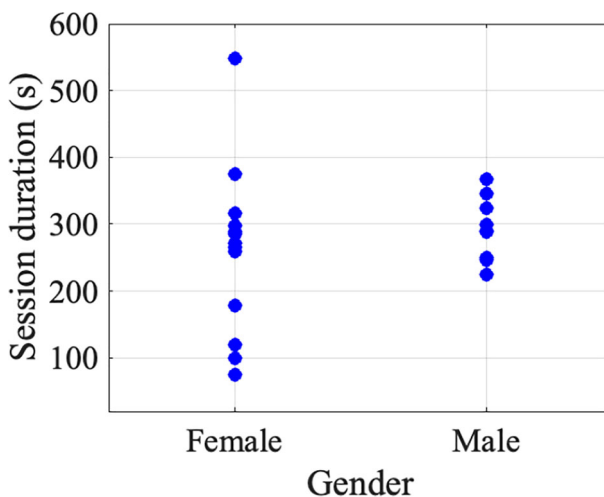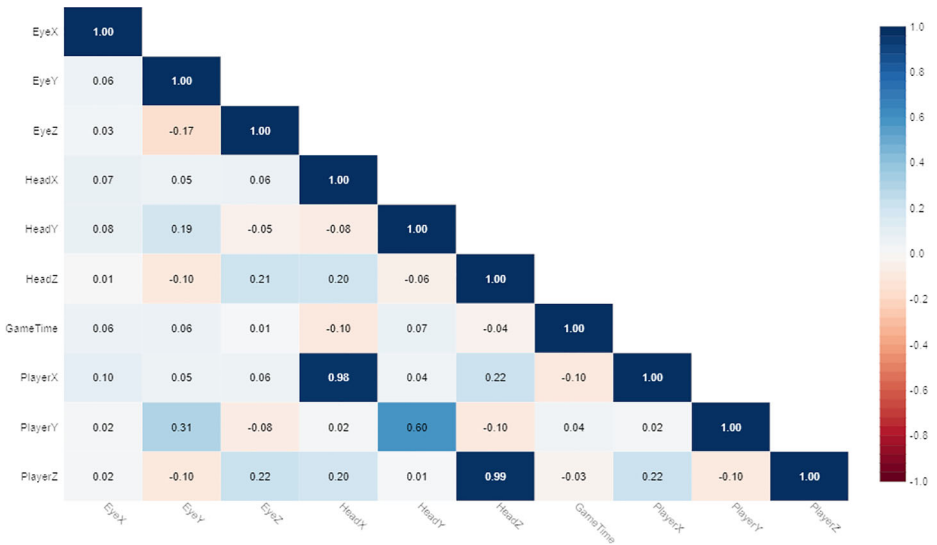

Fig. 6 VR viewing session duration

Fig. 7 Correlation matrix of measured data

completely different characteristics in the movements of participants and how the movements in VR differ from those in a physical gallery. We observed "wanderers" such as *p4473* and *p4646* who travelled both inside and outside of the artwork and covered a very large area. There were "explorers" like *p1958* and *p2654* who also moved extensively, but in a slightly more cautious way. Some participants like *p1679* preferred to stay near the start point and mainly explored the artwork by turning their head and eyes.

Head elevation *head_y* (Fig. 9) also shows active engagement with the artwork. Normal walking leads to small fluctuations in head elevation, while large and distinctive elevation changes indicate participants purposely lowering their heads. This seems to be to take a closer look at a much lower object (often near the ground level) or to bend over to avoid contact with a virtual "obstacle" while exploring. Figure 9(b) shows the head elevation from all participants. Each data series has its own "baseline" as determined by the height of the person wearing the headset. Clearly some participants demonstrated more enthusiasm than others in getting close and interacting with the artwork. The differences in user behaviours can be attributed to different personalities and backgrounds.

4.2 Eye orientation

As participants moved around or through the artwork, they indicated their attention on brushstrokes by how they move their heads and eyes in different directions. As human eyes are constantly moving, it is understood that there is noise and redundancy in the eye gaze data. For instance, when we move our attention from one object to a different object, our eyes may pick up many other objects in between. In the case of this experiment, there are 555,727 eye gaze records captured including groups of three types of signals: "Focus In" which marks the moment a participant started gazing at an object, "Normal Frame" which is a heartbeat signal registered every 30 milliseconds while the gaze on an object continues, and "Focus Out" which flags the moment the participant moved the attention away from an object. Therefore, a one second long gaze would appear in record as one "Focus
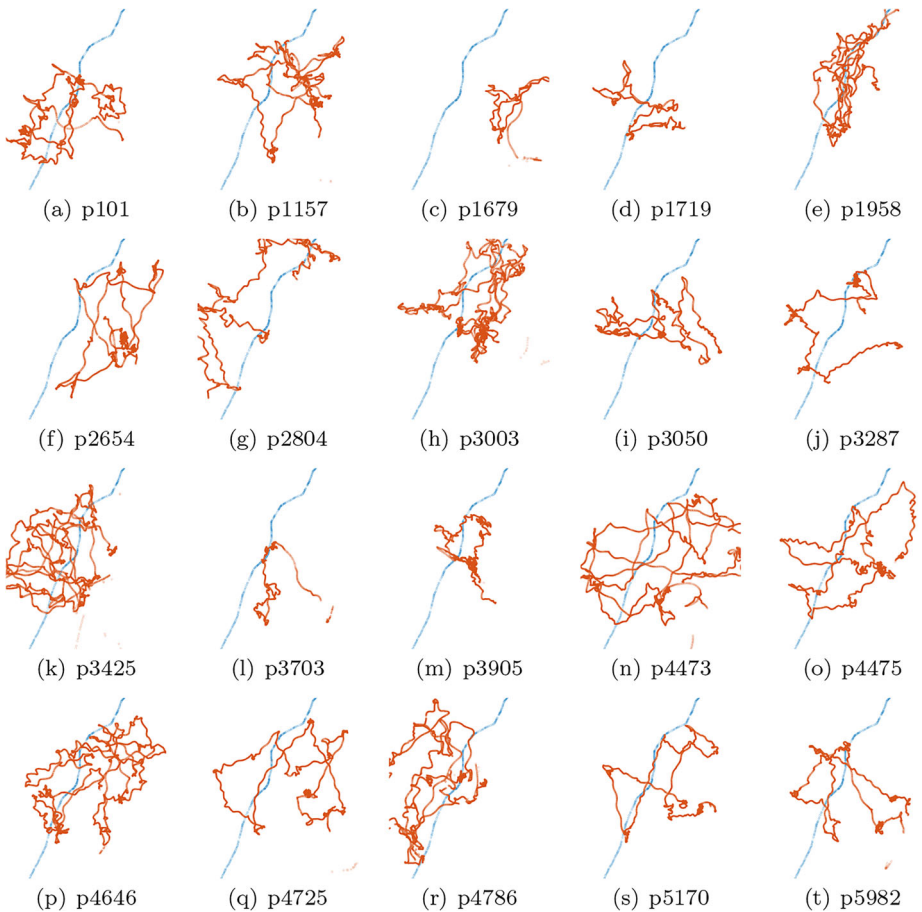
(a) p101	(b) p1157	(c) p1679	(d) p1719	(e) p1958
(f) p2654	(g) p2804	(h) p3003	(i) p3050	(j) p3287
(k) p3425	(l) p3703	(m) p3905	(n) p4473	(o) p4475
(p) p4646	(q) p4725	(r) p4786	(s) p5170	(t) p5982

Fig. 8 Traces of participants' movements (top down view)

In" followed by around 30 "Normal Frame" then one "Focus Out". A "Focus out" is normally followed by another "Focus In" when the users attention switched to a new object. Eye orientation information within each signal group is nearly identical. Hence all "Focus In" events (a total number of 59,928) were extracted to represent all gaze events. The events captured the head location of the participant (as shown in Fig. 8), the unique Brushstroke ID, as well as the three dimensional eye orientation data (eye_x, eye_y and eye_z) as shown on a sphere in Fig. 10.

