Foundations of Interaction in The Virtual Reality Medium

Danielle Marie Olson[1], Elisabeth Ainsley Sutherland[2], Cagri Hakan Zaman[3], D. Fox Harrell[4]

[1] Massachusetts Institute of Technology (MIT) Computer Science Artificial Intelligence Laboratory (CSAIL), Cambridge, MA
dolson@mit.edu
[2] Mediate VR, Cambridge, MA
ainsleys@mit.edu
[3] MIT CSAIL, Cambridge, MA
zaman@mit.edu
[4] MIT Comparative Media Studies Program and CSAIL, Cambridge, MA
fox.harrell@mit.edu

Synonyms

Artificial intelligence; Conceptual blending; Immersive design; Immersive systems; Interactive design; Spatial cognition; Virtual reality

Definitions

The *medium* of virtual reality (VR) can be defined as the unique set of practices and novel affordances that emerge when using VR and is distinct from work built for screen-based or other media.

1. Introduction: VR, a Technology and a Medium

VR systems bring together a family of technologies to create a convincing computer-graphical space around the user. These spaces can range from the photorealistic to the abstract and can include a host of objects, interactions, and effects. While many VR systems are primarily visual (the focus here), others can involve other sensory modalities such as haptic or olfactory feedback. Better understanding VR as a technology and as a medium will aid in distinguishing between media technologies, genres for using such technologies, and particular works. *Technologies* here refers to hardware and software that are researched, developed, and used. A *genre* is a style of using some input/output capabilities, and a *work* is an instance of a genre" (Goguen & Harrell, 2005). The *medium of VR* can then be defined as the unique set of practices and novel affordances that emerge when using VR. Genres of VR will continue to emerge as conventions are innovated and established. Works created that uses the *medium* of VR will be distinct from work built for screen-based or other media. This understanding of VR is especially important for understanding its role as a medium for works of computer-based art (Bates, 1992) such as videogaming and interactive narratives.

The *technology of VR* can be defined as any combination of devices and software that produces a sense of virtual *exteroception* and *proprioception* in a user. Exteroception refers to perception of an external environment. Proprioception refers to perception of one's own body (Fotopoulou, 2014).

These technologies most typically include: (1) real-time computer graphics, (2) dynamic visual interaction including: a panoramic image space with positional and orientation tracking, (3) sensory elements including: stereoscopic vision, a head-mounted viewing device, and spatial audio, and often (4) motor elements including peripheral devices such as handheld controllers. These elements combine to produce a sense of *virtual exteroception*, providing a sense of a tangible virtual environment, and of *virtual proprioception*, providing the sense of a tangible virtual self. VR can include less interactive work, such as 360-degree film, or highly-interactive work, that could include peripherals such as gloves equipped with sensors and treadmills to simulate physical movement.

This section introduces definitions of these various technologies, discusses their key aspects, and demonstrates examples of how each—individually and in combination—provide unique impacts on creating VR systems.

1.1. Real-Time Computer Graphics

Real-time computer graphics are familiar to most users from media forms such as computer games and animated feature films. However, while animated films often use pre-rendered computer graphic images (CGI), VR environments and models are rendered as-needed in response to user input such as gaze, position, movement, and gesture while users explore and interact with the virtual environment. VR experiences can be built from 3D models and environments implemented synthetically using CGI software. Alternately, VR imagery can be captured from the physical world through techniques such as photogrammetry and videogrammetry to build 3D models from photos and video. Finally, VR environments can be created using techniques integrating both synthetic modeling and physical world capture-based modeling. For example, in *Hospital with One Entrance,* artist Deniz Tortum uses a laser scanning device to capture the physical dimensions of a hospital's operating room and then imports this data into VR software to program interactivity (Tortum, 2016).

Key Graphics Technology Considerations

Computer graphics used in VR require a frame rate of up to 75 frames-per-second (fps) in order to ensure smooth movements and avoid noticeable *judder* to the user. Systems that fail to provide high quality resolution and frame rate cause end users to perceive *judder* and may result in motion sickness. Given these requirements, VR computer graphics systems must be equipped with powerful enough graphics processing unit (GPU) capabilities and have low system *latency*, or time between user input and updates in displayed graphics.

