Virtual transcendence experiences: Exploring technical and design challenges in multi-sensory environments

This item was submitted to Loughborough University’s Institutional Repository by the/an author.


Additional Information:

- This paper is in closed access until it is published.

Metadata Record: https://dspace.lboro.ac.uk/2134/33398

Version: Accepted version

Publisher: © ACM

Please cite the published version.
Virtual Transcendence Experiences: Exploring Technical and Design Challenges in Multi-Sensory Environments

Vanissa Wanick  
University of Southampton  
v.w.vieira@soton.ac.uk

Guilherme Xavier  
Pontifical Catholic University of Rio de Janeiro  
guix@puc-rio.br

Erhan Ekmekcioglu  
Loughborough University London  
E.Ekmekcioglu@lboro.ac.uk

ABSTRACT
In this paper, we introduce the concept of Virtual Transcendence Experience (VTE) as a response to the interactions of several users sharing several immersive experiences through different media channels. For that, we review the current body of knowledge that has led to the development of a VTE system. This is followed by a discussion of current technical and design challenges that could support the implementation of this concept. This discussion has informed the VTE framework (VTEf), which integrates different layers of experiences, including the role of each user and the technical challenges involved. We conclude this paper with suggestions for two scenarios and recommendations for the implementation of a system that could support VTEs.

CCS CONCEPTS
• Human-centered computing → Human computer interaction (HCI); Interaction paradigms; Collaborative interaction; Mixed / augmented reality

KEYWORDS
Transcendent experiences; Multi-sensory experiences; Virtual Reality; Immersive media; Multimedia storytelling; Digital consciousness

1 INTRODUCTION

Multi-sensory technologies allow us to interact with different aspects of digital embodiment, granting access to millions of experiences that could blend the physical and the virtual into one single experience. This experience can be shared and improve human collaboration, through the manipulation of remote objects [1]. Yet, there is still little knowledge about how to integrate new interactions mediated by multi-sensory and immersive technologies that could change the users’ perspective. It is in this context that we introduce and discuss the concept of Virtual Transcendence Experience (VTE).

The idea of VTE configures a new media ecology for people to navigate in immersive environments; it is more about the quality of perception than just a three-dimensional self-realisation of a customisable character, for example. In order to promote a virtual “transcendent” state, users might rely on sensorial illusions. Sensorial illusion is a key element in multi-sensory environments and it can be conveyed by the use of haptic devices, in order to make users understand their spatial awareness [2]. However, the concept of VTE goes beyond just connecting people into devices; it brings a new paradigm. VTE can set new aspects of interactive approaches towards meaningfulness, and it demands knowledge about technical and design challenges for making it happen.

Past technical and artistic related experiments from Jaron Lanier [3] tried to discuss the mix between cognitive and perceptual bias when mediated by computational devices. Two decades before Lanier’s VPL futuristic helmets, the 1962’s Morton Heilig Sensorama [4] was designed to bring audiences “inside” a simple visual demonstration; users could sit down in a tilting and vibrating chair, put their heads inside a hooded display, with breeze, fragrance and sounds, and watch a 3D image projection. In 1986, another kind of VTE was proposed by the artist David Rockeby, with his work A Very Nervous System [5]. In his work, David describes the usage of an interactive integration of space and sound; this created an interactive loop, described as a “constant transformation as the elements, human and computer, change in response to each other. The two interpenetrate, until the notion of control is lost and the relationship becomes encounter and involvement” [5]. Only 23 years later, Microsoft Xbox Kinect would try to subvert video-game manipulation and agency. Meanwhile, Second Life [6] was struggling to keep up with the expectations since 2003 as a new kind of communication among users. As a concept, VTE can be critical today, not just as a new set of work possibilities, but as dialogical and metaphorical way of doing things. However, content processing, network, interactivity quality, and interface challenges, must be addressed to permit users to have a VTE.

