Discussion on 3D Audio Technology Solutions and Standard Testing: Postprint

Authors: Yuhai Chen

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

Abstract

In the contemporary era, socioeconomic development continues unabated while science and technology advance at a rapid pace. Three-dimensional audio (3D Audio) technology has entered the public consciousness and is progressively integrating into the daily lives of the populace. The widespread application and promotion of three-dimensional audio (3D Audio) technology imbues auditory imagery with three-dimensional spatiality and directionality, thereby enabling individuals to perceive authentic sound presence within virtual environments. Consequently, an investigation into the technical solutions and standardized testing methodologies for three-dimensional audio (3D Audio) technology holds significant practical relevance.

Full Text

Exploration of 3D Audio Technology Solutions and Standard Testing

Abstract: In today's era of continuous socioeconomic development and rapid scientific advancement, 3D Audio technology has entered the public consciousness and gradually penetrated daily life. By endowing sound imagery with three-dimensional spatial and directional perception, this technology enables listeners to experience authentic sound presence within virtual environments. Consequently, exploring 3D Audio technology solutions and standard testing holds significant practical importance.

Keywords: 3D Audio; Technical Analysis; Standards

From the industrial age to the present, discussions on audio technology have remained integral to societal development. Audio technology has evolved from monophonic systems based on microphones and loudspeakers, to stereophonic audio, and subsequently to 5.1 and 7.1 surround sound systems, becoming increasingly sophisticated until reaching today's 3D Audio technology. The application and promotion of 3D Audio have further advanced and innovated the broadcasting, television, and journalism industries.

1.1 Definition of 3D Audio

3D Audio refers to systems that create distinctive sound imagery with three-dimensional spatial and directional perception through loudspeaker arrays or headphones. This enables listeners in virtual scenes to maintain clear spatial awareness and accurately localize sound sources, providing an auditory experience consistent with human psychoacoustic perception in the real world. Generally, 3D Audio is divided into three spatial layers. The first layer is at ear level, corresponding to the height of surround loudspeakers. The second layer is positioned at a 30-degree elevation from ear level, typical for sounds such as bird chirping and wind. The third layer represents sounds heard directly overhead, known as the "Voice of God" layer, exemplified by airplane roars and pre-storm thunder.

2.1 Channel-Based Implementation Analysis

To comprehensively understand 3D Audio, we must examine its implementation technologies. 3D Audio realization relies on three forms: channel-based, bed+object-based, and scene-based approaches. The channel-based approach involves preset channel configurations that directly map each channel's audio source to its designated loudspeaker position. While this method ensures sound achieves the intended effect, it requires listeners to select positions accurately to maintain sound within the listening area. This approach suffers from limited sound propagation range and requires playback within specific loudspeaker layouts.

2.2 Bed+Object Implementation Analysis

The bed+object approach also plays a crucial role in 3D Audio production. The sound bed carries basic environmental sounds and serves as the fundamental sound field audio signal. Since channels are closely tied to specific loudspeaker positions, implementing the bed+object approach requires predefined metadata for sound scenes. Audio produced through this method can achieve excellent sound effects, but necessitates accurate listener positioning and has limited propagation range, requiring playback within specific loudspeaker layouts.

2.3 Scene-Based Implementation Analysis

To meet increasingly sophisticated public demands for audio technology, scene-based implementation has gained popularity in 3D Audio production, achieving significant breakthroughs in sound capture and audio selection. In this model, sound transmits as pressure waves. For a given sound scene at a specific time, each point requires several pressure functions to be represented, meaning each point's pressure value plays a crucial role. If we know the pressure value at every point in space, we can perfectly reconstruct the spatial sound. To capture these pressure values, methods such as higher-order coefficients based on spherical harmonic transforms are used to describe pressure values at all points in 3D space. Since each point's pressure relates to neighboring points, microphone arrays can comprehensively collect sound, and mathematical principles can calculate and effectively derive these coefficients. To fully leverage scene-based audio production advantages, we must accurately obtain HOA (Higher-Order Ambisonics) coefficients, improve coding quality for spatial coefficients, and overcome implementation difficulties. A 3D audio codec framework can be designed based on VBAP (Vector Base Amplitude Panning) principles, using three channels as a group to synthesize virtual sound sources and downmix signals, adding time difference parameters during encoding, and employing a virtual sound image redistribution method based on linear equation solving at the decoding end to reconstruct signals consistent with the original channel configuration.