Key Graphics Design Considerations

Real-time graphics provide both constraints and affordances for designers. Latency and judder are examples of *constraints*.

System designers must negotiate these constraints, for instance *tethering* consumer VR headsets to powerful GPU-packed computers using cables. These challenges have inspired research and development efforts focused on designing more efficient hardware to enable higher-quality VR experiences on affordable mobile devices with less processing power. Given that VR systems enable the design of nuanced interfaces and experiences tailored to individuals based on their gaze, position, movement, and gesture over time, overcoming power and latency issues enables more impactful uses of the medium in educational, training, medical and therapeutic contexts (Chu and Cuervo, 2016; Lai, Z. et al., 2017).

1.2. Orientation and Positional Tracking

Users can navigate their visual environments more freely in VR than in other screen-based media. This is due in part to the combination of a panoramic virtual environment with orientational and positional tracking. Positional tracking and orientation tracking refer to the ability of VR systems to track user movement in the physical world, and translate that movement to the virtual environment. The degrees to which VR systems support translating user movement from the physical world to the virtual environment is referred to as the "degrees of freedom" (DOF) of the system. *Orientational tracking* means that the system tracks the rotational movement of a viewer's head, thus allowing a user to look freely in any direction in the virtual environment. All current VR systems enable orientational tracking. *Positional tracking* follows the user's translational motion, and positions the user in a 3-dimensional environment. Higher-end VR systems enable positional tracking. These types of user tracking, combined with the complete panoramic image space and a wide field-of-view create a visual environment for the user that is closely connected to natural perceptual activities. Moreover, in VR, the CGI spatial environment also implies shape, volume, location, and physics, encouraging users to look under, around, or through elements in space (Smith 2017).

Key Tracking Technology Considerations

Positional tracking in VR is generally achieved through sensors which may be internal or external to the head-mounted display or controller. Some hardware also uses embedded sensors to track this kind of motion. *Orientation tracking* is typically tracked using a combination of accelerometers, gyroscopes and magnetometers embedded into the hardware. For positional tracking, laser- or camera-based sensors can be used. Each of these technologies has tradeoffs for design: for example, camera-based systems can be more accurate, but require processing power and connections sufficient to avoid latency and have additional privacy considerations.

Key Tracking Design Considerations

Positional and orientation tracking activate the visual space outside the user's immediate field of view, allowing users to glance and reach at things in the periphery of their vision, and to lean or reach towards and away from elements in the scene. *Intentional looking* and *direct address* are two techniques for accounting for this in design. For example, consider a VR experience in which the spatial imagery around the user initially goes pitch black, and then lights flicker on a little bit later. The user might react by looking curiously around using and investigating the full panorama image space to see what changed. The user's investigation of the space is motivated and intentional, more active than receptive. *Direct address* describes the "two-way process of a user both seeing and being seen" by other characters within the environment, and by the virtual environment itself (Sutherland 2015). The tracking devices allow the system to change in response to a user's attention: for example, objects appearing only when the user is gazing elsewhere.

Placement of objects within the environment given the constraints of the positional tracking system and the *tethered* display is of key concern when designing for HCI in VR. For example, while activities such as "crouching" or "crawling" in VR games can provide an exciting sense of immersion, they can easily result in user fatigue if repeated too often during the experience.

User tracking and spatial design also has emergent effects in multi-user environments. For instance, nSpace is a project exploring collaborative aspects of VR for design tasks. This system uses a special sensor for hand tracking to visually represent the user's hands. This enables interaction with user interface components that exploit the 360 degrees visual representation. This enables users to move through virtual environments and provide more specific and relevant feedback on objects and instructions to other users using *subjective*—in relation to a user's body—rather than *objective* language, such as the absolute position of an element on a screen (Zaman et al., 2015).

1.3. Stereoscopic Vision, Head-Mounted Displays, and Spatial Audio

The combination of stereoscopy and head-mounted displays (HMDs) connects VR systems closely with natural visual perception. Stereoscopic vision in VR produces the illusion of depth and three-dimensional space and is achieved by displaying parallax angles of images through dual lenses. These dual images originate from different perspectives and slightly overlap, such that users fixating binocularly on a point will perceive elements images on the same relative coordinates as a single object (Tam et al. 2011).

