Current Status and Prospects of Television Audio Technology (Postprint)

Authors: Zhang Lei, Jiang Shijie

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

Abstract

With the continuous development of television video technology, television audio has evolved from mono and stereo to multi-channel and 3D surround sound formats. This paper examines the evolution trajectory of high-definition audio technology, introduces contemporary mainstream multi-channel and 3D surround sound formats, and elaborates on the control functionalities of metadata. Through a systematic review of existing technologies, it predicts that television audio technology will advance toward full-dimensional reproduction, and that with the continuous improvement of metadata functionalities, future television audio will demonstrate enhanced interactivity and flexibility.

Full Text

Current Status and Future Prospects of TV Audio Technology Development

Abstract: As television video technology continues to evolve, TV audio has progressed from mono and stereo to multi-channel and 3D surround sound formats. This paper examines the evolution of high-definition audio technology, introduces today's mainstream multi-channel and 3D surround sound formats, and elaborates on the control functions of metadata. By reviewing existing technologies, we predict that TV audio technology will develop toward full-dimensional reproduction, and with the improvement of metadata functionality, future TV audio will offer enhanced interactivity and flexibility.

Keywords: High-definition era; 3D surround sound; metadata; full dimension

Classification Code: G220.7

Document Code: A

Article ID: 1671-0134(2017)12-061-04

DOI: 10.19483/j.cnki.11-4653/n.2017.02.011

Since the invention of black-and-white television in the 20th century, television has become the mainstream medium for information collection, exchange, and dissemination. Television technology has undergone three qualitative leaps: in the 1950s, television transitioned from black-and-white to color, bringing audiences an unprecedented color visual experience; in the 1990s, with the development of digital technology and networking, television evolved from analog single-image transmission to an intelligent, interactive, and multi-purpose digital format; the third generation comprises digital high-definition television and 3D stereoscopic television, delivering audiences an immersive high-definition video experience with powerful and shocking audio effects. In the analog television era, TV systems could provide at most two-channel audio links. When television transmission shifted from analog to digital, TV audio also transitioned from mono to stereo and multi-channel formats. Particularly with the advent of the digital television high-definition era, multi-channel surround sound systems have become an inevitability.

1. Evolution Path of High-Definition Audio Technology

Human demands for TV audio reproduction quality have always grown with technological advancement. The most significant change is reflected in the increasing number of reproduction channels. Human hearing is a comprehensive response to multiple types of information. In addition to basic sound elements such as loudness, pitch, and timbre, it also includes factors that reflect spatial sound characteristics, such as sound source direction, distance, sound field size, and sound field color. The continuous growth in the number of TV audio system channels aims to reproduce as much of this information as possible.

In the analog television era, TV audio employed mono systems that used one or several loudspeakers to reproduce sound signals recorded by a single microphone. This reproduction system could only represent loudness, pitch, and timbre, and to some extent reflect the distance of sound sources, but could not represent important characteristics such as sound source spatial localization. Mono TV audio used analog signals transmitted via frequency modulation, which often produced several problems during transmission, storage, and conversion: (1) The dynamic range of audio signal storage media was only 40-50 dB, far below the maximum signal dynamic range of program sources (120 dB); (2) During signal editing and conversion (program editing, dubbing, and delay effect processing, etc.), sound quality deteriorated rapidly as the number of conversions increased.

To improve analog mono TV audio quality, digital audio development became an inevitable trend. In the digital television era, standard-definition digital TV featured mono or stereo digital audio signals, while high-definition digital TV featured multi-channel digital surround sound signals. The digitization of audio signals easily achieved a dynamic range greater than 90 dB. Additionally, digital

audio signals could undergo non-linear editing without increasing audio signal distortion.

For the stereo two-channel system of standard-definition digital TV, the "binaural effect" could be utilized to create "auditory illusions," achieving horizontal and vertical localization of front sound sources and producing relatively obvious representation of sound spatial characteristics. Although this reproduction system represented significant improvement in sound quality and listening effects compared to the analog mono era, there remained a large gap from actual sound fields. The main problem was that the lateral and rear region sound fields for listeners were not faithfully reproduced.

Entering the high-definition era, as TV picture resolution became increasingly higher, multi-channel surround sound systems for TV audio became essential to achieve better audio-visual enjoyment. Multi-channel surround sound systems pursue comprehensive representation of sound spatial characteristics, giving people a three-dimensional spatial impression. Currently, digital TV audio standards adopt 5.1 surround sound systems, specifically the 3/2/.1 configuration scheme. This method determines loudspeaker placement for reproduction according to ITU-R BS.775 recommendations [1], as shown in [Figure 1: see original paper].

