Fraunhofer IIS

TECHNICAL PAPER

STUDIO RECOMMENDATIONS FOR 3D-AUDIO PRODUCTIONS WITH MPEG-H AUDIO

Yannik Grewe
Fraunhofer Institute for Integrated Circuits IIS
Erlangen, Germany
yannik.grewe@iis.fraunhofer.de

Ulli Scuda
Fraunhofer Institute for Integrated Circuits IIS
Erlangen, Germany
ulli.scuda@iis.fraunhofer.de

Adrian Murtaza
Fraunhofer Institute for Integrated Circuits IIS
Erlangen, Germany
adrian.murtaza@iis.fraunhofer.de

Markus Kahelin
Genelec Oy
Iisalmi, Finland
markus.kahelin@genelec.com

Nuno Duarte
Olympic Broadcasting Services
Madrid, Spain
NDuarte@obs.tv

MPEG-H AUDIO

Fraunhofer Institute for
Integrated Circuits IIS

Management of the institute
Prof. Dr.-Ing. Albert Heuberger
(executive)
Dr.-Ing. Bernhard Grill
Prof. Dr. Alexander Martin
Am Wolfsmantel 33
91058 Erlangen
www.iis.fraunhofer.de

Contact
Mandy Garcia
Phone +49 9131 776-6178
mandy.garcia@iis.fraunhofer.de

Contact USA
Fraunhofer USA, Inc.
Digital Media Technologies*
Phone +1 408 573 9900
codecs@dmt.fraunhofer.org

Contact China
Toni Fiedler
Phone +86 138 1165 4675
china@iis.fraunhofer.de

Contact Japan
Fahim Nawabi
Phone: +81 90-4077-7609
fahim.nawabi@iis.fraunhofer.de

Contact Korea
Youngju Ju
Phone: +82 2 948 1291
youngju.ju@iis-extern.fraunhofer.de

* Fraunhofer USA Digital Media Technologies, a division of Fraunhofer USA, Inc., promotes and supports the products of Fraunhofer IIS in the U. S.

CONTENTS

1 INTRODUCTION

1.1 About this document

3D-Audio or immersive audio mixing for home delivery is done using loudspeakers in an audio control room or near-field mixing environment. Such a room offers the most accurate spatial reproduction of the sound image.

The following document describes the main structural requirements and technical specifications for a 3D-Audio production environment. It details the best practice for mixing and reproduction in a flexible manner for loudspeaker reproduction systems ranging from 1.0 up to 7.1+4H channel layouts. It offers consultation for room geometry and room acoustics, loudspeaker positioning and electroacoustic performance, 3D-Audio monitoring and mixing capabilities and provides recommendations for related literature. It is not meant to substitute any aspect of professional studio planning, including, but not limited to interior design, room acoustical conceptions, studio signal flow, construction works or general studio operation. However, the acoustical and technical recommendations provided in this document should produce a satisfactory result for most professional environments. Further, the document gives a high-level preview of the MPEG-H 3D-Audio standard and highlights related requirements for a studio production environment.

This document addresses professional broadcast producers and engineers who have experience with audio control rooms and their operation. It is not intended to be a guideline to build up a studio complex from scratch. Facilities are encouraged to engage with Fraunhofer IIS or consult a professional studio designer to ensure that the best possible loudspeaker placement and reproduction environment is achieved.

1.2 About 3D-Audio and object-based audio

3D-Audio consists of audio channels and/or audio objects that can be defined by the content producer to design an audio scene with sounds above and around the listener. An audio scene comprises audio content and additional information which is affixed in the shape of metadata. This metadata is interpreted by a renderer, which allows the audio channels and audio objects to be flexibly reproduced over the target reproduction system. Using audio objects and combining them with channels enable the listener to interact with the content by using the standard TV remote control. Simple adjustments, such as increasing or decreasing the prominence of dialogue in relation to other audio elements, to more advanced scenarios are possible. Listeners may choose from different languages or commentators or even change the position of audio objects. In MPEG-H Audio (for details, see Chapter 6.1), the creation of related metadata and respective parameters is always under the full control of the content creator.

1.3 Contact

You need further assistance or want to talk to Fraunhofer IIS regarding your 3D-Audio production facility? Please contact Fraunhofer IIS:

amm-info@iis.fraunhofer.de
Fraunhofer Institute for Integrated Circuits IIS
Audio & Media Technologies Division
Am Wolfsmantel 33 | 91058 Erlangen | Germany

2 TECHNICAL BACKGROUND AND SPECIFICATIONS

In the following, technical background, terminology and fundamentals are detailed for the understanding of later topics.

2.1 Loudspeaker nomenclature

The nomenclature *m.n* + *h*H or *m.n.h* has been introduced in the literature where *m* is the number of loudspeakers in the horizontal plane, *n* is the number of LFE channels and *h* is the number of overhead or »height« loudspeakers [1] [2]. In MPEG standards, ambiguity in loudspeaker configurations is avoided by the use of an index table in the MPEG Coding Independent Code Points (CICP) standard [3]. In Table 1, some examples of this nomenclature with the related number of loudspeakers and corresponding MPEG CICP index are presented. Figure 1 shows an example of a 5.1+4H (or CICP16) loudspeaker reproduction setup.

