Audio-Visual Experience Design for Virtual Reality Environments: Postprint

Authors: Ouyang Shengjie

Date: 2023-10-08T00:00:00+00:00

Abstract

This paper integrates the experiential characteristics of virtual reality with the mechanisms of human visual and auditory sensory systems to analyze the fundamental elements of audio-visual experience design in virtual reality environments, and derives guiding principles for virtual reality audio-visual system design from this analysis.

Full Text

Abstract

This paper integrates the experiential characteristics of virtual reality with the mechanisms of human visual and auditory sensory functions to analyze the fundamental elements of audio-visual experience design in virtual reality environments, deriving guiding principles for VR audio-visual system design.

Keywords: Virtual Reality; Audio-Visual Experience; Immersion
CLC Number: TP391.9
Document Code: A
Article ID: 1671-0134(2018)05-091-03
DOI: 10.19483/j.cnki.11-4653/n.2018.05.028

---

1. Characteristics of Audio-Visual Experience in Virtual Reality

The concept of Virtual Reality (VR) originates from humanity's exploration and simulation of the real world, representing our enduring pursuit of recreating reality through artistic and technological means. Initially a sophisticated technology primarily used for simulation training in military and aerospace applications, VR has gradually become accessible for everyday human use. Today, people can experience the enchanting, dreamlike qualities of VR worlds through head-mounted displays or dedicated VR experience zones, with lifelike audio-visual effects that transport users into seemingly alternate realities.

In 1993, Heim proposed that "VR is something that is real in effect but not in fact," delineating seven characteristics of VR: simulation, interaction, artificial reality, immersion, telepresence, full-body immersion, and network communication. In 1994, Burdea et al. published *Virtual Reality Technology*, summarizing VR's essential features with three "I"s—Immersion, Interaction, and Imagination—which also constitute the experiential demands of VR audio-visual systems.

Immersion is directly proportional to the degree of simulation of the real world; higher fidelity results in less cognitive interference and stronger immersion. As previously noted, vision and hearing are among the most critical components of VR systems, with all authentic visual and auditory information that humans can perceive serving as an important reference for achieving VR immersion. Furthermore, immersion reflects the user's sense of being contained within and integrated into the VR environment, emphasizing from a perceptual standpoint that VR environments should constitute a three-dimensional space. Consequently, VR audio-visual systems must satisfy human needs for spatial information regarding depth and orientation.

VR immersive experience derives not only from three-dimensional spatial perception but also from interactive behaviors between users and the environment. VR's interactive characteristics can engender a sense of immersion within the behavioral system. This interactivity refers not only to providing feedback for users' conscious operations but also to simulating real-world changes in response to unconscious behaviors. For instance, when a user turns their head, the display adjusts accordingly to simulate natural field-of-view changes.

Imagination represents VR's manifestation of "originating from reality yet transcending it," emphasizing the designer's dominant role. Virtual reality is essentially a pathway for humans to seek sensory simulation and replication, a process that also involves transcending the physical properties of the real world. From an audio-visual perspective, VR "allows us to observe and track the real world with unprecedented richness of detail and depth," and through technological means to "arrange" people's audio-visual activities, thereby further satisfying users' aesthetic demands and creating emotional resonance.

2. Audio-Visual Experience Design for Virtual Reality

Guided by VR's three "I" characteristics, audio-visual experience design should proceed from two perspectives: the physiological mechanisms of human perceptual activity and experiential demands. The visual component primarily encompasses three aspects—visual fidelity, spatial vision, and position/orientation tracking—while the auditory component mainly involves scene-accompanying sound and directional information.

2.1 Visual Component

VR's simulation of the real world includes target objects, relationships and interactions between objects, and the laws governing their development and changes. Correspondingly, VR visual systems for human vision can be divided into visual fidelity, spatial vision, and position/direction tracking.