As is with the walk, the eye orientation data exhibited significantly different behaviours between the participants. In general, the participants who were more active in moving their location were more likely to have looked in different directions, as demonstrated by *p3425*. Participants didn't travel deep into the artwork, for instance, *p1679* often concentrated their attention on one half of the sphere as the artwork was mostly in front of them. Some dense areas at both poles of the sphere were also observed which indicates that the participants spent time investigating brushstrokes that are much lower than their eye level and also objects directly above them. No general eye orientation pattern was demonstrated across the entire participant group though some participants seem to share similar viewing preferences.
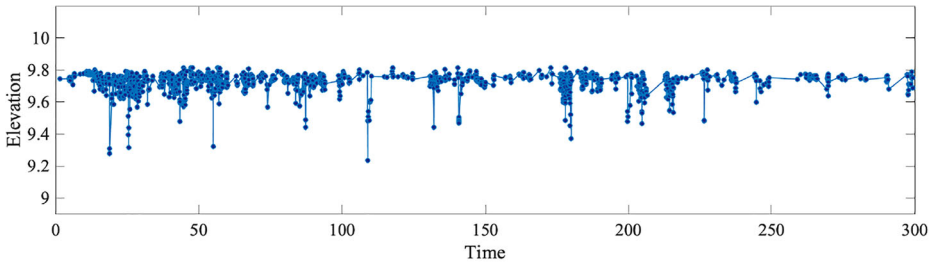
(a) One participant
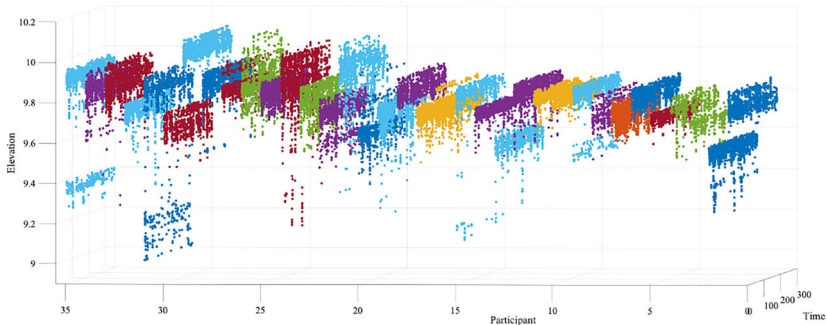
(b) All participants

Fig. 9 Head elevation during experiment

The diversity of audience behaviour is also observed at brushstroke level. Figure 11 shows the eye orientation for four different brushstrokes with data from participants separated by colour. Many brushstrokes were viewed by the participants from different angles. Individual participants also looked at the same brushstrokes multiple times from different viewpoints. We surmise that this can be attributed to a combination of 1) the inherent qualities of VR i.e., 3D artwork and free audience movements (around or through objects) which encourages more exploration and diverse viewing angles; 2) how the artist choreographs audience interactions via design choices, such as composition, colour, texture. For instance, some brushstrokes are partially visible from an angle such that participants need to change their location to discover occluded content. This is further continued where shape and lighting configuration also allow some brushstrokes to change their appearance when viewed from different angles.

4.3 Colour encounter

Overall the data showed that the participants encountered 64 different colours in the VR painting. To understand this further, we studied how different participants explored those colours and whether there were different preferences to colour by gender.

Figure 12 depicts the number of colours seen by each participant, their viewing duration and gender where a larger circle indicates a higher number of colours viewed. In general it was noticeable that a longer viewing session led to more colours being encountered, with females being more likely to encounter more colours under the same condition as males.
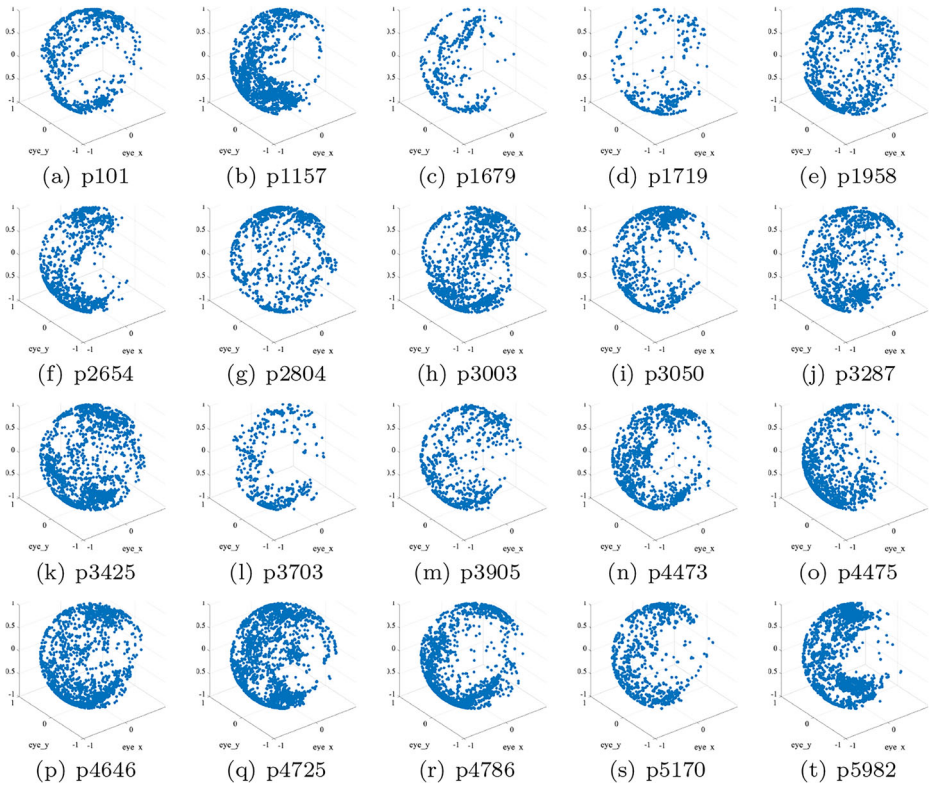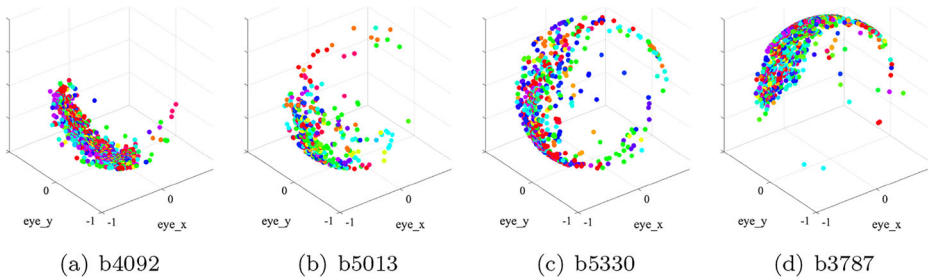
Fig. 10 Eye orientation

Fig. 11 Eye orientation on brushstrokes from participants (grouped by marker colours)
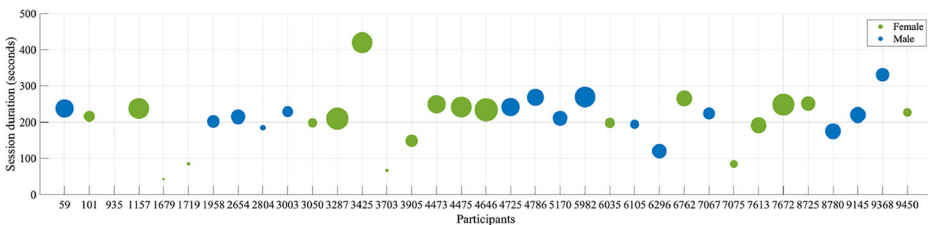

Fig. 12 The number of colours seen by each participant by their viewing duration and gender

We also investigated if there was a colour preference demonstrated by gender. The Pearson correlation between female's and male's eye gazes on top colours are 0.46 for gaze count and 0.68 for gaze duration. The moderate but significant correlations suggest that female and male share some colour preferences though some major differences exist. The correlation measures treat colours as independent categories and neglect their colour similarity as perceived by human. Future work will look into alternative methods that take into account human perception. Figures 13 and 14 show the top 15 colours encountered for female and male, measured by the proportion of gaze duration (how long each colour was looked at) and gaze count (how many times a colour was looked at) on brushstroke colours. The pie charts' colours resemble the actual brushstroke colour in the VR painting. It is important to point out that colour is only one of the many deterministic factors for user attention. Other factors include size, shape, structure, brushstroke type, distance to the viewer and composition within the artwork. In spite of the complexity of human attention, in this comparison between genders, it seems female participants demonstrated a preference for green, while male participants viewed red brushstrokes more often in terms of both duration and count.