Description	Alternative naming	Number of loud-speakers	MPEG-H 3D-Audio CICP index
1.0	Mono	1	1
2.0	Stereo	2	2
3.0	Left-Center-Right	3	3
4.0	Quad	4	10
5.1	3/2 Surround / ITU-R Rec. BS.775 [4]	6	6
5.1+2H	5.1+2H	8	14
5.1+4H	Auro 9.1, Dolby Atmos Home	10	16
7.1+4H	Auro 11.1, Dolby Atmos Home	12	19

***Table 1**: Examples of common loudspeaker reproduction setups and alternative naming.*

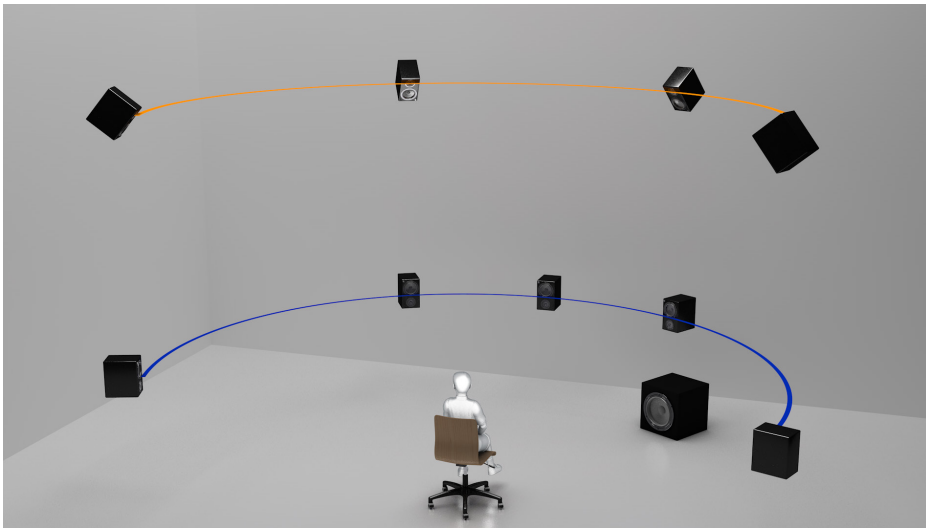

***Figure 1**: 5.1+4H loudspeaker reproduction layout according to CICP 16 in ISO/IEC 23091-3 [3]. The height of the middle layer is 1.2 m.*

2.2 Room design terminology

There are two common types of loudspeaker layouts in the context of a professional studio environment: **equidistant** and **orthogonal**. Both layouts can represent an appropriate mixing environment and are in-line with common studio recommendations.

The choice between them is solely based on the available room geometry, optimal mixing position, available space and additional equipment located in the room.

In an **equidistant layout**, the distance from the listening point to each loudspeaker is approximately equal, resulting in a spherical loudspeaker placement. The advantage of this arrangement is that there is, in theory, no level and time offset between loudspeakers. However, given the fact that each room has a unique geometry, an equidistant layout may not be achievable.

In contrast, in an **orthogonal layout** loudspeakers are placed in a cubic fashion. This is usually done when the room length is significantly greater than its width.

2.3 Acoustic terminology

To ensure an accurate mixing environment, the room should meet several requirements. A good guidance can be found in standards such as ITU-R Rec. BS.1116 [5], as well as ITU-R Rec. BS.2051 [6]:

- The maximum **noise floor level** (e.g. NR25, NR15, NR10) according to Rec. ITU-R BS.1116
- The demanded **reverberation time** in the middle frequency range can be calculated with the Sabine formula as shown in [7] and provides useful indications for a rough planning.
- Strong **early reflections** between obstacles and boundary walls should be avoided and thus suitably treated with absorptive or diffusive material.

Further guidance on acoustics can be found, e.g. in [8] as well as in web-based room acoustic calculator tools, such as www.hunecke.de/en, www.sengpielaudio.com or www.mcsquared.com.

3 ROOM DESIGN

3.1 Size and geometry

To ensure an optimal reproduction environment, it is recommended to meet minimum and maximum criteria of the room size and its geometry as presented in Table 2. As each room can have a unique geometry, it is recommended to also take the dimensions of the intended loudspeaker reproduction system into account rather than the actual dimensions of the room.

Parameter	Value
Minimum room height	2.75 m
Minimum room width	4.0 m
Minimum room length	5.0 m
Minimum room volume	55 m^3
Max. distance loudspeaker to sweetspot	4.0 m

Table 2: Recommended minimum size and geometry of a listening room.

If the recommended criteria of the room size and its geometry can not be fulfilled, e.g. in an OB-van or production container, the minimum dimensions should not be less than the ones presented in Table 3.

Parameter	Value
Minimum room height	2.4 m
Minimum room width	2.35 m
Minimum room length	3.2 m
Minimum room volume	18 m^3

Table 3: Minimum size and geometry of a listening room

3.2 Reverberation and insulation

The background noise level in the control room should not exceed NRC 25 [5] [9] [10]. NRC 15 is to be aspired. To be assumed:

a) highest exterior noise,

b) all technical devices in the room at full operation,

c) air conditioning at 80%.