Considering this, it is possible to ask: (1) How are/ should interactive elements and storytelling methods be used in VTE environments? (2) What are the emerging approaches and arising research questions? In order to address these questions, we start by reviewing the main technical challenges that apply to a VTE environment. This is followed by a review of concepts related to
the design applications of the VTEs, including approaches related to Virtual Reality (VR) and quality of experience (QoE). After reviewing the main technical and design challenges of a VTE, we introduce the Virtual Transcendence Experiences framework (VTEf), with a collection of technical and design challenges, together with different applications and user interaction in order to make VTEs possible. The contribution of this paper lies on the exploration of technical and design issues related to VTEs, through the development of the VTEf and the exploration of two scenarios.

2 TECHNICAL & DESIGN CHALLENGES

2.1 Challenges with content processing

Today’s production, streaming and rendering pipelines have all been scaled, at best, for 4K video, let alone the 8K Ultra-HD standard. Hence, current efforts in VR/360° content processing focus on delivering the best visual quality under realistic hardware and networking settings. The majority of today’s VR content representation and processing efforts are tiles-based and exploit adaptive streaming paradigm. Other options are view-port aware bitrate adaptation [7], MPEG’s Dynamic Adaptive Streaming over HTTP (DASH) [8] and a tile-based processing using rectangular tiles for high-resolution media [9]. Another option could be a non-tile based but viewport-adaptive, 360° video delivery system by creating spherical scene representations with various Quality Emphasis Regions (QER) [10]. A comprehensive overview of the usage of tiles as specified within the state-of-the-art video codecs to support bandwidth efficient adaptive streaming of omnidirectional media is published by Bitmovin [11] together with various streaming strategies. With an attempt to maintain the interoperability across fast emerging HW and streaming platforms for immersive media, MPEG initiated the Omnidirectional Media Format (OMAF) [12] in 2015 that governs the stages of end-to-end VR media consumption, such as acquisition, stitching, packing, coding, segmenting and delivery over HTTP.

2.2 Challenges with networking

Various ways of transmitting VR media are present, such as through a remote cloud hosted server, resembling most of today’s media streaming solutions; through servers facilitated at network edges; and through local platforms (e.g., local server with graphics processors and high-speed Wi-Fi, or other near-field communications). Although the latency profile of edge-supported streaming is significantly better than cloud-streaming, it can still be inferior to what is required for comfortable VR viewing with free head and body movements. While 100ms~150ms latency can be deemed reasonable for most converational real-time applications and gaming, such as first-person shooting [13], a response time of 10ms~20ms is appropriate for responding to head rotations. It is possible to further mitigate the damaging effects of the latency profile. Novel viewport popularity-based tile-grouping and content-aware advance head movement prediction techniques are useful approaches, such as the degree of head rotation predictability in relation with the past head rotations, through regression and neural network [14]. There is also the possibility to explore a hybrid use of multiple transmission methods (e.g., broadcast and unicast) for scalability [15]. In this approach, the common viewport regions are transmitted as a broadcast stream while the residual regions for each individual viewpoint of interest are transmitted as unicast streams.

2.3 Challenges with interactivity and Quality of Experience

The 360° video projection format, which delivers an immersive experience with 3 degrees-of-freedom (DoF) including yaw, pitch and roll, is widely considered the introductory VR media. A more advanced step comprises 3DoF+ applications. The focus is still on 360° media with 3DoF, but with additional depth clues, which would allow moving the viewpoint in a limited space, providing some degree of motion parallax. Emerging plenoptic camera systems (e.g., Lytro Immerg [16]) are capable of recording the directions of the light rays falling on their sensors unlike traditional cameras. Hence, strong parallax effects can be reconstructed to advance the level of immersion delivered by 3DoF. Free Viewpoint Video (FVV) allows one to freely navigate a 3D scene with full parallax. This means that three additional degrees of freedom are incorporated for advanced interactivity, such as translational directions (left/right, forward/backward, up/down). This constitutes the highest degree of VR media, widely referred to as 6DoF media. However, 6DoF calls for new formats of video representation and compression to facilitate seamless 3D navigation in a virtual environment. Recent efforts concentrate on omnidirectional immersive media services. A novel subjective study for omnidirectional video streaming looked into covering the impact of stalling in omnidirectional media viewed through HMDs and its comparison with the traditional TV-based video consumption. Overall, the Quality of Experience (QoE) associated with 360° videos can be considered as perceptual quality, degree of presence, acceptability, and cyber-sickness. [17]. A QoE study [18] on immersive gaming experiences with Oculus Rift observed that in general the higher level of perceived immersion also demonstrated to be directly proportional to perceived usability. Also, authors in [19] explored the implications on the processing of rendered video to prevent cyber-sickness in a first-person omnidirectional video. They showed that sudden changes of the speed in the translational camera motion and excessive vibration can cause cyber-sickness.