4.4 Types of eye gaze

During the experiment we observed different levels of attention demonstrated through the participants eye gaze, varying between tens of milliseconds to seconds. Gazes of different durations may correspond to different viewing habits. Some gazes are part of a quick scan while others can be a close inspection of artwork details.

Using K-Means, an effective and commonly used method for clustering, the gaze data was separated into distinctive groups based on the distribution of their numerical values. Gaze events and their duration was gathered from all sessions, then K-Means method was applied with a range of configurations. Figure 15(a) shows the Within Cluster Sum of Squares (WCSS) results when data is split into different numbers of groups. WCSS evaluates the effectiveness of clustering by measuring the squared average distance of all gaze duration within a cluster to the cluster centroid. As the number of clusters increase, the WCSS decreases, which means better clustering at the data level though too many clusters
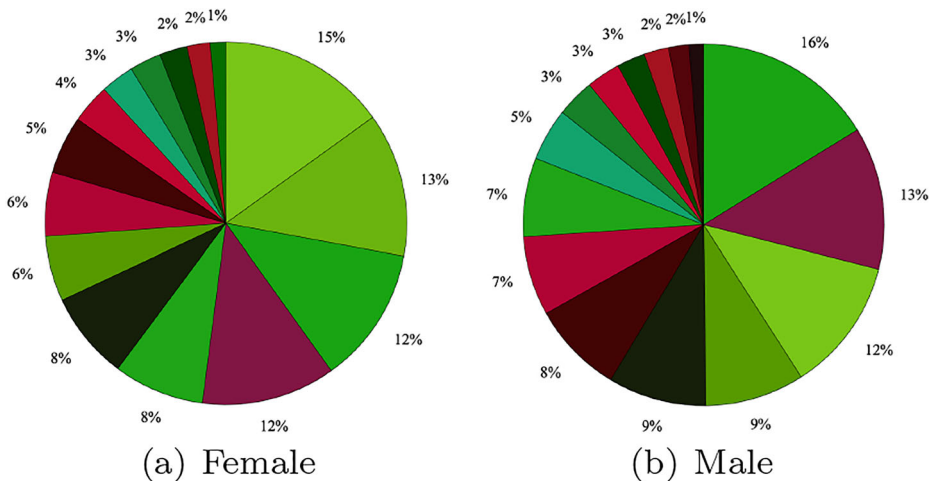

Fig. 13 Top 15 colours by gaze duration (pie charts show actual brushstroke colour)
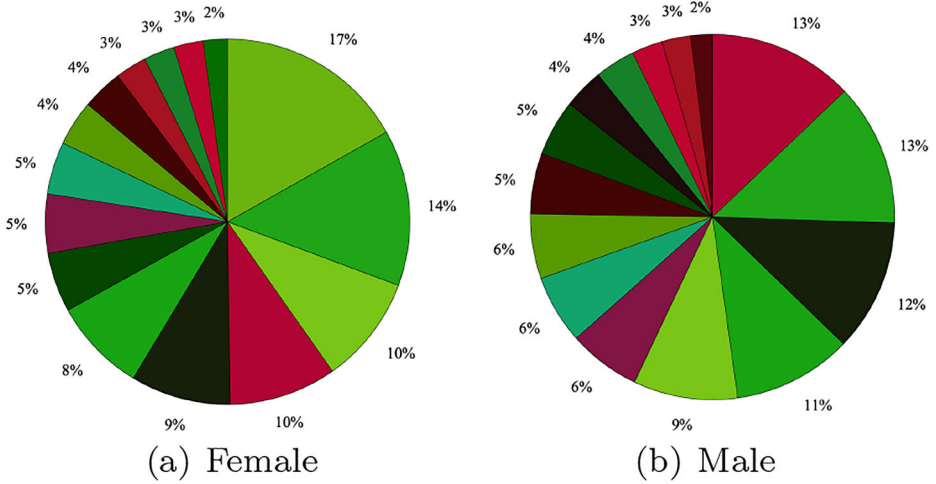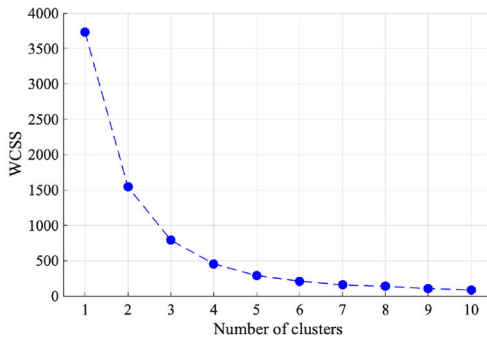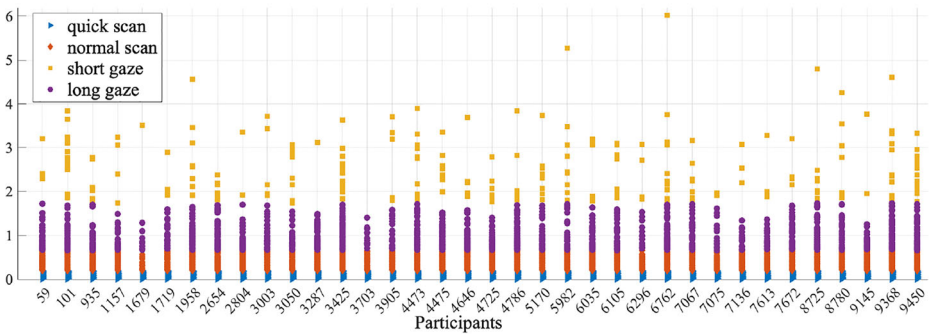
(a) Female   (b) Male

Fig. 14 Top 15 colours by gaze count (pie charts show actual brushstroke colour)

(a) Eye gaze clustering

(b) Gaze type from participants

Fig. 15 Eye gaze in user experiment

prevent a clear understanding of the levels of user attention. With four clusters, the clustering has a good balance between domain interpretability and performance. Hence, we split the level of gaze into four degrees of user attention: quick scan, normal scan, short gaze, and long gaze with the centroid of 0.047, 0.338, 0.9530, and 2.488 seconds respectively.

Participants' gaze was grouped into four clusters and plotted in Fig. 15(b). The figure indicates that all participants demonstrated a large amount of quick scan and normal scan activities while they explored the artwork. The data clearly indicates different levels of long gaze activities across this group of participants. This shows how a piece of VR painting may elicit different experiences as participants chose their own ways to interact with the artwork. Such understandings of art encounter preferences via gaze analysis can become a main contributor for experimenting with new artwork that adapt to individual preferences.

5 User behaviour classification

While participants' eye gaze can be clustered into groups of distinctive natures, each individual participant demonstrated a unique encounter with the experimental abstract VR painting, as was communicated by their different eye orientation and body movement. There are strong indications that some of these differences are attributed to participants' background such as gender, age, personality and related experiences/skills. If a non-intrusive quantitative measurement of human behaviour is used to accurately infer the background of the participants, it will be possible to deliver focused encounters and improved experiences for different user groups. This section provides a first attempt to classify art audiences by gender, age group, VR experience, gaming experience and art background based on six features: eye_x, eye_y, eye_z, head_x, head_y, and head_z.