The necessity of a room-in-room solution is to be assessed. In such a case, an acoustic protective double door construction is also required. All cables and ducts, independent of their position, must be accessible at all times. The construction of wall ducts must fulfill highest possible acoustical insulation. Typical sound pressure levels in the control room while listening might easily reach 110 dB_{SPL}. Particular attention has to be paid to proper insulation at frequencies below 100 Hz (structure-borne noise).

In terms of reverberation decay time, RT_{60} is taken according to Rec. ITU-R BS.1116 [5] at 63 Hz, 125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz and 8 kHz and the results should fall within the tolerance limits relative to the average value T_m (see Figure 2). Strong discrete reflections should be appropriately treated using diffusion or absorption to reduce coloration of the reproduced audio signals.

The room shall not have audible sympathetic resonances (buzzing or rattling). This can be verified by using a two-minute sine sweep at listening level with equal sweep time per octave.

3.3 Air conditioning

The air conditioning needs to be designed in such a fashion that four people and the equipment may work in optimal conditions, taking different outside air conditions into account. Regarding the postulated NRC 25 limit (NRC 15 or NRC 10 is to be aspired to), none of the single parts of the A/C must exceed 10 dB_{SPL} A-weighted at a distance of 0.5 m; which means, at a performance level of 80 %, NRC 10 must be fulfilled. According to this, the ventilation outlets need to be constructed as large as possible and should not exceed 0.5 m/s air velocity. The exchange capacities per hour must fulfill five times the cubic content of the room.

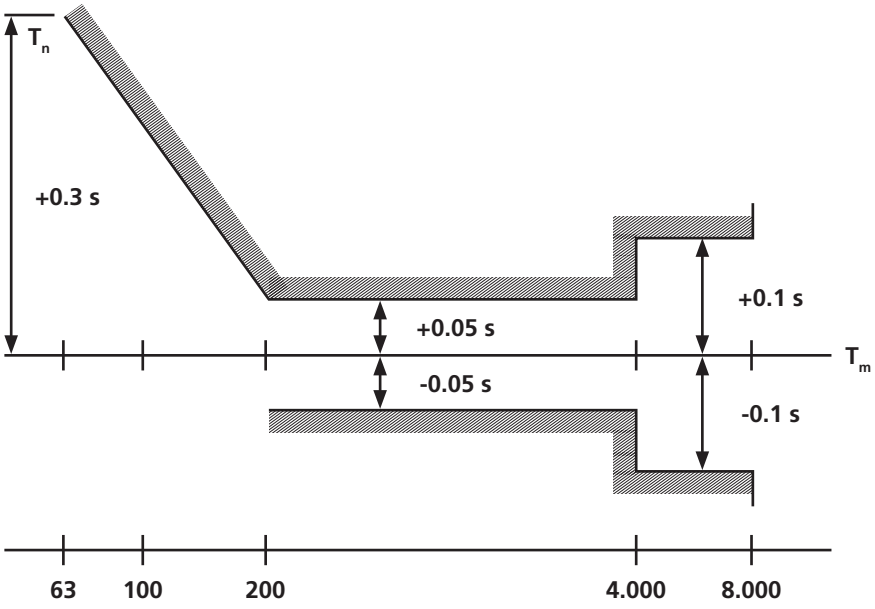

Figure 2: Reverberation time range of tolerance as described in Rec. ITU-R BS.1116 [5]

According to the values in Table 3 there is an average/maximum heat load of 4–5 kW/11kW which results in a thermal dissipation power of 6kW for the A/C. For the machine rooms similar values are assumed. Absolute values obviously depend on the installed equipment.

#	Description	Counts	Single load max (W)	Overall load max (W)	Single load avg (W)	Overall avg (W)
1	Persons	4	150	600	100	400
2	Loudspeakers	12	400	4800	100	1200
3	Subwoofer	2	1000	2000	200	400
4	Mixing desk	1	500	500	250	250
5	Equipment, fanless	1	1000	1000	500	500
6	Equipment, others	1	1000	1000	500	500
7	Lights	1	1000	1000	300	300
8	Other?					
			Overall	10900	**Overall**	3550

Table 3: Estimation of the maximum and average heat load in the studio compartment. Actual numbers depend on the installed equipment and may vary.

3.4 Lighting

If the mixing environment offers reproduction of accompanying picture, special attention should be paid to the lighting in the room.

Lighting at the seating position should come from overhead and be dimmable between 10 and 500 lux without a noticeable change in color temperature. To allow for the display of material at different frame rates, lights should not have any modulation below 1000 Hz. The use of DC-powered LED lamps with a filtered power supply is one means to accomplish this.

Room lighting should not directly fall on the display screen.

4 LOUDSPEAKERS AND REPRODUCTION LAYOUTS

For 3D-Audio productions, loudspeakers are the most important tool:

- to evaluate the immersive mix,
- to listen to individual microphone signals,
- to listen to individual channels, objects or interactivity presets enabled by the MPEG-H Audio Next Generation Audio system,
- to compare the immersive mix versus a version that has been rendered for 5.1 or stereo reproduction.