Another important recent stream of activities in this direction is the real-time estimation of viewers’ QoE levels exploiting physiological signals [20], [21], [22]. This category of activities aims at tracking singular or combinations of physiological responses (e.g., heart rate, electro-dermal activity, and electroencephalogram) in the course of immersive media experience. By employing regression and classification models on the tracked data and referring to the real QoE values reported by the participating users, the authors aim to generate automatic and non-intrusive QoE prediction models.

2.4 Challenges with User Interface Design
User interface has acquired three levels of interaction: graphical user interface (GUI), tangible user interface (TUI) and radical atoms [23], [24]. Another level is through Brain-Computer Interfaces (BCI) [25], which can be translated into three categories related to control: active, reactive and passive [26]. Challenges with adaptive interactive experiences in BCI are the sense of control, which can affect their immersion [25].

Other aspect to consider is the use of multi-sensory technology from different sources. In VR, it is crucial to understand the position of the camera and the angles where the main information is displayed. Users could also interact with wireless devices that function like interactive gloves [27]. Although this could be an alternative, users have reported that in a daily basis it can be extremely tiring, which shows that studies in this area still require more investigation. In VTEs, it is also expected that the interface design would function as a conversational tool with all users with no disruptions. Thus, a seamless interaction is necessary.

2.4.1 Challenges with metaphors and conceptual models

With the aim to investigate the relationship between real and digital worlds, Jetter et al. [28] developed the concept of Blended Interaction, which briefly explains how digital structures are blended in terms of analogies, schemas and metaphors. For example, the save button is a blended structure since it combines a floppy disk illustration with the action of saving; this means that users who have never experienced saving a file in a “real” floppy disk can still understand the saving metaphor. Challenges with QoE of adaptive experiences in this case could be related to the translation of the same metaphor and conceptual models to different users.

2.4.2 Challenges with usability and cyber-sickness

One of the main concerns about usability within immersive environments is cyber-sickness [29]. In scenarios in which the user looks at immersive omnidirectional videos (ODV) from a first-person perspective, it is possible that users might experience cyber-sickness since the video was recorded from another person’s PoV and then translated to a larger screen. In order to address this issue, researchers have designed a prototype, which creates two design dimensions: one from the perspective of the wearer, who records the interaction, and the other from the spectator’s point of view [30]. This “re-experience” is similar to the virtual transcendence; however, challenges related to this approach are the synchronisation of motion and the viewing angle of the viewer and the spectator. Another challenge related to cyber-sickness is identifying if the user has the symptoms before interacting with the system or not. This anticipation has led to the development of heuristics designed for usable VR environments. Challenges related to this are usually technical, involving stable FPS, realistic motion, minimisation of rotation, having a static head-up display (HUD) and introducing interactions in small chunks, so users can get immersed gradually in the environment [31]. Another concern is that if the environment has little connection with a natural context, users might feel lost and disoriented [32]. A good practice could be the use of the Point & Teleport technique to make users move within the VR; however, challenges regarding cyber-sickness might persist, regardless the orientation of the user [33].

