

Introduction.

Since the dawn of television, sound has been an integral part of programming, from the analog origins using FM monophonic audio to the advances in stereo FM to the digital video era that enabled programs with surround audio (5.1, 7.1, etc). The next advancement in audio technology adds additional channels (height) to surround audio and provides improved flexibility in how programs are produced, distributed, and rendered at the consumer device. This audio technology, known as Next Generation Audio (NGA), includes new tools such as audio objects, dialog enhancement, personalization, and advanced audio metadata.

What is Immersive Sound?

We hear sound in all dimensions: in front, behind, to the left or right, and above and below. Mono sound comes from only one direction, presumably in front. Stereo sound can produce mono-center and separate left-right sound fields. As the name suggests, surround sound creates a sound field enveloping the listener front and back, left and right. Immersive sound is the complete compilation of sound in all directions to provide a hemisphere sound field for the listener. Immersive sound can be approximated by adding speakers above a surround set of speakers or using up-firing soundbars; however, rather than the immersive sound being defined solely by speaker locations, it can also use audio objects with spatial metadata, which can be static as well as dynamic, to specify their location in the sound field. The concept of audio objects is integral to Next Generation Audio encoding techniques such as Dolby AC-4, MPEG-H, and DTS UHD.

Examples of immersive sound can be sampled here (requires iPhone or iPad):

<https://apps.apple.com/us/app/media-hyperium-mh-ii/id1600588295>

What are Audio Objects?

Audio objects can be considered speaker channels unrelated to a fixed speaker position. When the NGA program is rendered at the consumers' device, a specific audio object can be placed anywhere in the acoustic space, directed by audio metadata. This spatial metadata can be dynamic, allowing an audio object to be coupled with an on-screen object, such as a flying bird or moving vehicle, with the sound following that object. The audio object can also be personalized by the viewer with a specific language, announcer, or accessibility feature such as an audio description. The rendering of the audio at the consumer device can take advantage of the audio metadata to optimize the playback to the capabilities of the device, be it an advanced AVR (Audio Video Receiver) or a smartphone/tablet with a binaural audio processor for an immersive experience with headphones.

Immersive Sound Production

With the implementation of a new audio system, from mono to stereo, stereo to 5.1, and 5.1 to immersive, there are opportunities and challenges to use the latest tools that enable a more compelling program and provide additional information to the viewer. The Guidelines contain many documented case studies of how to produce an immersive audio program, from drama to sport to special events.

The current TV live production infrastructure (mainly based on SDI workflows) is capable of producing and distributing up to 16 channels of next-generation audio (e.g., Immersive and personalization). This typically consists of a single pre-rendered mix of 5.1 or 5.1.4 (5.1 + 4 heights) with additional stereo or mono objects representing additional dialogs for multiple languages, audio description, or effects. Typically, the configuration changes dynamically during a live event transmission; for example, stereo advertisement clips may be introduced anytime.

Audio Metadata

Audio metadata is integral within the immersive audio ecosystem. While each of the NGA codecs has its own metadata profile, there is a metadata schema standardized by the ITU called the Audio Definition Model (ADM). ADM can be carried along with the audio data in production using a serialized version of ADM (S-ADM) that can be encapsulated in SDI and carried as a data packet in ST 2110 IP transport. Dolby has developed mezzanine audio compression for AC-4 with two variants, AC-4i and ED2, that encapsulate the audio metadata with the compressed

production audio. Fraunhofer has developed a method of carrying audio metadata as an audio signal called Control Track[®], which occupies one digital audio channel embedded in the SDI ancillary data stream or ST 2110 IP audio transport along with the uncompressed audio program channels.

What are the Benefits of Immersive Audio using an NGA CODEC¹?

The NGA audio codecs, AC-4, MPEG-H, and DTS-UHD, can all efficiently encode an immersive audio program. A rich audio metadata stream can provide multiple presentation descriptions, dynamic audio objects, loudness control metadata, and compressed audio. This metadata can enable viewers to select their preferred presentation, dialog track, and language and apply a suitable loudness profile (night mode, wide dynamic range, earbuds mode, etc). The broadcaster can tailor the presentations to optimize for the end device, whether a TV, soundbar, or mobile device (smartphone or tablet). Further, each codec has features that improve dialog intelligibility for people with hearing impairments or for listening in noisy environments.

Where has NGA and Immersive sound been adopted?

NGA has been widely adopted for use with Ultra HD. Examples of immersive audio programming are:

2018 Olympics Korea-KBS/MBS Korea
2018 Eurovision Song Contest- RTP Portugal
2018 Pyeongchang Olympics-NBCU
2018 World Cup-NBCU
2018 European Athletics Championship-BBC, France Télévisions, RAI, and ZDF
2019 Rock in Rio-TV Globo Brazil
2019 Eurovision Song Contest - EBU
2018/2019 French Tennis Open- France Télévisions
2020 European Football-Telewizja Polska
2022 São João do Nordeste-Globo Recife
2022 World Cup-TV Globo Brazil
2023 Rock in Rio-TV Globo Brazil
2023 Carnival Rio de Janeiro and São Paulo-TV Globo Brazil
2023 Eurovision Song Contest-BBC

Guidelines Information

The Guidelines Rainbow books have a wealth of background information on immersive audio, NGA audio codecs, and case studies on using immersive audio in UHD programs. Relevant sections are listed below:

Red Book-Introduction to Ultra HD, Section 10.2 on NGA Codecs

Yellow Book- Beyond Foundational Technologies, Section 7.2 Next Generation Audio

Green Book- Ultra HD Distribution, Section 8.2 Audio and 8.5.2 Final Distribution Processing-Audio

Blue Book- UltraHD Production and Post Production, Section 11.3 Channel Based Immersive Audio, 12.2 Ultra HD-Next Generation Audio, Section 12.2.3 NGA Production Tools and Workflows

Indigo Book -Ultra HD Technology Implementations, Section 9 Monographs on NGA

NB: Copies of the Guideline Books can be found at: <https://ultrahdforum.org/guidelines/>

¹ CODEC is an abbreviation for enCOder/DECoder