As in legacy production studios, the selection, placement and configuration of loudspeakers is always a compromise between available space, room geometry and budget. As other components such as equalizers or microphones can be replaced rather easily, installed loudspeakers usually stay in place for many years. Thus, loudspeaker planning should be done carefully and with an open mind for further developments to come.

4.1 Loudspeaker positions

For immersive audio productions, height layer of loudspeakers is required. Usually, this is done by adding four loudspeakers above the surround loudspeaker layout, which results in a 5.1+4H or 7.1+4H loudspeaker layout (see Figure 3 and 4 (Configuration D in Rec. ITU-R BS.2051 [6] (Table 1))).

In order to achieve a clearly immersive sound reproduction, the elevation angle between the middle layer of loudspeakers and the upper layer of loudspeakers should be within a range of 30° and 55°, preferably at 37° (see Figure 5). For a loudspeaker radius of 2.5 m, a minimum ceiling height of 3.5 m is needed, preferably 3.75 m. The recommended height of the middle layer is at 1.2 m which is the average listener's ear level when sitting on a chair. However, if needed the height of the middle layer speakers can be adopted according to the actual height of the listener's ear level.

The main listening position is the reference point or sweet-spot, ideally located in the geometric center of the loudspeaker reproduction layout. Following an equidistant reproduction layout, all loudspeakers should have the exact same distance to the sweet-spot. Where it is not possible to follow this loudspeaker layout, the distance difference between the loudspeakers has to be compensated for in terms of level and delay (see Chapter 4.3).

The recommended range and ideal horizontal angles of the center loudspeaker depend on the layout type and room geometry and might vary. In terms of vertical arrangement, it is strongly recommended to place the center loudspeaker at ear-height of 1.2 m. In image-related audio productions, it is advisable to place the center loudspeaker behind the center of an acoustically transparent canvas used to project the image. This implies that the middle loudspeaker layer is at the height of the center of the image projection. Or, put vice versa, that the center of the projected image shall be at listener's ear-height (1.2 m). Depending on the size of the image display, it might be necessary to slightly lower the center loudspeaker. Thus, the center loudspeaker can be angled within a range to maximum 20° from sweet-spot below the screen (see Figure 5).

All loudspeakers should be angled towards the sweet-spot, which means that the upper layer loudspeakers have to be tilted downwards (see Figure 5). Where this is not possible, the sweet-spot has to be located within the dispersion angle of the loudspeakers. The subwoofer should be placed off-center in the front and preferably on the floor. See Table 4 and 5 for recommended channel order, loudspeaker labels and their positions.