Visual fidelity refers to the simulation of real-world visual imagery through VR technology. In terms of visual information construction, static and dynamic forms of simulated objects and environmental landscapes constitute the main content of VR imagery, requiring reasonable and distinguishable digital modeling. Regarding visual information transmission, the imaging capabilities of head-mounted displays also affect visual fidelity, primarily as technical factors. First is the field of view: human monocular vision spans approximately 200 degrees horizontally and 130 degrees vertically, with the central 20 degrees being the most sensitive region. Therefore, VR visual systems should ensure imaging quality within the central 20-degree field while expanding peripheral vision as much as possible. Second are frame rate and resolution: VR visual systems generate real-time imagery through sequential frame images that the brain integrates into continuous motion via visual persistence. Due to limitations in camera and display performance, VR imaging exhibits slight frame rate lag and sensor-display delay, making overly complex imaging processes lead to VR scene distortion.

Spatial vision refers to the human ability to form spatial depth perception through vision, enabling judgment of distance between oneself and objects of interest. Spatial depth perception derives primarily from four basic types of depth cues. First, static depth cues, which can be obtained from static images, such as object positioning, clarity, and relative size. Second, motion depth cues, derived from dynamic images, such as perceiving an object approaching when it appears to grow larger. Third, physiological depth cues, generated by the brain monitoring the accommodation of eye muscles. Fourth, binocular disparity cues, where the horizontal separation between eyes causes any point in a scene to project onto different horizontal positions in each eye's image, enabling depth perception. The first two require reference points within the three-dimensional environment, obtaining cues based on differences between target and reference objects. The latter two utilize multiple mechanisms of the human visual system, generating effects more internally within the user.

Position/direction tracking represents VR's imitation of real-world visual rules and is the dynamic response of VR imaging to user behavior, creating behavioral system immersion. This primarily manifests when VR users turn their heads or eyes, with the head-mounted display adjusting according to rotation angles to simulate perspective changes in real environments. This not only creates a sense of presence by closely linking the user with the virtual environment through visual activity but also enriches the user's visual spatial perception by simulating perspective changes in the relationships between objects in the VR environment. This requires both precise and timely capture of eye and head movements by software technology without hardware hindrance to natural eye and head motion, and VR display systems with sufficient computational imaging capability to minimize delay in simulated perspective changes.

In summary, several guiding principles for VR visual experience design can be derived: For simulated objects, digital modeling should ensure clearly distinguishable static features and dynamic behaviors following reasonable motion principles, while considering hardware limitations and utilizing mature, stable implementation methods within technical constraints. The human eye's sensitive region spans approximately 20 degrees of central vision; to reduce computational burden on head-mounted displays, peripheral regions can be appropriately simplified. For example, in Whaley VR film productions, effective information is often concentrated within a defined virtual stage, with three-dimensional lighting effects extending the periphery to control technical complexity while compensating for display field-of-view limitations.

Regarding relationships between objects and between objects and environments, spatial vision is crucial for transforming two-dimensional image information into spatial depth perception. When experience design requires virtual environments to convey spatial distance, emphasis or exaggeration of positional, size, clarity, and motion changes between individual objects or among multiple objects can be employed. For instance, the 2012 CCTV Spring Festival Gala performance of the song "Because of Love" incorporated falling petal effects that created an expansive stage space through exaggerated petal movements and focal shifts between performers.

Concerning the changing patterns of objects and environments, timely capture of eye and head movements and VR system feedback are paramount. Beyond developing gaze tracking technology, attention should be paid to display device comfort and stability. Current VR head-mounted devices fall into three categories: all-in-one VR headsets with built-in processors, smartphone-based VR boxes, and simplified cardboard versions. All-in-one headsets are often bulky, causing eye and neck muscle fatigue during extended use, while VR boxes, though lightweight, suffer from unstable smartphone positioning. These issues affect VR visual systems' ability to simulate natural eye and head movements and their visual feedback.

2.2 Auditory Component

Hearing is the second most important human perceptual channel after vision, providing approximately 15% of our sensory information about the objective world. Auditory information serves two functions: first, as real-time scene-accompanying sound that creates superimposed visual and auditory effects for audio-visual fusion; second, as three-dimensional sound information that supplements details beyond the visual field, enabling monitoring and identification of information from any direction, not just within the visible area.