5.1 Data pre-processing

As discussed in Section 4.2, the features were continuously sampled during the experiment which resulted in 555,727 raw samples points. From the raw data, we extracted 59,928 samples where participants switch their attentions to a brushstroke. This data extraction process is essential to remove data redundancy. Because participants' viewing duration was not controlled, the samples are not evenly distributed among participants. In order to minimise bias in input data to the modelling process, 15,750 sample points (450 for each participant) are used for modelling. Table 1 shows descriptive statistics of the data. Questionnaire responses (Fig. 5) were used as a ground-truth labels for model training and validation.

Table 1 Features for data modelling

Features	Notes	Minimum	Maximum	Mean	Std
eye_x	Eye orientation (Cartesian coordinates X-axis)	-0.99997	0.99990	-0.20253	0.43767
eye_y	Eye orientation (Cartesian coordinates Y-axis)	-0.99935	0.99973	0.10302	0.40633
eye_z	Eye orientation (Cartesian coordinates Z-axis)	-0.99996	0.99999	-0.08484	0.76453
head_x	Head coordinates in physical space (X-axis)	-17.50724	-9.02000	-13.78706	1.22423
head_y	Head coordinates in physical space (Y-axis)	6.21650	10.11267	9.78626	0.19700
head_z	Head coordinates in physical space (Z-axis)	6.37128	17.35471	8.97618	1.04039

Table 2 list some raw measurements data taken from the first few seconds of a VR viewing session.

5.2 Data modelling

Our session-based schema considers each of the six measurements as sequential data that independently captured the participants' behaviour during a viewing session. This means that every column in Table 2 is an independent input used to classify participants. The order of data within each sequence is retained and treated as an important feature. So visiting location A then location B is different from visiting location B first. A use case of such a session-based model would be the model observing how a person's activities change over time then predict the person's background.

Artificial neural networks (ANNs) are very effective for complex machine learning tasks such as time series analysis, image classification, object detection, data processing, etc. In particular, different types and configurations of ANNs have been successfully used for human behaviour and mobility modelling in recent years. R. Zhu et al. used an ensemble model of convolutional neural networks (CNNs) to classify human activities based on mobile sensing data [46]. In [22], a relationship between human mobility and personality was established using deep neural networks. Multiple deep learning models were used

Table 2 User behavioural data (excerpt)

eye_x	eye_y	eye_z	head_x	head_y	head_z
-0.764599	0.4071314	-0.4996187	-12.54632	9.87301	8.309494
-0.6116759	0.6332223	-0.4741834	-12.54806	9.885844	8.333403
-0.6475974	0.6544955	-0.3901766	-12.51895	9.907124	8.296976
-0.6247082	0.6315342	-0.4592134	-12.53775	9.901265	8.325699
-0.600112	0.6160315	-0.5102419	-12.54874	9.898056	8.341338
-0.6221098	0.6034535	-0.4987954	-12.54431	9.901594	8.338136
-0.6538676	0.6154199	-0.4401109	-12.52308	9.907201	8.312715
-0.5598859	0.6049045	-0.5662138	-12.53829	9.903415	8.336252
-0.5145928	0.5948046	-0.6175599	-12.53389	9.904068	8.33251
-0.5426044	0.6208919	-0.5657312	-12.50965	9.912107	8.294859
-0.5358268	0.6092777	-0.5844959	-12.51862	9.909679	8.307343
-0.5243732	0.6081678	-0.5959446	-12.5252	9.907698	8.315166
-0.5276814	0.6109888	-0.5901119	-12.52225	9.908517	8.309989
-0.6765658	0.6154901	-0.4042251	-12.50916	9.911985	8.287548
-0.752409	0.5629492	-0.3419628	-12.52749	9.907016	8.312791
-0.7641846	0.571066	-0.2998287	-12.51686	9.909601	8.296782
-0.7675983	0.5718319	-0.2894368	-12.51374	9.91063	8.291984
-0.7610472	0.565481	-0.3178392	-12.52108	9.908808	8.302503
-0.7607328	0.5657485	-0.3181158	-12.52104	9.908838	8.302495
-0.738611	0.551627	-0.3874907	-12.53943	9.904744	8.326933
-0.7537291	0.5594133	-0.3448617	-12.52739	9.907848	8.311537
-0.7703838	0.5601763	-0.3044409	-12.51443	9.911286	8.295473
-0.6810743	0.542149	-0.492128	-12.54407	9.903443	8.336664

for skeleton-based human activity and hand gesture recognition [31]. A survey paper [1] reviewed several CNN-based methods for predicting viewers' eye gaze based on images sampled from front-facing cameras. Therefore ANNs are well-suited for our complex tasks of analysing user attention from both movements and eye gaze.

A range of deep learning classification models are built as means to explore how user behaviours are attributed to their personal background. The modelling is based on in-the-wild tracking data captured directly from user experiments and user survey answers, which are different from other datasets that are engineered and labelled for a specific and narrow machine learning task such as image classification. A good model performance indicates a strong link between the tracked user behaviour patterns and the associated background information such as gender or art knowledge.

We aim at comparing the performance of three typical neural network structures: Feed-forward Dense Network (FDN), Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) [18]. Keras and Tensorflow are the frameworks chosen for the model implementation [20, 41].

The FDN takes six inputs and concatenates them before passing the data through three Dense layers with Dropouts to mitigate over-fitting (Fig. 16). The CNN structure adds a two-layer 1D convolutional base with MaxPooling to each of the six inputs before they are concatenated (Fig. 17). This should allow the network to exploit patterns in input sequential data. The patterns may reflect participants' distinctive movements such as "leaning forward to closely inspect a paint brushstroke then moving back" and "turning their head from left to right to see more of the painting". LSTM, as a Recurrent Neural Network (RNN) design, conducts multiple iterations of learning and uses outputs from previous steps to inform the current iteration. This allows the network to discover temporal dynamics in the sequential data. LSTM can exploit this additional "context" to process human behaviour. Two LSTM layers with recurrent Dropout were applied to each input layer as shown in Fig. 18.

As a standard practice, the output shape of each network structure adapts to each classification task. For gender prediction, the output shape is set to 1 (binary) via a *sigmoid* activation function at the last layer, as all participants chose either female or male in the pre-test questionnaire. Other predictions have categorical outputs via a *softmax* activation
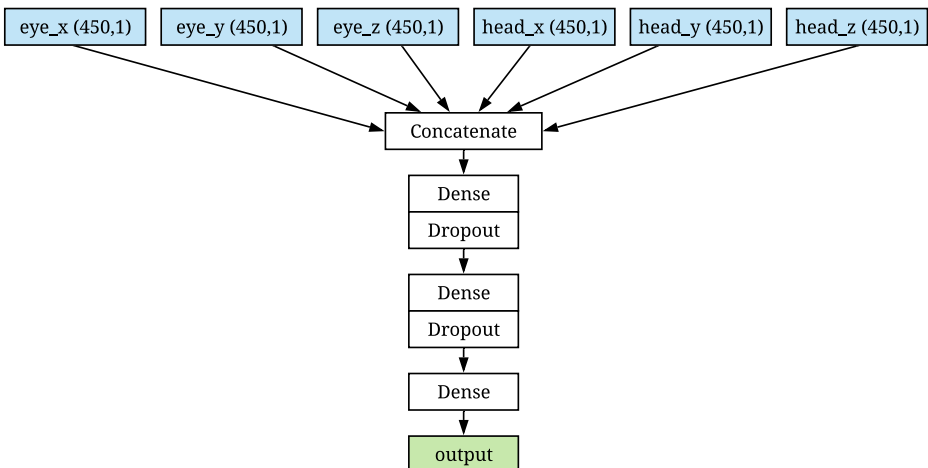