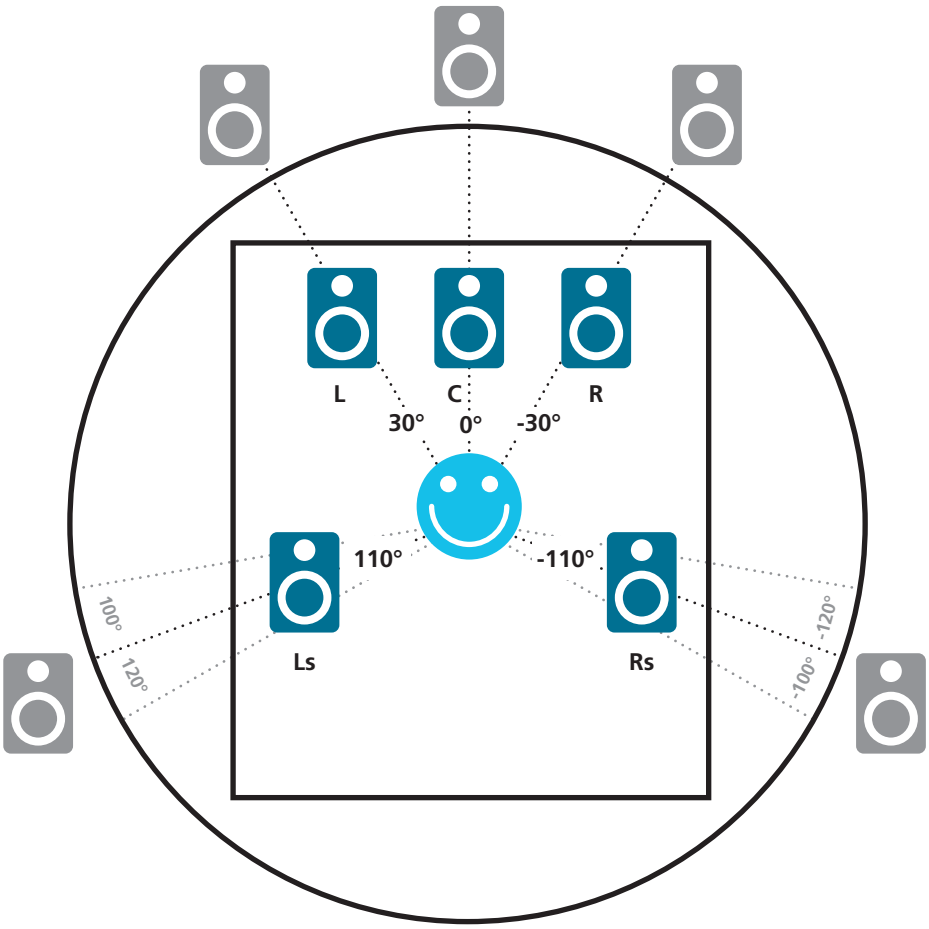

Figure 3: Top view of loudspeaker placement for the middle loudspeaker layer for a 5.1+4H reproduction layout according to Rec. ITU-R BS.775 [4]. Height of middle layer is 1.2 m.

#	Name	Alternative Names		Azimuth	Elevation
1	Left	L	M+30	+30°	0°
2	Right	R	M-30	-30°	0°
3	Center	C	M0	0°	0°
4	LFE				
5	Left Surround	Ls	M+110	+100° ... +120°	0°
6	Right Surround	Rs	M-110	-100° ... -120°	0°
7	Front Left Height	Lh	U+30	+30° ... +45°	+30° ... +55°
8	Front Right Height	Rh	U-30	-30° ... -45°	+30° ... +55°
9	Left Surround Height	Lsh	U+110	+100° ... +135°	+30° ... +55°
10	Right Surround Height	Rsh	U-110	-100° ... -135°	+30° ... +55°

Table 4: Recommended channel order and labels for a 5.1+4H loudspeaker configuration [11].

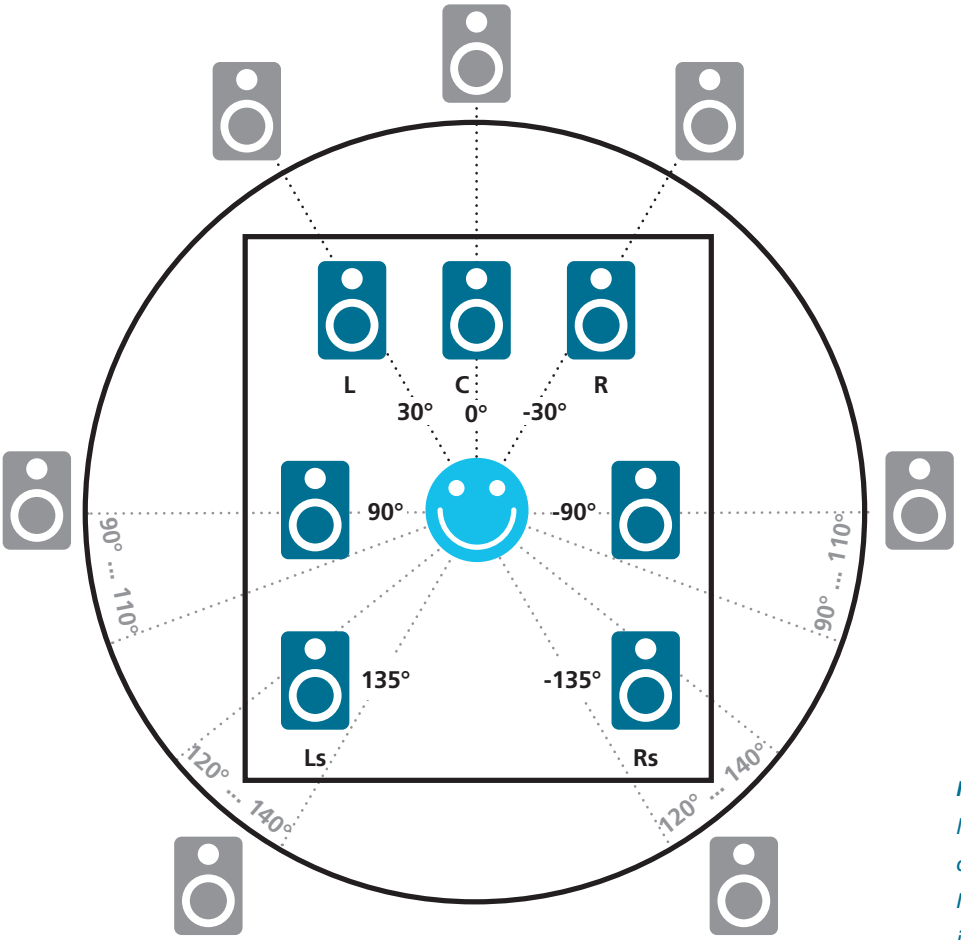

Figure 4: Top view of the middle layer loudspeaker placement for a 7.1+4H reproduction layout according to CICP 19 in ISO/IEC 23091-3 [3]. Height of middle layer is 1.2 m.