In VR systems, sound information as scene accompaniment directly stimulates the user's auditory senses, with experiential demands focusing on sound simulation and synchronization between sound and image. Sound simulation requires computer-generated audio to approximate reality as closely as possible, modeling changes in sound velocity across media of different densities, reflection, refraction, and the superposition and mixing of multiple simultaneous sound sources. The concept of audio-visual synchronization originates from film art, referring to sounds produced by people, objects, or environments in the image that coordinate with the visual content. Maintaining synchronized audio-visual fusion significantly enhances system realism.

The transformation of sound information into directional information is achieved through human "monaural effects" and "binaural effects." The monaural effect, or "pinna effect," refers to how sound waves reflected by the asymmetric pinna reach the eardrum with spectral characteristics related to the sound source's direction. This means humans can preliminarily identify sound direction and localize it using a single ear. The binaural effect (Duplex Theory) refers to how the distance between ears on opposite sides of the head creates different path lengths for sound from a source, resulting in different time delays and attenuations, thereby generating auditory spatial perception. Therefore, when simulating sound environments, VR auditory systems can: first, simulate virtual surround effects by altering time differences between target sound sources and environmental sounds reaching the ears; second, adjust the timing and intensity of sound signals received by left and right ears according to eye and head rotation angles to simulate sound source distance and direction relative to the user's head center.

Thus, to achieve VR auditory system effects as both scene accompaniment and spatial awareness enhancement, sound data must be prepared from three perspectives: simulated sound sources for target objects; environmental sounds for current virtual scenes, including multiple sound superposition and reverberation and propagation characteristics across different media; and dynamic response to user head movement. The first two can leverage currently mature surround sound technology to enhance realism and presence, while the latter again emphasizes VR system's tracking and feedback of user head activities.

3. Summary and Outlook

Virtual reality audio-visual experiences transcend traditional human-computer interaction that relies on symbolic mediation, directly engaging innate perceptual activities. By simulating the stimulation and interaction of real environments on human senses, VR constructs a digital world derived from reality yet subject to human intervention. Through simulated replication of perception, VR audio-visual immersion places users' minds and bodies within digitally constructed virtual spaces—an attempt to reproduce the real world using media technology and an embodiment of media evolution along the trajectory of "virtualization."

As VR technology develops, audio-visual experiences will not only break the boundaries between reality and imagination on the basis of simulation but also deliver aesthetic experiences that transcend actual reality, liberating human understanding, emotion, and imagination from real-world constraints, and even transforming people's perspectives on self, world, and spacetime.

References [1] Zhou Kui. Media construction of virtual reality: A perspective from media technology history[J]. Modern Communication (Journal of Communication University of China), 2013, 35(08): 29-33. [2] Zhao Qinping. A survey of virtual reality[J]. Science China (Series F: Information Sciences), 2009, 39(01): 2-46. [3] Peng Danling. Modern theories of perception[J]. Philosophical Trends, 1989(07): 23-27. [4] David Gelernter. *Mirror Worlds: or the Day Software Puts the Universe in a Shoebox...How It Will Happen and What It Will Mean.* Oxford University Press, USA, 1993. [5] Jiang Qingquan. Visual mechanism and field-of-view design of head-mounted displays[J]. Television Technology, 2004(04): 81-84. [6] Xu Weizhong, Liu Hui, Tan Zheng. Development and research of 3D stereo display systems[J]. Journal of Image and Graphics, Vol.2, No.2, 3, Mar.1997, PP.144-148. [7] Luo Fuyuan, Wang Xingren, Peng Xiaoyuan. Sound rendering technology and its application in virtual environments[J]. Journal of System Simulation, 1999(05): 364-367. [8] Liu Guangming, Liu Xi, Dou Weibei. Research and simple implementation of virtual surround sound generation[J]. Audio Engineering, 2003(04): 43-47. [9] Stevens SS, Newman EB. The localization of actual sources of sound[J]. Am J Psychol, 1936, 48: 297-306. [10] Wang Shengjiu. Research on virtual auditory space and virtual surround sound technology and its DSP implementation[D]. Southeast University, 2006.

(Author Affiliation: Institute of Internet Information, Communication University of China)

Note: Figure translations are in progress. See original paper for figures.