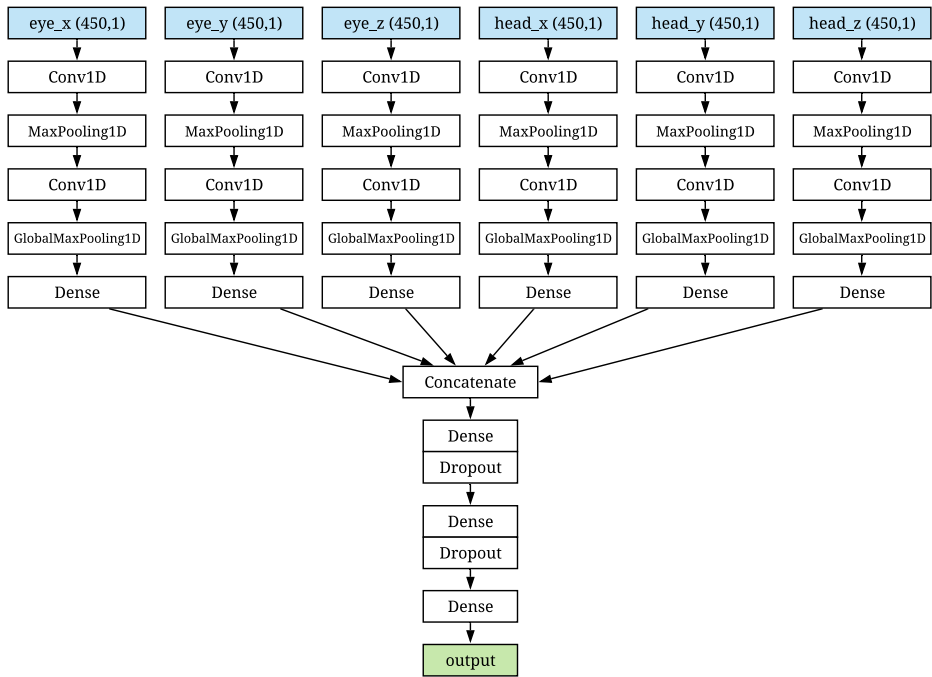**Fig. 16** Session-based FDN network

Fig. 17 Session-based CNN

function and the shapes are configured to match the number of label categories. The designs use Categorical crossentropy or Binary crossentropy as the loss function when tuning model weights.

Since the dataset contains numeric values of different data ranges and scales, the data is normalised for each of the six measurements. Because a strict rule was not placed on how long a viewing session should be, some participants spent more time than others. For session-based models, the input data from all 35 participants must be in the same shape. Therefore, the first 450 Focus-In events when participants started gazing a brushstroke were sampled from each participant for the modelling. The deep learning training on each of the three network structures was carried out 20 times independently. Each run started with a random 70/30 split on the dataset so that 70% of the data was randomly selected for training while the rest were used for validation. It was made sure that the data from any participant will fall in either the training set or the validation set. This ensured that the participants in the training set were entirely unknown to the validation process. A maximum of 200 epochs for each run was set but Keras' "EarlyStopping" callbacks was used to stop the training process and restore the best weights when the validation loss had stopped decreasing in several epochs. The training was carried out on an HP Workstation with a Nvidia TITAN RTX graphics card for GPU acceleration.

5.3 Results and discussions

Figures 19 and 20 shows the boxplots (from 20 runs) of the Top1 and Top2 accuracy of our models predicting gender, age group, game experience, VR experience, and abstract art
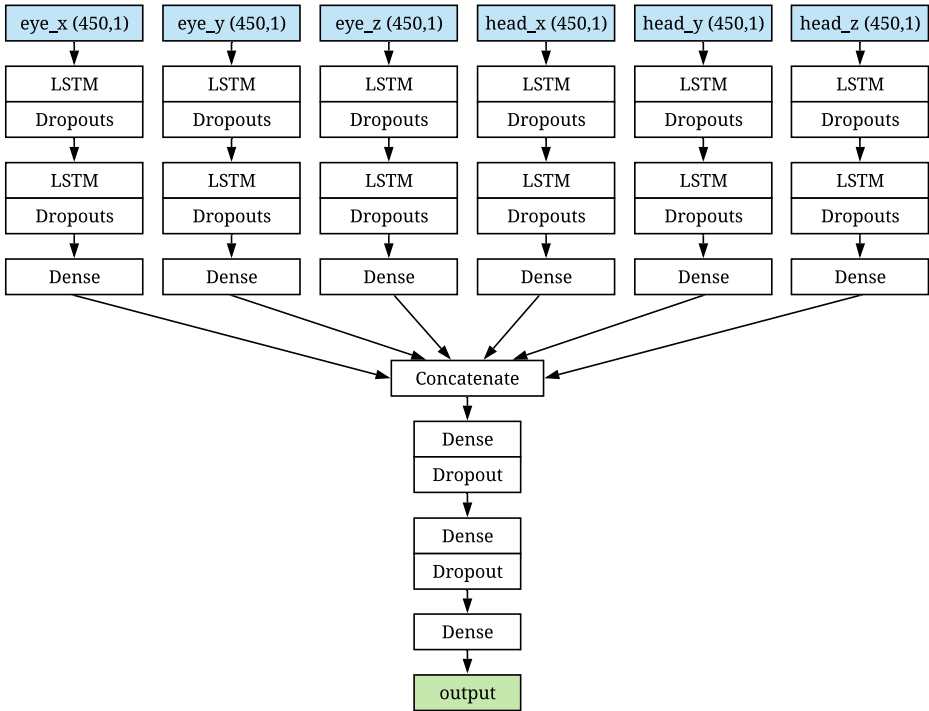
Fig. 18 Session-based LSTM

knowledge. The CNN model exhibits the best performance in all predictions. The median of the results given by the CNN model is 0.82, 0.64, 0.46, 0.55 and 0.55 for Top1 accuracy. Top2 accuracy of some viewer background predictions rises above 0.9. For a dataset of 35 participants, the results are excellent for gender prediction and good for age prediction. This indicates a very strong cross-gender difference in patterns of head and eye movements when participants interact with VR artwork. A gender classification model can help customising services based on viewers' preferences and help researchers to study any gender bias in VR-related technologies and application design. Age is more difficult to predict using behavioural patterns, though its accuracy is significantly higher than a random prediction (with an accuracy of 0.25). Meanwhile, the results from our experiment indicate that participants with different levels of game experience, VR experience, and abstract art knowledge did not show significantly different behavioural patterns as they moved between different parts of the artwork.