#	Name	Alternative Names		Azimuth	Elevation
1	Left	L	M+30	+30°	0°
2	Right	R	M-30	-30°	0°
3	Center	C	M0	0°	0°
4	LFE				
5	Left Surround	Ls	M+135	+120° ... +140°	0°
6	Right Surround	Rs	M-135	-120° ... -140°	0°
7	Left Side Surround	Lss	M+90	+90° ... +110°	0°
8	Right Side Surround	Rss	M-90	-90° ... -110°	0°
9	Front Left Height	Lh	U+30	+30° ... +45°	+30° ... +55°
10	Front Right Height	Rh	U-30	-30° ... -45°	+30° ... +55°
11	Left Surround Height	Lsh	U+135	+100° ... +135°	+30° ... +55°
12	Right Surround Height	Rsh	U-135	-100° ... -135°	+30° ... +55°

Table 5: Recommended channel order and labels for a 7.1+4H loudspeaker configuration [11].

As mentioned above, the standard surround loudspeaker layout should be positioned at seated ear-height of 1.2 m. However, it might be necessary to elevate the surround loudspeakers (Ls, Rs for 5.1+4H and Ls, Rs, Lss, Rss for 7.1+4H) due to room use or architectural limitations. To ensure adequate signal separation to the elevated loudspeaker layer, the surround loudspeakers should be elevated following the guidelines as described below.

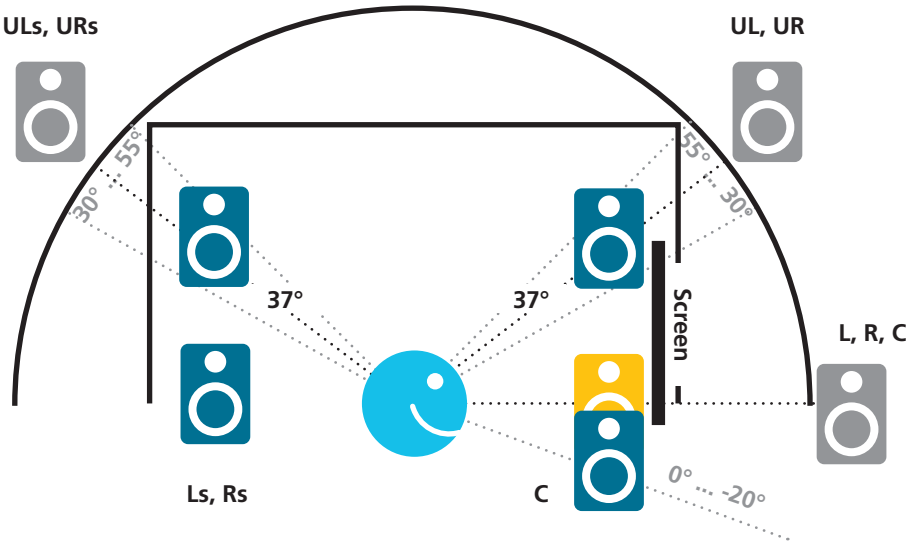

Figure 5: Side view of loudspeaker placement for a 5.1+4H reproduction layout according to ISO/IEC 23091-3 [3]. Height of middle layer is 1.2 m

The surround loudspeakers – namely Left Surround and Right Surround in 5.1+4H and 7.1+4H reproduction layouts – can be:

- slightly elevated but no greater than 15° from the sweet-spot.

Left Surround Side and Right Surround Side in 7.1+4H reproduction layouts can be:

- slightly elevated but no greater than 15° from the sweet-spot.

4.2 Loudspeaker specifications

The selected loudspeaker model should fulfill all requirements of regular studio monitors. The frequency response for the loudspeakers without subwoofer should extend from 50 Hz to 18 kHz with a tolerance of 3 dB. The subwoofer should be capable of reproducing frequencies between 20 Hz and 120 Hz.

The self noise of the loudspeakers should be as low as possible to be compliant at least with the NRC 25 specification [5]. Maximum acoustical power output should match the loudspeaker distance and room size with a headroom of 20 dB above reference level for non-distorted reproduction. The loudspeakers for middle layer and upper layer should be the same model to allow a smooth transition when panning a sound through the space. If space or budget does not allow the same model for the upper layer loudspeakers, a model from the same product line whose technical specifications are as close to the middle layer loudspeakers as possible should be chosen.

4.3 Loudspeaker calibration

All loudspeakers should be aligned in terms of level, delay and frequency response regarding the sweet-spot. The difference between the levels of any loudspeaker should not exceed ±0.5 dB. The difference between the delays of any loudspeaker should not exceed ±0.1 ms. The absolute value of delay adjustment should not exceed a threshold of 10 ms. The LFE signal reproduced over the subwoofer should have an additional 10 dB boost compared to the full-range loudspeakers. For alignment, it is recommended to measure the sound pressure level Z-weighted using pink octave-band noise at 1kHz for the Center and 63Hz for the LFE channel. Please note: It is only the LFE channel that should be boosted. Signals generated using bass management shall not be affected by this boost (see Figure 6).

Industry standard acoustic measurement tools should be used to make sure that all loudspeakers perform equally. Possible software solutions are: Easera, Dirac, Smaart, Fuzzmeasure, Room Eq Wizard, etc.