HMDs must also be equipped to deliver spatial audio that is synchronized with the visual experience. Finally, many companies and researchers are working towards enabling eye-tracking in consumer VR systems.

Key Stereoscopy, HMD, and Spatial Audio Technology Considerations
As discussed above, most consumer VR headsets are *tethered* to powerful GPU-packed computers with high power consumption and thermal output. However, active work is being done to develop more efficient HMDs that can deliver higher-quality VR experiences on more efficient and wireless devices. HMDs can also deliver spatial audio which simulates natural human localization techniques using sound clues. The widespread use of consumer speech recognition systems, work in 3D binaural sound reproduction and spherical microphones are advancing the quality of spatial audio capture and playback (Jarrett 2017).

While eye-tracking adds an additional computational load to systems, it can enable *foveated rendering* techniques, which are processes for reducing the workload on the system by high-quality rendering of the graphics the user is looking at and lower-quality rendering of the visuals in the user's peripheral vision (Guenter et al., 2012; Padmanaban, 2017).

Key Stereoscopy, HMD, and Spatial Audio Design Considerations
Stereoscopic depth enables VR to convincingly situate users in CGI spatial environments, resulting in a sense of a physical relationship to objects and characters.

Tethered HMDs limit users' range of translational motion, but research and releases to the consumer VR market are trending towards improving phone-based, wireless, all-in-one HMDs. As barriers to entry for VR HMDs become lower, HMD design must anticipate and address the "brick in the face problem" which results from the opaque quality of the headset.

Given that the eyes provide a crucial means of nonverbal communication in social contexts including gaming, eye tracking technology can be used as input to achieve better customization. Other VR experiences such as *The Enemy*, a journalistic VR artwork by photographer Karim Ben Khelifa addressing global conflict (Kennedy, 2016; Lacey, 2016), allow for dynamic changes to the experience's narrative (e.g., events and dialogue) and staging (e.g., lighting and mise en scène of the virtual environment) based on the users' embodied input to the system using artificial intelligence. For example, features including users' translational motion, head motion, directional orientation, and proximity to each other and to virtual characters are used as proxies to track users' attention, nervousness, and biases (e.g., asymmetries of attention and nervousness). This real-time user tracking triggers feedback by the system including changes to behaviors of the non-player characters, appearance of the users' avatar, voice-overs, and the virtual cloud cover (thereby impacting the lighting). Eye-tracking capabilities can also enhance the evocative potential of this genre of VR experiences. In social VR, eye-tracking technology has been used in conjunction with computer vision algorithms which align and blend a 3D face model with a camera's video stream of the user (Frueh et al., 2017). These strides demonstrate how artificial intelligence will help to enhance connection and interaction in multiple-user VR scenarios and third-party gameplay viewers.

Furthermore, spatial audio can be a powerful tool to present sounds from any direction, control user attention, give users cues on where to look, and provide an immersive VR experience. Spatial audio in VR has the potential to be applied as a powerful design tool for evaluating planned architectural designs in combination with soundscapes prior to physical construction (Echevarria Sanchez et al., 2017). Tools such as *Mediate VR* which enable the evaluation of space and soundscape designs through remote, asynchronous, voice-driven collaboration. Forward-looking designs of HMDs must be able to stream and play spatial audio in real-time, with tools such as *TheWaveVR* introducing the concept of social platforms which host immersive VR music concerts.

1.4. Peripheral Devices

VR peripheral devices enable additional forms of user input, output, and interaction in the immersive environment. Although not the focus of this article, it is important to consider key types of such devices.

Key Peripheral Device Technologies

Peripheral device input can include the positional tracking of hands and a variety of controller-like inputs. Peripheral device technologies includes *haptic technologies*, which describes a form of human-computer interaction involving touch. These peripherals range from controllers or simple touch screen devices with button-based controller paradigms, to a joystick, remote, or mouse, as well as devices which enable gestural interaction ranging from the 1990s “data glove” (Premaratne, 2014), treadmills, to *Leap Motion, Microsoft Kinect,* or the *Myo* armband. Peripherals can also include biometric devices such as wireless wristbands that monitor real-time physiological signals for affective computing interfaces, EEG-based biometrics for brain-computer interfaces and more. Sensory-output based peripherals may also include vibrating floors and mats, electrical muscle stimulation for muscle-computer interfaces, and olfactory output devices.