Increasing the number and diversity of participants can improve the model performance and support the training of more complex neural network structures. A larger dataset may help discover deeper behavioural patterns for background classification. However, VR experiments are resource intensive and equipment-dependent. Consumer HMDs are becoming popular at homes but eye-tracking capabilities are only available in specialised HMDs. One possible solution is data augmentation. Data augmentation is a common tool used in image classification to improve model performance over small training data. Future work will experiment with different data augmentation techniques such as adding
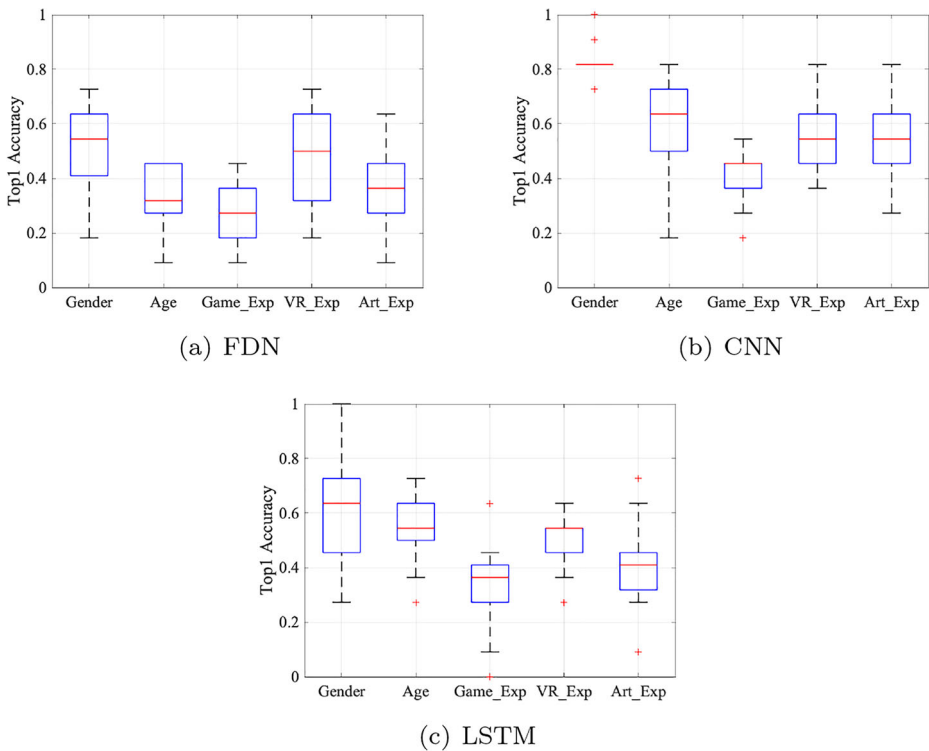
(a) FDN

(b) CNN

(c) LSTM

Fig. 19 Performance of session-based models (Top1 accuracy)

a human behaviour noise model to expand the training data to accommodate more complex ANN structures. The data augmentation may also generalise findings from individual experiments.

Some data labels such as game experience, VR experience, and art knowledge came from participants in the form of self-assessment. As the scoring being subjective, participants may apply different standards while reporting their background. In future work, small proficiency and aptitude tests can be planned to complement survey data with objective metrics in order to help establish any influence of user background to VR art viewing behaviours.

6 User attention visualisation

6.1 VR heat map

One of the main objectives of this work is to develop a tool to visualise user attention on VR artwork as a feedback channel for VR content creators. A common method to visualise human attention is a 2D heat map of gaze intensity superimposed on the original content. The heat map is often semi-transparent so that the original objects remain partially visible. However it becomes problematic when associating a 2D heat map to 3D brushstrokes, especially in densely populated areas where brushstrokes intertwine. Therefore, we propose VR heat map as a new method for user attention visualisation.
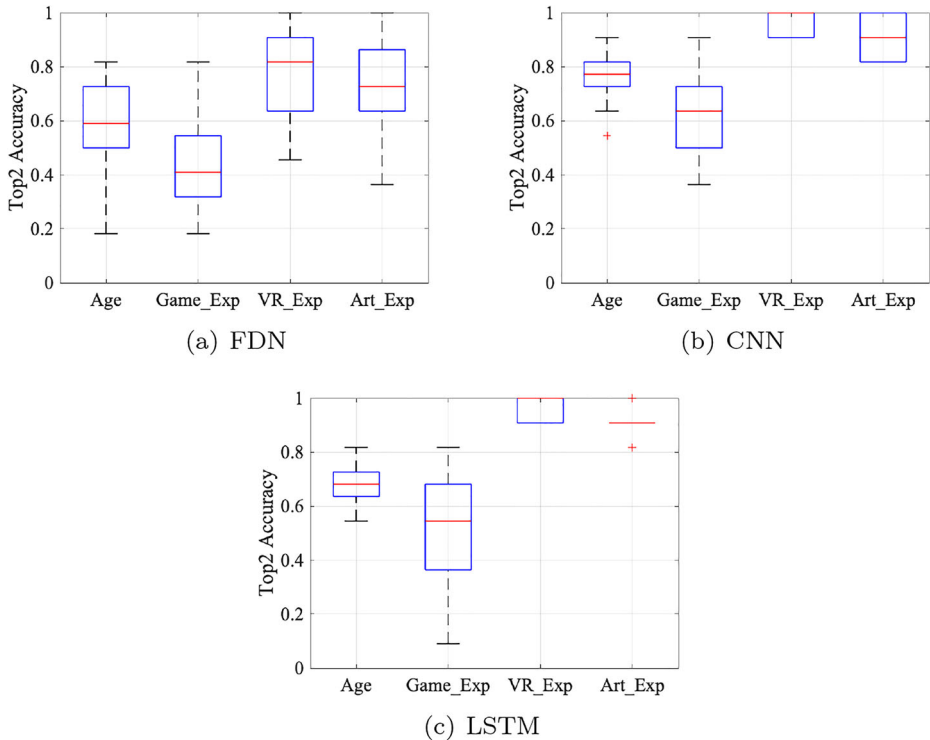
(a) FDN

(b) CNN

(c) LSTM

**Fig. 20** Performance of session-based models (Top2 accuracy)

Recent advances in information visualisation within game design are used to improve the gaming experience. For instance, player health and ammunition levels are not displayed as an overlay at a corner of the screen but integrated as part of the game character or an item of equipment that they carry. In alignment with this idea from game design, this work experimented with two methods of visualising the viewers' attention in VR by controlling the opacity and colour saturation of the brushstrokes based on the level of attention received from 35 participants. The implementation was done by a customised script based on Tilt Brush Unity Toolkit [43].

Figure 21(a) shows the results when there is a single opacity threshold applied to the artwork by altering its alpha channel. In this particular example, the opacity of a brushstroke is set to 1 if it has received more than 1 second of attention per viewing session (35 seconds in total for the experiment). Otherwise, the opacity is set to 0. This screening method clearly demonstrates which parts of the artwork viewers were mostly interested in on average, but the removal of other parts makes it difficult to analyse the users' interest in its context. In addition, finer control of opacity was also tested. However most Tilt Brush brushes are natively opaque and cannot be set semi-transparent directly.

For the colour saturation-based visualisation, the R,G,B values of brushstrokes are altered (Fig. 21(b)). Any brushstroke that received user attention above a threshold keeps its original colour. Other brushstrokes are de-saturated based on how far their received attention is from the threshold. The process factored in, the original brightness of brushstrokes

(a) Opacity-based attention visualisation
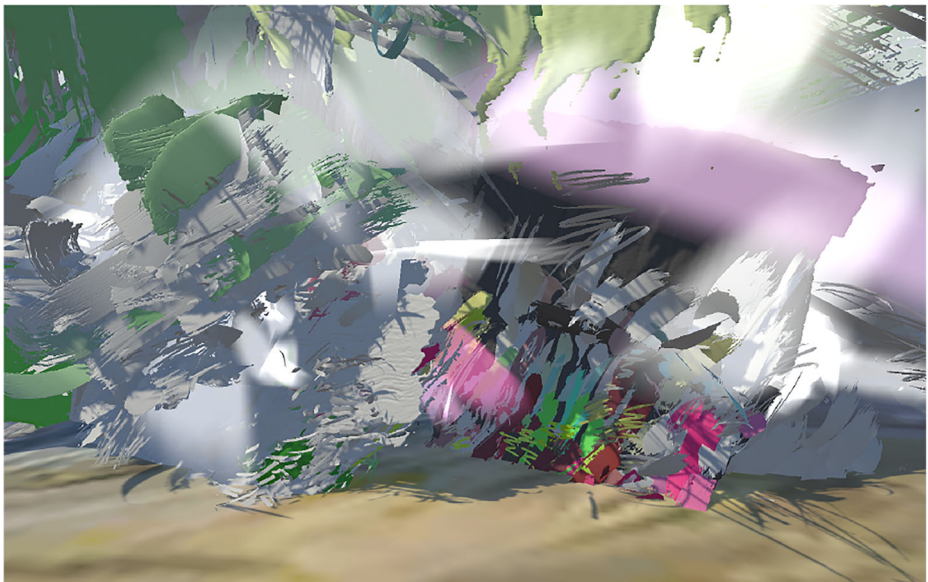
(b) Saturation-based attention visualisation

Fig. 21 User attention visualisation

and human visual system's response to colours, in order to preserve the aesthetic of the artwork as much as possible. This means a bright colour will retain brightness after being de-saturated or converted to a neutral colour. The saturation-based visualisation preserves the structure of the artwork, whilst revealing how user attention moves from one "hotspot" to another.