Loudspeaker calibration and alignment can be also performed using an integrated loudspeaker management system, that uses digital signal processing onboard each loudspeaker and a system management software to measure and optimize the system performance. Using loudspeaker management system enables easy setup recall and ensures best possible signal quality by eliminating the need for external signal processing hardware. Also level calibration and system time alignment can be performed at the same time. Possible solution is Genelec Smart Active Monitoring range of studio monitors with Genelec Loudspeaker Manager software [12].

Pink noise covering a frequency range from 500 Hz to 2 kHz at -23 LUFS on any of the middle or upper layer loudspeakers should result in 73 dB_{SPL} (RMS, measured C-weighted slow in the sweet spot with an omnidirectional measurement microphone) [13]. The playback chain should be set up in a way that it allows playback at a defined and reproducible reference level. It is used to set a reference gain (usually 0 dB) for the level adjustments in a listening session.

Equalizing the loudspeakers should be done with care and should result in a balanced frequency response between all loudspeakers in the sweet-spot. It is advisable to make the corrections in the low frequency range (below 300 Hz) only. The correction should be applied to all loudspeakers in the same way, see [14] and [15]. Target curves might be applied as relevant, such as the Brüel & Kjaer optimum curve or, for rooms exceeding 125 m³, the modified x-curve, as appropriate [16]. However, it is recommended to reproduce reference material to evaluate the timbre and consistency of the aligned loudspeakers and compare in other environments to ensure a proper calibration process. Make sure to avoid strong early reflections from the floor, ceiling or side walls. Apply acoustic material to absorb or diffuse reflections, if needed.

Please note: The loudspeakers should be filtered before level alignment.

4.4 Bass management

It is recommended to use a subwoofer for the reproduction of the low pass filtered LFE channel. Keep in mind to apply the +10 dB boost at the right point in the signal chain (see Figure 6) [4]. For most medium and small sized monitor loudspeakers, it is recommended to use bass management. For this, add up the low pass filtered signals from the middle and upper layer channels. The crossover frequency is usually set to a value between 80 Hz and 120 Hz depending on the size of your main loudspeakers.

The risk when working with improper bass management is that the mixer cannot evaluate the amount of low frequencies present on the middle or upper layer channels. With small upper layer loudspeakers unable to reproduce frequencies below 50 Hz, there can be significant low frequency energy (induced by wind or microphone handling, etc.) with strong effect on the aesthetics of the mix without the engineer taking note thereof. Good bass management helps to prevent such cases.

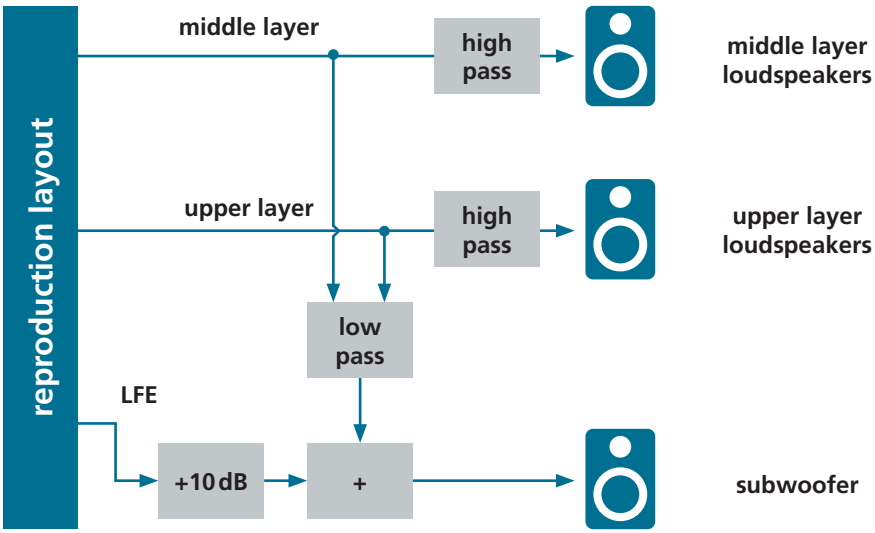

Figure 6: Bass management and subwoofer signal diagram.

5 MIXING AND MONITORING

3D-Audio production environments require specific software and hardware for mixing, monitoring and rendering. This includes, but is not limited to, the ones mentioned below.

5.1 3D-Audio mixing tools

- Industry standard computer hardware with macOS or Windows operating system
- Digital Audio Workstation (DAW) host application with 3D-Audio mixing capabilities such as Avid ProTools version 12 or higher, Steinberg Nuendo version 7 or higher, Merging Technologies Pyramix version 11 or higher or Cockos Reaper version 5
- Mixing console or DAW software plugin that can generate 3D-Audio panned signals, such as Solid State Logic System T broadcasting console, Fraunhofer MPEG-H Audio Authoring Plugin (MHAPi), New Audio Technologies Spatial Audio Designer plugin (SAD), DSpatial Reality Builder or Dear Reality DearVR Pro plugin
- Mixing console or DAW control surface with a minimum of 8 faders
- 3D-Audio monitoring controller to control the overall level of the reproduced mix (see Chapter 6.1)
- A separate monitoring path for 2.0, 5.1 or 7.1 reproduction (which either generates a downmix of the 3D-Audio mix, outputs an alternative rendering for a smaller loudspeaker configuration or bypasses the rendering and mastering workstation)

5.2 Monitoring controller

A feasible monitoring controller for 3D-Audio monitoring should satisfy the following recommendations:

- Overall level control with dim functionality of the loudspeaker signals.
- Solo and mute for individual loudspeaker channels and loudspeaker layers.
- Cue-Control.
- A separate monitoring path for 2.0, 5.1 or 7.1 reproduction either generated using a downmix of the 3D-Audio mix, using the alternative renderer output or bypassing of the rendering and mastering workstation.