3 VIRTUAL TRANSCENDENCE EXPERIENCE FRAMEWORK (VTEF)

Virtual Transcendence Experience (VTE) allows users to see the world with the eyes of others but through participatory actions, enabled by immersive and multi-sensory technology. Thus, with the aim to develop such experiences, we have considered the possible technical and design challenges that could emerge while developing VTEs. With that, we have designed the Virtual Transcendence Experience framework (VTEf) in order to guide the development of VTE applications. As shown in Fig.1, the VTEf embraces different layers of user experience in multi-sensorial environments. Its main components are:

- **Actors**: a) A camera wearer/sender of PoV; b) Spectator(s)/participant(s);
- **Design aspects**: a) Wearable omnidirectional camera (e.g. a regular omnidirectional camera fitted at the top of a stick that is part of the cloth, or a series of cameras fitted around a headband worn by the person sending the PoV); b) User interface for the spectators; c) Metaphors and conceptual models; d) Usability and cyber-sickness;
- **Technical aspects**: a) Capturing of the omnidirectional media with high visual accuracy; b) Real-time processing (compression/upload); c) Content-aware networking considering streaming optimisation

**Metadata**: All information accompanying the multimedia (e.g., time, engagement levels, information about usage, emotions)

**Points of interaction** (e.g. multi-sensory inputs, including speech and gesture-based interfaces)

The main scenario of the anticipated VTEf encompasses the following steps: a) The camera wearer captures an environment in 360° using a mobile camera system; b) Connected to the system with a VR headset and the user interface, a spectator can transport their visual consciousness to the event using controls and choosing special items (e.g., images, affective memory); c) As the spectator collects the elements from the event, the process is followed and discussed by a network of other
spectators/participants, who comment on the decision-making process and wait for their own participation or participate in real-time; d) Metadata is gathered from the interactions between multiple users; it could also function as a key performance indicator (kpi), in order to assure QoE. The co-participation in the event mirrors the spectator and artist, investigating a new expositive category that can be translated to other environments, such as education, sports or entertainment (e.g. musical concerts, movies, etc.). In order to visualise the application of the VTEf, we considered two likely scenarios. For the first scenario, a real-time interaction was considered, in which multiple users can engage in one task, transcending and sharing experiences in real time. In the second scenario, we explored the influence of pre-recorded experiences. This could be very helpful in building immersive environments, like games or movies in the guidance of pre-structured narratives. This could also be used in order to navigate through memories. For example, users could record their experiences and revisit them after years, living their memories again and possibly looking for different alternatives from that past reality. These scenarios are explained in the following sections.

3.1 Scenario 1 - Hiking: live content

The environment in this exemplary scenario considers the action of hiking. This action was chosen due to its unique points of interaction, which could be remote, since the wearer/sender would be in a remote, open area that is not necessarily well connected. This scenario provides two points of interaction: hiking from the visual perspective of the wearer/sender and the spectator(s). In this scenario, users need to make decisions and interactions (e.g. the wearer can take actions while hiking, whereas spectators can give suggestions to the wearer); spectators need to interact with the environment as well. There is more freedom in this scenario and more communication between the spectators (who may also become participants) and the wearer. The wearer might choose to pose questions to the live spectators in order to improve his/her experience, for example, by asking where to go, or for tips to survive, etc. This would evoke a sense of collaboration from the perspective of the spectators. There is a direct verbal interaction between the wearer and the spectators to guide and participate in the wearer’s action. Depending on the complexity of the environment of the action, the wearer may also wear Augmented Reality (AR) goggles in addition to the mobile camera setup, which would allow him/her to visualise key interaction elements by the spectators (i.e., not only hear) as an overlay on the real sight through the mountains (see Fig. 2).

**Metadata**: perceived presence, perceived usability, emotions, sense of control, engagement levels

**Design challenges**: adjusting the video stabilisation of the wearer/sender and the spectators through different camera angles in order to avoid cyber-sickness; user interface interactions (e.g. voice commands, gestures)

**Technical challenges**: Local mobile processing and compression of omnidirectional video, live upload of the video to the cloud-based content management entity for distribution, content-aware processing for bandwidth-efficient mass distribution (e.g., viewpoint streaming).