Key Peripheral Device Design Considerations

In combination with the technologies presented in the previous sections, peripheral devices present additional design opportunities and challenges. For example, while handheld peripherals and gloves can enable users to manipulate virtual objects, developers must account for the effect of users seeing their own hands. Systems that enable standing in combinations with tracked handheld peripherals affords the user the ability to be able to reach into virtual environments and do things, requiring nuanced handling of embodied input.

By leveraging biometric devices that provide real-time physiological data or manipulate the sensations of the user’s body, an additional layer of immersion is added to the experience. For example, recent research has used muscle stimulation through gentle electrical impulses as a new approach to rendering the haptics which afford the repulsion of a wall or gravity pulling down the weight of a heavy box (Lopes et al., 2017).

2. Creating VR Experiences

2.1. Introduction

This section presents two theoretical approaches that can underpin and motivate the design of VR experiences. The first, constraints and affordances, is crucial to organize approaches to the many new interaction paradigms and sheer variety of design choices. The second, conceptual metaphors and blends, suggests a way to bring metaphorical thinking into the virtual environment in a way that can be used both for creative purposes and for efficient design.

2.2. Constraints and Affordances in VR

The technologies that enable the possibilities of immersive VR environments also present both *sensory constraints* and *physical/motor constraints. Sensory constraints* are limitations in being able to provide visual information to the user. This may result in a lack of framing around the content of the experience, or a lack of self-representation of the user's body in space. Many VR systems focus on visual and auditory constraints. *Physical/motor constraints*, which are limitations in the physical capabilities of the hardware design, setup, of ability for users to move in the physical world. This may result in a lack of freedom in users' movements and interactions with their physical surroundings, which may result in safety risks.

The psychologist James Gibson provided a useful term when discussing what technologies enable or constrain: he defines an *affordance* as what the environment offers or furnishes to the user (Gibson, 1977, 1979). In these terms, VR is a unique medium with the potential for novel interaction mechanisms given the particular constraints and affordances of the technology. In VR, the designer has a close hand in creating and varying the *perceived affordances* presented to individuals within the experience (Norman, 1999a; Norman, 1999b).

The design choices of developers, in conjunction with the aforementioned technological elements of the medium, also shape the *sensory affordances* of VR systems. Especially of note are their *visual*, *tangible*, and *gestural affordances*. *Visual affordances* are visual cues which invite users for action in the form of visual guides or floating user interface (UI) elements. For example, in any given VR experience, users can have either *no body*, *an object instead of a body*, a *partial body*, their *own body*, *another person's body, or even multiple bodies*. *Tangible affordances* are provided by peripherals which enable *haptic* interaction and cues in the environment, such as different vibration patterns being transmitted to the users' fingertips through controllers or point and click devices. *Gestural affordances* provide cues which invite users for gestural interaction, such as cues to virtually touch or pick up objects.

2.3. Conceptual Metaphors and Blends in VR

VR experiences and tools can be thought of as a *performance* between users and their virtual environment (Laurel, 2014). One potentially effective way to design such mediation is to build *interface metaphors*, a set of visuals which build on existing notions and actions to facilitate meaningful and intuitive interactions in VR. Interface metaphors are in turn grounded in conceptual metaphors, which are mappings between ideas (including mental images) that are grounded in sensory-motor action (Lakoff and Johnson, 1980). For example, forefingers and closed hands are typically associated with pointing and grabbing, hence, selecting virtual objects may be achieved by virtually pointing and gathering virtual objects may be accomplished by virtually grabbing.

They key to designing strong interface metaphors is keeping them intuitive to the user. Users should be able to infer coherent actions based on these metaphors unless they are explicitly designed for creative effects (for instance, an interaction mechanism that is difficult to perform can be useful to represent a difficult action in a VR game). This can be very challenging in such an immersive environment, given that the user can do and perceive many things. Anticipating the user's inferences is important, and by taking an environmental approach to design one can begin to think this way. Likewise, for creative projects, anticipating users' expectations can allow for the creation of surprising and meaningful narratives.