For both VR heat map options, artists use a "Compare" feature in game engine to switch between original paintings and augmented paintings in order to assess visual attention on the artwork. Artists can change their position and view by operating a controller or by walking. Future work will include dedicated user experiments for VR heat map designs.

6.2 Verbal response

The verbal responses during the VR encounter can also provide valuable feedback to the content creators. The headset's microphone turned out to be an ideal choice to capture participants' voice with low background noise due to its close proximity to the source. Since there was no instructions for the participants to provide in-test feedback, the recordings reflect their genuine feelings and emotions.

Google Cloud Speech-to-Text API was used to automate the conversion from recorded audios to text. Figure 22 shows a word cloud of the verbal responses after the removal of stop words and filler words such as "yeah". Most keywords are linked to general feelings towards the painting especially colours ("blue", "green", "purple"), lights ("bright"), scale ("big", "little", "space"), and direction ("right", "down"). As young adults, participants also used words such as "god", "crazy", "wow", "scary" and "cool", and "like" to express their feelings towards the artwork. It was noticed during the experiment that a lot of the verbal responses were between the participants and the bystanders. In some cases, the "cross-reality" conversations seemed to have encouraged some participants to explore the artwork as they became a performer.

6.3 Post experiment interview

The post experiment interviews include a set of questions about the overall experience, means of user navigation, some features under development followed with open-ended

Fig. 22 Word cloud of audio recordings during experiment

discussions. Participants were asked to name the most memorable elements of their VR encounters and any other suggestions. All participants reported very positive experiences with the VR painting. Some comments were entirely unanticipated. Participant 5170 commented that he loved the experiences of "going into the paint and looking up", which really gave him "a sense of scale". He preferred walking to explore the artwork more than the idea of "flying" using VR controllers because "tethering to the ground" gave him "a point of reference". Participant 4475 felt that she was "in a virtual jungle" and "loved the lights shining through the paint". She also suggested "more pink trees" and "tropical music" in the background. Participant 1157 "felt a bit lost in the painting... but in a good way...". The feedback reflects her walking patterns shown in Fig. 8(b) where she spent little time adventuring inside the artwork. Participant 4646 "really liked a glowing triangular area in the middle of nowhere" because it reminded her of a song. She also suggested "swimming in the paint" as a better navigation method compared to walking and flying.

Overall, the post experiment interviews confirmed the artist author's intuition built upon their own encounters learning about how to make and navigate through an abstract VR painting. Those impressions being that here is a medium that has the potential to be immersive in a way currently beyond that of a traditional physical painting. These types of works could open up the field of painting to new audiences, such as the gaming community, with the possibility of expanding existing audiences by introducing them into new realms, that is Fine Art in VR.

7 Discussions

The user experiment resonate with what many artists already do while creating physical artwork. It's common practice for artists to observe the audience/viewer/participant when encountering their artworks, to hide themselves from view and quietly watch the dynamics at play. In general, artworks go through many phases of testing starting at the artists' studio, through various stages of experimentation, continued into more public spaces such project spaces or at screenings etc., which takes place long before any much grander launch is planned. This experiment keenly demonstrated how these kinds of quantitative processes of analysing responses could be a valuable tool to add to the process of VR content creation.

Our system enables VR artists to observe an audience's encounter of their artwork. The observation is augmented by user attention visualisations and insights of audience behavioural patterns from machine learning. An artist can study how an artwork is perceived by different genders, age groups, and different levels of art knowledge. For instance, if visitors of a certain gender or cultural background are not engaging with the artwork as much as their counterparts, artworks, VR environment or physical exhibition can be adapted to accommodate their viewing preferences. Some adaptations can be pre-programmed using game engines.

Game engines and game design practices are likely to see an increasing adoption by the VR artists. Besides offering an environment to accommodate artwork, the game engines can empower content creators to exploit dynamic elements that respond to how the artwork are perceived by audiences. Game design theories such as the "three Cs" (character, camera and control) and level design may also assist artists to choreograph audience interactions in alternate reality.

During the experiment, many participants reached their hands out and tried to stroke the artwork. In post experiment interviews, most participants welcomed the idea of "touching" the artwork. While analysing the hand gesture data captured in the experiment was being carried out, the observations from the experiments indicate that multi-sensory experience may have a pivotal role in VR artwork. For abstract painting, this can lead to new experimental designs for how brushstrokes respond to user interactions with sound, visual deformation, and haptic feedback [23, 40].

We plan to conduct additional user experiments using different VR artworks, participants of different backgrounds, user navigation techniques, and VR social interaction methods to generalise our findings. While the user behaviour models are specific to artwork, exhibition style, cultural and other contextual factors, the system, tools, and methodologies used for our experiments can be used for similar research. The system can be released as a standalone application to benefit HCI researchers, artist, VR exhibitors and wider communities. Our public GitHub repositories [29] are continuously updated with the latests software toolkit and open datasets. Future work will also extend the analysis of user's visual attention especially the causes and impacts of different eye gaze levels (quick scan, normal scan, short gaze, and long gaze) on VR objects.

8 Conclusions

Alternate realities are becoming the new pathways for content creation, distribution and audience engagement in the creative communities. Understanding how audiences explore and interact in new forms of media is essential to realise the full potential of alternate realities. Using a purposely built abstract VR painting and an experimentation system, our user experiment captured eye gaze and body movement patterns from 35 participants while interacting the artwork. The results can help expanding the knowledge base of user attention and interactions in VR especially in the context of fine art. Deep learning-based modelling showed how predictions of viewers' background can be made using behavioural data to potentially personalise the user experience.

Data Availability All data underpinning this publication are openly available from the University of Northampton Research Explorer at https://pure.northampton.ac.uk/en/datasets/user-attention-and-behaviour -in-virtual-reality-art-encounter.

Declarations

Conflict of Interests Murtada Dohan, Alison Goodyear, Gary Hill, Cleyon Johns and Andreas Mauthe declare that they have no conflict of interest.