![Virtual Transcendence Experience framework (VTEf)](image)
3.2 Scenario 2 – Interactive Movie: pre-recorded experiences

This scenario is not real-time since a user could record his/her experience through different viewing angles. For example, in a movie, actors could record their own PoV through 360° cameras. After recording, spectators could be able to switch between first person views at any point in time. With that possibility, spectators could immerse themselves more effectively to the anticipated story by feeling the mood of the actor with more accuracy (see Fig. 3). Those experiences can also be non-linear with multiple story branches depending on the viewer’s choices. In this case, different from point-and-click adventure games, participants might influence each other’s experiences in a non-linear way; it could function like a multiplayer choice story game, in which spectators can choose who they are based on the actors’ PoV. This could allow spectators to become the actors from the movie and to be able to choose their actions within the movie narrative.

**Metadata:** perceived presence, perceived usability, emotions, sense of control, engagement levels

**Design challenges:** adjusting the point of view of the wearer/sender and the spectators through different camera angles in order to avoid cyber-sickness; ensuring a minimal usability control that could be perceived with no interruptions from each recorded experience; permitting users to have a sense of control on their own experience without the limitations of the recorded experience; keeping users engaged at the same level free from interruptions; maintaining the quality of experience for each user; hiding multiple 360° cameras used in the shooting of a scene.

**Technical challenges:** Different from the former scenario, this scenario doesn’t necessarily impose challenges from live content processing and compression perspectives. The main challenge in this perspective is the availability of multiple views for streaming and the need for bandwidth optimisation in the process of channel switches.

4 DISCUSSIONS AND FUTURE WORK

The Virtual Transcendence Experience framework (VTEf) introduces a new concept and paradigm for the implementation of this sense of transcendence, mediated by multi-sensory technology. The VTEf has five main components: actors, metadata, points of interaction, design and technical challenges. The VTEf provides a holistic overview of elements that affect Quality of Experience (QoE) in immersive environments mediated by multi-sensory inputs.

The first discussion about the VTE is about engagement protocols. We may imply enthusiasm not just in the moment devices are concealing contextual realities in favour of virtual ones but also the motivations behind specific behaviours; gamification, as practice and meta-interactive project, could be one of those protocols. Gamification is the use of game elements in non-gaming situations [34]. The term is applied to certain engagement perspectives that demands users to attend otherwise boring activities, moved by some sort of social realisation and fun core principles. Despite the fact that the success of VTE does not solely rely on sensorial fidelity more advanced stages will require a greater feel of reality. These range from formulating sophisticated representation and compression frameworks for not only omnidirectional video, but also for emerging volumetric video formats, to deploying new cloud and network-edge assisted content processing mechanisms accounting for a diverse ecosystem of user display devices of varying capabilities. Considering these, in our future work we consider to look at:

**People with disabilities interacting in each scenario:** How could someone with a certain sensory impairment have a transcendent experience as proposed? What would be the limitations and adaptations needed for each multi-sensory technology in order to capture enough information to be shared with other users?

**Security and privacy in shared experiences:** Parts of the information shared between users would carry personal traits in multiple modalities, which could be used in digital identity theft if
skilfully processed. Stronger encryption protocols and on-source de-identification mechanisms could prevent such attempts.

**Diversity in cultural background and locations:** How could people from different cultures and locations interpret and respond to the interactions of other users in a VTE environment? What are the main guidelines of intuitive and effective interactions?

**Ethical issues:** How can the distribution of experiences with bad intentions be prevented in real-time?

**Human consciousness in digital environments:** What is the definition of digital consciousness? What are the theoretical and practical implications of this type of consciousness?

**Gamification:** How could the well-defined principles of gamification be incorporated within a practical VTE framework to have a broader impact?

**Quality of experience:** Understanding and developing models that could solve cyber-sickness issues in multi-user environments for VTE.

Our next step is to implement the described VTEf. There has been a growing interest of the incorporation of VR in events that lead to a democratisation of the experience. Immersive environments can also evoke memories that can become part of the individual’s “real” memory [35]. Thus, this illustrates the potential of VTE to transform people’s memories. There is also a transformation of consciousness, which permeates the relationship among artists, designers, audiences, artworks and people, showing a need for new rearrangements and patterns for VR as a medium [36]. These social and theoretical conditions address some questions related to new experiences. It is expected that the conception of VTE as object of investigation can accollade research on theory and practices. We also expect that this study will be of interest to artists, engineers and designers alike.

**REFERENCES**