2.4 3D Audio Reproduction Environment Analysis

The 3D Audio reproduction environment plays a vital role in 3D Audio production, particularly prominent in Japanese broadcasting. Japan's 22.2 multichannel audio reproduction system employs 22 full-range loudspeakers and 2 low-frequency loudspeakers, though fewer loudspeakers can also be used. The number of loudspeakers significantly affects the optimal listening area size—the fewer the loudspeakers, the smaller the sweet spot, showing a positive correlation. In recent years, this technology has been widely applied in China's audio production and in cinemas, continuously improving audio quality.

2.5 Binaural Rendering Technology

The rise of Virtual Reality (VR) has further advanced 3D Audio technology application. VR 3D Audio playback generally employs binaural methods. Currently, major VR 3D Audio technologies include Ambisonics based on physical sound field reconstruction and spherical harmonic decomposition, Binaural recording technology, and HRTF (Head Related Transfer Function)-based reconstruction. Binaural rendering technology enables 3D Audio reproduction through both loudspeakers and headphones. To fully realize this technology's performance advantages, we must decode pre-produced audio source files, allowing ordinary stereo headphones to experience 3D Audio. Binaural rendering technology has matured 3D Audio, satisfying public spiritual needs and aligning with development trends.

3.1 Establishment of the 3D Audio Task Force

To further promote and apply 3D Audio technology in China, the AVS Technology Application Joint Promotion Working Group was established in March

2012 by the Ministry of Industry and Information Technology's Electronic Information Department and the State Administration of Press, Publication, Radio, Film and Television's Science and Technology Department. This working group facilitates further development of China's broadcasting, film, and television new services, maturing domestic innovative technologies. On February 19, 2016, China established the 3D Audio Task Force, led by China Central Television. This task force has advanced 3D Audio program production, transmission, distribution, and reproduction technology in broadcasting and television.

3.2 Development of 3D Audio in China

In recent years, stereo and multichannel audio coding technologies have developed rapidly, finding widespread application across many fields. 3D Audio development in China requires time for verification. The 3D Audio Task Force completed requirements specification on April 15, 2016, followed by a call for proposals on April 16, 2016. The task force held its first working group meeting subsequently. Final proposals were submitted on January 19, 2017, and standard verification was completed in October 2017. Currently, the 3D Audio Task Force is conducting test summaries and developing the standard documentation.

3.3 Discussion on 3D Audio Broadcasting Standards in China

China's current 3D Audio broadcasting industry standards basically adopt the bed+object implementation approach, requiring testing at both the audio-video compression encoding end and receiving terminal. Therefore, we must clarify supported objects, specifically coordinate positioning in 3D production, volume levels, and object sound presence. We must also effectively implement metadata control, channel description, and program-related information description. Furthermore, post-production must support objects, requiring at least 64 objects in post-production environments and at least four objects in live environments to create favorable atmospheric conditions and satisfy listeners' higher-level audio enjoyment. 3D Audio's transmission method has transformed sound effects from static to dynamic transmission. 3D Audio can create excellent sound environments, meeting listeners' high-level demands and promoting broadcasting and television industry development, thus enabling technology to better serve human society.

References: [1] Wang Rui, Qin Yili, Zhang Jiandong, et al. 3D Audio Technology Solutions and Standard Testing [J]. Radio & Television Technology, 2017(12). [2] Liao Chuanqi, Tu Weiping. Research on 3D Audio Coding Technology Based on Spatial Position Information [J]. Computer Engineering, 2017, 43(1): 303-308. [3] Zhang Yang, Zhao Junzhe, Wang Jin, et al. Current Status and Development of Key 3D Audio Technologies in Virtual Reality [J]. Audio Engineering, 2017, 41(6): 56-62. [4] Huang Yichao. Development of Stereo and Multichannel Audio Coding Technology [J]. Research, 2016(17): 00011-00011.

Author Information: Chen Yuhai, Jiangxi Radio and Television Station