References

1. Akinyelu AA, Blignaut P (2020) Convolutional neural network-based methods for eye gaze estimation: a survey. IEEE Access 8:142581–142605. https://doi.org/10.1109/ACCESS.2020.3013540
2. Battisti F, Di Stefano C (2018) Virtual reality meets degas: an immersive framework for art exploration and learning. In: 2018 7Th european workshop on visual information processing (EUVIP), pp 1–5
3. Caserman P, Garcia-Agundez A, Konrad R, Göbel S., Steinmetz R (2019) Real-time body tracking in virtual reality using a vive tracker. Virtual Real 23(2):155–168
4. Chambel T, De Simone F, Kaiser R, Ranasinghe N, Van den Broeck W (2018) Altmm 2018 - 3rd international workshop on multimedia alternate realities. In: Proceedings of the 26th ACM international conference on multimedia, MM '18, pp 2109–2110. Association for Computing Machinery, New York. https://doi.org/10.1145/3240508.3243718
5. Chen S, Gao L, Lai Y, Rosin PL, Xia S (2018) Real-time 3d face reconstruction and gaze tracking for virtual reality. In: 2018 IEEE Conference on virtual reality and 3d user interfaces (VR), pp 525–526
6. Dohan M, Mu M (2019) Understanding User Attention In VR Using Gaze Controlled Games. In: Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video, TVX '19, pp 167–173. Association for Computing Machinery, New York, https://doi.org/10.1145/3317697.3325118
7. Eads JR The Great 3D Wave. https://gifville.wordpress.com/tag/james-r-eads/. Accessed 15 July 2021
8. FOVE: FOVE 0. https://www.getfove.com/. Accessed 15 July 2021
9. Goodyear A Paint Park - An experimental installation of immersive virtual and physical paintings. https://www.mkgallery.org/whats-on/paint-park/. Accessed 15 July 2021
10. Goodyear A Phygital works. https://www.alisongoodyear.co.uk/copy-of-palette-works. Accessed 15 July 2021
11. Goodyear A, Mu M (2019) Abstract painting practice: Expanding in a virtual world. In: Adjunct Proceedings of the ACM International Conference on Interactive Experiences for Television and Online Video (TVX). figshare. Collection. https://doi.org/10.6084/m9.figshare.c.4515005.v4
12. Grant D America's virtual museums take on new significance as covid-19 lockdown deepens. https://www.theartnewspaper.com/news/covid-19-pushes-museums-to-embrace-the-virtual-world. Accessed 15 July 2021
13. Grau O (2003) Virtual art: from illusion to immersion MIT press
14. Hammady R, Ma M, Strathern C, Mohamad M (2020) Design and development of a spatial mixed reality touring guide to the egyptian museum. Multimed Tools Appl 79(5):3465–3494
15. Hashemi SH, Kamps J (2018) Exploiting behavioral user models for point of interest recommendation in smart museums. New Rev Hypermedia Multimed 24(3):228–261
16. Hayes J, Yoo K (2018) Virtual reality interactivity in a museum environment. In: Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology, VRST '18. Association for Computing Machinery, New York, https://doi.org/10.1145/3281505.3281620
17. Heidenreich SM, Turano KA (2011) Where does one look when viewing artwork in a museum? Empir Stud Arts 29(1):51–72
18. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural computation 9(8):1735–1780
19. HTC VIVE Pro Eye. https://www.vive.com/uk/product/vive-pro-eye/. Accessed 15 July 2021
20. Keras: The Python Deep Learning library. https://keras.io/. Accessed 15 July 2021
21. Kevin S, Pai YS, Kunze K (2018) Virtual gaze: Exploring use of gaze as rich interaction method with virtual agent in interactive virtual reality content. In: Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology, VRST '18. Association for Computing Machinery, New York, https://doi.org/10.1145/3281505.3281587
22. Kim DY, Song HY (2018) Method of predicting human mobility patterns using deep learning. Neurocomputing 280:56–64. https://doi.org/10.1016/j.neucom.2017.07.069. Applications of Neural Modeling in the new era for data and IT
23. Lecuyer A (2017) Playing with senses in vr: alternate perceptions combining vision and touch. IEEE Comput Graph Appl 37(1):20–26
24. Lin EL Girl with a Pearl Earring 3D. https://sketchfab.com/models/d8d676cbfea44a2db2569fabc4ed6562. Accessed 15 July 2021

25. Lopez S, Yang Y, Beltran K, Kim SJ, Cruz Hernandez J, Simran C, Yang B, Yuksel BF (2019) Investigating implicit gender bias and embodiment of white males in virtual reality with full body visuomotor synchrony. In: Proceedings of the 2019 CHI conference on human factors in computing systems, pp 1–12
26. Lugrin JL, Kern F, Schmidt R, Kleinbeck C, Roth D, Daxer C, Feigl T, Mutschler C, Latoschik ME (2018) A location-based vr museum. In: 2018 10Th international conference on virtual worlds and games for serious applications (VS-games), pp 1–2. https://doi.org/10.1109/VS-Games.2018.8493404
27. Mack K (2017) Blortasia: a virtual reality art experience. In: ACM SIGGRAPH 2017 VR Village, pp 1–2
28. Marwecki S, Wilson AD, Ofek E, Gonzalez Franco M, Holz C (2019) Mise-unseen: Using eye tracking to hide virtual reality scene changes in plain sight. In: Proceedings of the 32nd Annual ACM symposium on user interface software and technology, UIST '19, pp 777–789. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3332165.3347919
29. Mu M Virtual Reality and user attention research at University of Northampton. https://github.com/UoN-VR. Accessed 23 July 2021
30. Museum SD dreams of dalí. https://thedali.org/dreams-of-dali-2/. Accessed 15 July 2021
31. Núñez JC, Cabido R, Pantrigo JJ, Montemayor AS, Vélez JF (2018) Convolutional neural networks and long short-term memory for skeleton-based human activity and hand gesture recognition. Pattern Recogn 76:80–94. https://doi.org/10.1016/j.patcog.2017.10.033
32. Patney A, Salvi M, Kim J, Kaplanyan A, Wyman C, Benty N, Luebke D, Lefohn A (2016) Towards foveated rendering for gaze-tracked virtual reality. ACM Trans. Graph. 35(6). https://doi.org/10.1145/2980179.2980246,
33. Parker E, Saker M (2020) Art museums and the incorporation of virtual reality: Examining the impact of vr on spatial and social norms. In: Convergence. https://doi.org/10.1177/1354856519897251
34. Pfeil K, Taranta EM, Kulshreshth A, Wisniewski P, LaViola JJ (2018) A comparison of eye-head coordination between virtual and physical realities. In: Proceedings of the 15th ACM Symposium on Applied Perception, SAP '18. Association for Computing Machinery, New York, https://doi.org/10.1145/3225153.3225157,
35. Pfeuffer K, Geiger MJ, Prange S, Mecke L, Buschek D, Alt F (2019) Behavioural biometrics in vr: Identifying people from body motion and relations in virtual reality. In: Proceedings of the 2019 CHI conference on human factors in computing systems, CHI '19. Association for Computing Machinery, New York, https://doi.org/10.1145/3290605.3300340
36. Rahman Y, Asish SM, Khokhar A, Kulshreshth AK, Borst CW (2019) Gaze data visualizations for educational vr applications. In: Symposium on Spatial User Interaction, SUI '19. Association for Computing Machinery, New York, https://doi.org/10.1145/3357251.3358752,
37. Raz G (2019) Virtual reality as an emerging art medium and its immersive affordances. In: The palgrave handbook of the philosophy of film and motion pictures, pp 995–1014. Springer
38. Sargezeh AB, Tavakoli N, Daliri MR (2019) Gender-based eye movement differences in passive indoor picture viewing: An eye-tracking study. Physiol Behav 206:43–50. https://doi.org/10.1016/j.physbeh.2019.03.023
39. Smith LF, Smith JK, Tinio PP (2017) Time spent viewing art and reading labels. Psychology of Aesthetics. Creative Arts 11(1):77
40. Sonar HA, Gerratt AP, Lacour SP, Paik J (2020) Closed-loop haptic feedback control using a self-sensing soft pneumatic actuator skin. Soft Robot 7(1):22–29
41. Tensorflow: An end-to-end open source machine learning platform. https://www.tensorflow.org/. Accessed 15 July 2021
42. theguardian: https://www.theguardian.com/culture/2020/apr/08/art-virtual-reality-coronavirus-vr. Accessed 15 July 2021
43. Tilt Brush Toolkit. https://github.com/googlevr/tilt-brush-toolkit. Accessed 15 July 2021
44. Tobii: Realistic virtual vision with dynamic foveated rendering. https://blog.tobii.com/realistic-virtual-vision-with-dynamic-foveated-rendering-135cbee59ee7. Accessed 15 July 2021
45. Zhou H (2019) Virtual reality in the art museum. In: SIGGRAPH Asia 2019 Doctoral Consortium, SA '19. Association for Computing Machinery, New York, https://doi.org/10.1145/3366344.3366441
46. Zhu R, Xiao Z, Li Y, Yang M, Tan Y, Zhou L, Lin S, Wen H (2019) Efficient human activity recognition solving the confusing activities via deep ensemble learning. IEEE Access 7:75490–75499. https://doi.org/10.1109/ACCESS.2019.2922104