To obtain more stable sound image localization and a larger listening area, multi-channel surround sound systems have expanded from 5.1 channels to 7.1, 10.2, and even 22.2 channel systems, greatly enriching sound reproduction capabilities. While 10.2 and 22.2 channel surround sound systems remain in the experimental stage, 7.1 channel surround sound has become the audio format for the vast majority of high-definition films. The typical loudspeaker layout is shown in [Figure 2: see original paper]. The 7.1 channel surround sound system adds a set of surround loudspeakers (Lb and Rb) while moving the 5.1 channel system's surround loudspeakers (Ls and Rs) forward, dividing the surround sound field into three parts: front side (30°-90°), rear side (90°-150°), and rear (150°-180°), further enhancing the continuity and envelopment of surround sound field imaging.

Currently, most television stations adopt Dolby Digital and Dolby E technology for digital high-definition TV program recording and transmission, as shown in [Figure 3: see original paper]. In television production, Dolby E technology [4] is used, which can transmit up to 8 channels of digital audio signals through a single AES digital audio pair. Since most television broadcast infrastructure (video recorders) only has 2-channel or 4-channel audio processing capabilities, Dolby E technology allows multi-channel signals to be stored on existing equipment without requiring equipment updates for multi-channel audio signal transmission. Additionally, Dolby E encoded audio frames have a duration of 40 ms, identical to video frames, facilitating synchronized audio-visual editing. Before television broadcasting, Dolby Digital encoding compresses and encodes multiple formats (5.1 channel, mono, stereo) of digital audio signals into a pair of Dolby Digital bitstreams for MPEG encoding/multiplexing. At the user re-

ceiving end, set-top boxes decode the Dolby Digital bitstream to restore the pre-encoded audio signals for input to users' home theater systems.

The DTS format, developed by DTS, Inc., is primarily applied in professional cinema, home theater, and pure music fields. This format employs coherent acoustics coding technology with compression ratios between 2.9:1 and 4.3:1, sampling rates between 8-192 kHz, and quantization precision between 16-24 bits. First used in *Jurassic Park* on June 11, 1993, this system employed separate audio and picture methods, with DTS bitstreams recorded on CD-ROM. Therefore, simply playing the CD containing digital audio signals separately yields multi-channel surround sound effects. If multi-channel music is recorded on a CD in DTS format and played with an ordinary CD player connected to a DTS decoder at its digital output, surround sound music can be obtained. This type of disc is called DTS-CD.

2. Multi-Channel Surround Sound Formats in the High-Definition Era

Multi-channel surround sound formats gradually became popular in the early 1990s, and to date more than a dozen multi-channel surround sound formats exist. Below, typical surround sound formats are organized and elaborated according to increasing channel numbers.

2.1 5.1 Channel Surround Sound Format

Currently, the 5.1 channel surround sound format is the most popular system. Common surround sound formats include Dolby Digital and DTS. The Dolby Digital format, developed by Dolby Laboratories [3], is primarily applied in professional cinema, broadcast television, and home theater. This format employs AC-3 coding technology with compression ratios up to 10:1 and supports maximum data transmission rates of 640 kbit/s.

2.2 7.1 Channel Surround Sound Format

In the film domain, with the popularization of BD discs, high-definition films have become mainstream. Major companies have strived to develop their own new surround sound systems to compete in the high-definition field. Dolby Laboratories first demonstrated the new Dolby Digital Plus system at the Tokyo AES conference in 2004, subsequently launching the Dolby True HD system [7]. Both systems are multi-channel audio formats designed for high-definition disc format development. DTS, Inc. also launched the DTS HD system [6]. Currently, these three audio formats have all been incorporated into BD disc audio standard formats. These new systems all start from 7.1 channels and are upwardly expandable. Furthermore, sound reproduction is not limited to the horizontal dimension; through loudspeaker configuration, vertical dimension localization can be reproduced, greatly enhancing sound expressiveness.

2.3 3D Surround Sound Formats with Height Channels

Currently, people have begun pursuing "holographic stereo" effects for sound reproduction. The DC28 Digital Cinema Technology Committee established by SMPTE has formulated channel capacity and loudspeaker configuration methods for future digital cinema (SMPTE 428.3M protocol), as shown in [Figure 4: see original paper]. This configuration describes the setup of 20 channels, with 16 channels arranged in the horizontal dimension and 4 channels serving as vertical arrays to enhance height 层次感.

Japan's NHK company has launched a 22.2 channel surround sound system [7] for Ultra-High-Definition TV (UHDTV), as shown in [Figure 5: see original paper]. The entire system has 10 loudspeakers located at the horizontal plane of the listener's head, 9 loudspeakers above the listener's head level, and 3 loudspeakers plus 2 subwoofers set below the listener's head level. Although this system remains in the promotion and experience stage, it demonstrates that digital TV audio systems will develop toward multi-channel full-dimensional directions.