2.4. Conclusion

Virtual reality is an important set of technologies and media. This is because, as the name indicates, its technologies seek to make use of many of the modes of interaction that humans have in the "reality" of the physical world while enabling the new "synthetic" possibilities of the virtual. It has also been said that objects in the physical "real" world can lose their reality due to layers of mediation (Baudrillard, 1995) and experiences in virtual environments can have real physical world impacts ranging from bullying to discrimination (Harrell and Lim, 2017). In light of these complexities, VR's technologies and their constraints and affordances must be account for while creatively making, subverting, and adapting them for impactful and powerful experiences.

Cross-References

- Enhancing User Immersion and Virtual Presence in Interactive Multiuser Virtual Environments through the Development and Integration of a Gesture-Centric Natural User Interface Developed from Existing Virtual Reality Technologies https://link.springer.com/chapter/10.1007/978-3-319-07857-1_72
- Minimal Virtual Reality System for Virtual Walking in a Real Scene https://link.springer.com/chapter/10.1007/978-3-319-40349-6_48
- An Overview of Virtual Reality/Virtual Environments for Education and Training https://link.springer.com/chapter/10.1007/978-1-4899-0038-8_1
- Early Prototyping of 3D-Gesture Interaction within the Presentation-Gesture-Dialog Design Space https://link.springer.com/chapter/10.1007/978-3-319-07230-2_2
- A Replicated CSCW Framework for Multi-user Interaction in 3D Collaborative Design https://link.springer.com/chapter/10.1007/978-3-540-73111-5_117
- Immersion in Cardboard VR Compared to a Traditional Head-Mounted Display https://link.springer.com/chapter/10.1007/978-3-319-39907-2_25
- Research on Lexical Items Presentation Based on Conceptual Metaphor with Development of Information https://link.springer.com/chapter/10.1007/978-3-642-23339-5_65
- Gaze-Contingent Volume Rendering https://link.springer.com/chapter/10.1007/978-3-540-75412-1_9
- Object-Based Spatial Audio: Concept, Advantages, and Challenges https://link.springer.com/chapter/10.1007/978-1-4614-8373-1_5

References

1. Bates, J.: Virtual reality, art, and entertainment. Teleoperators and Virtual Environments archive Volume 1 Issue 1, Winter 1992 Pages 133-138 (1992)
2. Fotopoulou, A.: The virtual bodily self: Mentalisation of the body as revealed in anosognosia for hemiplegia. *Consciousness and Cognition*, *33*, 500–510 (2015)
3. Tortum, H.D.: Embodied Montage: Reconsidering Immediacy in Virtual Reality (2016)
4. Chu, D., Cuervo, E.: FlashBack : Immersive Virtual Reality on Mobile Devices via Rendering Memoization, *20*(4), 291–303 (2016)
5. Lai, Z., Hu, Y. C., Cui, Y., Sun, L., & Dai, N.: Furion: Engineering High-Quality Immersive Virtual Reality on Today's Mobile Devices. *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking - MobiCom '17*, 409–421 (2017)