In recent years, companies such as Dolby, DTS, and Barco have respectively launched Dobly Atmos, DTS:X, and Auro-3D 3D surround sound reproduction systems. Dolby Atmos and DTS:X have introduced four reproduction formats for home theater: 7.1.4, 7.1.2, 5.1.4, and 5.1.2. [Figure 6: see original paper] shows the DTS:X 7.1.4 reproduction system. Auro-3D home theater systems include 9.1 and 10.1 formats, with [Figure 7: see original paper] showing the Auro 3D 10.1 format reproduction system. Due to the addition of height channels, spatial sound source localization is enhanced and three-dimensional spatial perception is effectively expanded, bringing audiences a more realistic experience.

3. Control Functions of Metadata in Multi-Channel Formats

To ensure correct transmission and reception of digital TV audio signals, metadata must be used as a control mechanism throughout the complete chain from production to broadcasting and reception. Metadata is essentially data about audio data—audio description and control parameters such as downmix parameters and dialogue normalization parameters. Different multi-channel coding technologies employ metadata for audio signal control.

3.1 Downmix Metadata

Downmix metadata is designed to enable multi-channel programs to be viewed and heard by stereo and mono users. Since the LFE (low-frequency effects) channel records audio signals primarily used for atmosphere rendering, only the five main channels (L, R, C, Ls, Rs) are used when multi-channel programs are downconverted to stereo or mono.

[Figure 8: see original paper] and [Figure 9: see original paper] respectively show the algorithms provided by Dolby Digital and MPEG 2-AAC [8] for downmixing multi-channel programs to stereo.

3.2 Dialogue Normalization

When playing audio from different sources, inconsistent loudness often occurs during program switching. The played audio may be different program types on the same channel (news, advertisements) or different broadcast channels. To solve this problem, reference levels are typically encoded directly into the audio data stream during encoding. Normal spoken dialogue subjective sound pressure level is usually used as the reference level.

At the decoding end, the reference level value is used to set reproduction volume in the reproduction system. System volume control is generally set according to the loudness expected by listeners or according to sound pressure level (SPL). Applying reference level values mainly includes the following functions: achieving the subjective sound pressure level of dialogue expected by listeners, and indicating the reference level value of dialogue sound pressure level in audio signals. Thus, regardless of the program type being decoded, listeners can set dialogue volume to maintain consistent subjective loudness of dialogue.

As shown in [Figure 10: see original paper], several different audio signals have inconsistent average dialogue levels as numerically marked. When switching between different programs, inconsistent levels lead to different subjective sound pressure levels. If film dialogue level is reduced by 3 dB to -30 dB as the subjective sound level (the reference level value), applying the reference level value will automatically normalize dialogue levels to -30 dB when switching to different program types, achieving consistent subjective perception, as shown in [Figure 11: see original paper].

4. Conclusion

In the analog television era, the role of sound in TV systems was somewhat neglected, consistently referred to as TV "accompaniment audio." As television has evolved from analog standard-definition to digital high-definition, ultra-high-definition, and stereoscopic TV, audio has become extremely important in supplementing video during the pursuit of visual presence. The development of sound from mono and stereo to multi-channel surround sound formats has become an inevitable trend. We believe that with further development of digital TV audio technology, sound will achieve full-dimensional reproduction. Moreover, as metadata functionality becomes increasingly sophisticated, future TV audio technology will incorporate greater interactivity and flexibility.

[1] ITU-R Recommendation BS.775-1. Multichannel Stereophonic Sound System with and without Accompanying Picture. International Telecommunication Union, Geneva, Switzerland, 1992-1994.

[2] Wang Shusen. Audio Technology Application in Olympic Broadcast International Common Signal Production [J]. Electroacoustic Technology, 2008, 32(5): 4-10.

[3] ATSC A/52B:2010. Digital Audio Compression Standard (AC-3, E-AC-3). Advanced Television Systems Committee, Washington, D.C., 22 November 2010.

[4] Kong Xiaolei. Introduction to Dolby E Technology. Electroacoustic Technology, 2003(5): 61-63.

[5] Roger Dressler. Dolby Audio Coding for Future Entertainment Formats. Dolby Laboratories, Inc., San Francisco, CA, White Paper, 2006.

[6] DTS-HD Audio Consumer White Paper for Blu-ray Disc and HD DVD Applications. DTS, Inc., White Paper, 2006.

[7] Kimio Hamasaki, Toshiyuki Nishiguchi, Reiko Okumura, Yasushige Nakayama, Akio Ando1. A 22.2 Multichannel Sound System for Ultrahigh-Definition TV (UHDTV). SMPTE Motion Imaging Journal, 2008, 117: 40–49.

[8] ISO/IEC 13818-7. Information technology - Generic coding of moving pictures and associated audio information-Part 7: Advanced Audio Coding (AAC).

(Author Affiliation: Beijing Television Station)

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