6. Jarrett, D.P.: Theory and Applications of Spherical Microphone Array Processing, Switzerland (2017).
7. Guenter, B., Finch, M., Drucker, S., Tan, D., Snyder, J.: Foveated 3D graphics. ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH, Volume 31 Issue 6 Article No. 164 (2012)
8. Kennedy, R.: Meeting 'the Other' Face to Face. The New York Times. https://www.nytimes.com/2016/10/30/arts/design/meeting-the-enemy-face-to-face-through-virtual-reality.html (2016)
9. Lacey, S.: Face to face with "The Enemy". http://news.mit.edu/2016/face-face-with-the-enemy-karim-ben-khelifa-1202 (2016)
10. Frueh, C., Sud, A., & Kwatra, V.: Headset Removal for Virtual and Mixed Reality. *In Proceedings of SIGGRAPH '17*. (2017)
11. Premaratne, P.: Historical Development of Hand Gesture Recognition. In: Human Computer Interaction Using Hand Gestures. Cognitive Science and Technology. Springer, https://doi.org/10.1007/978-981-4585-69-9_2 (2014).
12. Lopes, P., You, S., Cheng, L.-P., Marwecki, S., & Baudisch, P.: Providing Haptics to Walls & Heavy Objects in Virtual Reality by Means of Electrical Muscle Stimulation. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 1471–1482 (2017)
13. Gibson, J.J.: The Theory of Affordances, Hillsdale (1977)
14. Gibson, J.J.: The Ecological Approach to Visual Perception, Boston (1979)
15. Lakoff, G. and Johnson, M.: Conceptual Metaphor in Everyday Language. The Journal of Philosophy Vol. 77, No. 8 (Aug., 1980), pp. 453-486 (1980)
16. Laurel, B.: *Computers as Theatre*, Crawfordsville (2014)
17. Padmanaban, N., Konrad, R., Stramer, T., Cooper, E. A., & Wetzstein, G.: Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays. *Proceedings of the National Academy of Sciences*, *114*(9), 2183–2188 (2017)
18. Echevarria Sanchez, G. M., Van Renterghem, T., Sun, K., De Coensel, B., & Botteldooren, D.: Using Virtual Reality for assessing the role of noise in the audio-visual design of an urban public space. *Landscape and Urban Planning*, *167*(May), 98–107 (2017)
19. Tam, W. J., Speranza, F., Yano, S., Shimono, K., & Ono, H.: Stereoscopic 3D-TV: Visual comfort. *IEEE Transactions on Broadcasting*, *57*(2 PART 2), 335–346 (2011)
20. Goguen, J. A., & Harrell, D. F.: Foundations for Active Multimedia Narrative. *Igarss 2014*, *1*(1), 1–5 (2014)
21. Sutherland, E. A.: Staged Empathy: Empathy and Visual Perception in Virtual Reality Systems Empathy and Visual Perception in Virtual Reality Systems (2015)
22. Thiel, T.: Where Stones Can Speak: Dramatic Encounters in Interactive 3-D Virtual Reality. *Third Person: Authoring and Exploring Vast Narratives*, 153–178 (2009)

23. Smith, L.: 360 photos vs. VR. Bevel. https://bevel.space/news/2017/8/1/is-it-virtual-reality-360-photos-vs-vr (2017)
24. Using Virtual Reality for assessing the role of noise in the audio-visual design of an urban public space https://doi.org/10.1016/j.landurbplan.2017.05.018
25. Schröter, J.: 3D: History, Theory, and Aesthetics of the Transplane Image. International Texts in Critical Media Aesthetics, New York (2014).
26. Johnson, M.: The body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason, Chicago (1987)
27. Johnson, M.: The meaning of the Body: Aesthetics of Human Understanding, Chicago (2007)
28. Harrell, D.F.: Phantasmal Media: An Approach to Imagination, Computation, and Expression, Cambridge (2013)
29. Jenkins, H.: Game Design as Narrative Architecture. First Person: New Media as Story, Performance, and Game. Eds. N. Wardrip-Fruin and P. Harrigan. Cambridge, Mass. MIT Press, 2004
30. Galloway, A.R.: The Interface Effect, Cambridge (2012)
31. Grau, O.: Virtual Art: From Illusion to Immersion, Cambridge (2003)
32. Norman, D.A.: Affordance, conventions, and design. Magazine interactions Volume 6 Issue 3, (1999a)
33. Norman, D.A.: Affordance, conventions, and design. Magazine interactions Volume 6 Issue 3, May/June 1999 Pages 38-43 (1999b)
34. Zaman, C.H., Yakhina, A., Casalegno, F.: nRoom: an immersive virtual environment for collaborative spatial design. CHIuXiD '15, 10-17 (2015)
35. Baudrillard, J.: Simulacra and Simulation, 1st edition, 17th printing edition, Ann Arbor: (1995)
36. Harrell, D.F., Lim, C.: Reimagining the Avatar Dream: Modeling Social Identity in Digital Media. Magazine Communications of the ACM Volume 60 Issue 7, July 2017 Pages 50-61 (2017)