Possible hardware solutions are: JBL Intonato24, AVID MTRX, QSC Qsys or Soundweb.

Studio monitor controller functionality can be also performed using an integrated loudspeaker manager system, that performs these tasks using system control software, control network to the loudspeakers and onboard signal processing inside each studio monitor. Software based system control enables easy recall of any desired setup, file backup of all critical settings and provides a graphical user interface for the user. Possible system solution is Genelec Loudspeaker Manager (GLM).

6 MPEG-H AUDIO

6.1 About MPEG-H Audio

MPEG-H Audio is a group of international standards developed by the ISO/IEC Moving Picture Experts Group (MPEG) which contains (among others) the HEVC video compression standard and the 3D-Audio compression standard. Developed by MPEG following an extremely competitive and collaborative process, the MPEG-H 3D-Audio standard specifies two profiles – »Low Complexity« and »Baseline« – which allow decoding and rendering of immersive content while enabling advanced personalization features within the computational limits of today's consumer devices and mobile applications. Audio objects may be used alone or in combination with channels for efficient delivery and reproduction of immersive sound. The use of these audio objects allows for interactivity or personalization of a program by adjusting the gain or position of the objects during rendering in the MPEG-H Audio decoder.

MPEG-H Audio has been carefully designed for enhancing the broadcast and streaming applications with exciting new features that enable broadcasters and content producers to create and distribute interactive and immersive audio. The system is already deployed and supported in devices across existing broadcast chains, from Authoring and Monitoring Units (AMAU) for realtime monitoring and content authoring, A/V contribution and emission encoders, to consumer devices such as TV sets and immersive soundbars.

Compared to legacy audio compression standards in TV broadcasting such as AAC, the MPEG-H Audio system is a complete integrated audio solution, which besides its advanced immersive and personalization features, includes rendering and downmixing functionality, advanced loudness and Dynamic Range Control management (DRC) and a unique system design for connectivity across multiple devices.

MPEG-H Audio metadata contains all necessary information for reproduction and rendering in arbitrary reproduction layouts and for ensuring the best audio experience on any platform.

***Figure 7**: Logo of the MPEG-H Audio trademark program.*

The MPEG-H Audio system is already adopted by ATSC, DVB, TTA (South Korean TV) and SBTVD (Brazilian TV) TV standards and it is the audio codec selected for ATSC 3.0 broadcasting in South Korea where it is used on-the-air since UHD broadcasts began on May 31, 2017.

To indicate to consumers that products implementing the MPEG-H Audio System will inter-operate correctly, Fraunhofer operates the MPEG-H Audio System trademark program. Devices eligible for the »MPEG-H Audio« trademark carrying the MPEG-H Audio trademark program logo (see Figure 7) have been tested for compliance with the MPEG-H Audio System trademark program.

In addition, MPEG-H Audio powers Sony's 360 Reality Audio immersive music initiative. This makes it possible for artists and music creators to produce an immersive musical experience by positioning sound sources such as vocals, chorus and instruments in space to perfectly match the creative and artistic intent. When playing back the resulting content, users can enjoy music that immerses them in sound from every direction.

6.2 MPEG-H Audio workflow

The production and transmission of MPEG-H Audio introduces new concepts compared to a legacy production. In the following, a summary of the MPEG-H Audio workflow is presented. For more detailed information, accompanying documents are provided showing insights in the MPEG-H Audio production and transmission chain.

Besides 3D/immersive audio, another key feature of MPEG-H Audio is the advanced interactivity option, enabled by audio objects transmitted separately from channel components. For example, during the broadcast of a football game, an MPEG-H Audio stream may carry a channel component (e.g., stereo, 5.1 or 5.1+4H) and additional audio objects for the »main commentator,« the »home team announcer« and the »away team announcer«. Thus, a rich set of MPEG-H Audio metadata is required to describe the related information of the broadcast signal. This metadata contains, for instance, information such as limits on the viewer's interactivity options, definition of pre-configured versions of the mix called »presets« that the user at home can choose from, or labels for each object to be displayed to the viewer. Casual viewers, particularly during the initial introduction of interactive audio options, will most likely benefit from simple interactivity that is limited to a »single-button-push« on their remote control.

The MPEG-H Audio system has been designed especially for such use cases, and content creators can define multiple presets and explore new creative options. A broadcaster can prepare mixes (including the default or main mix of the program) using authoring tools that specify an ensemble of gain and position settings for objects to create preset mix selections that can be presented on a simple menu to the user. Even more control of the audio elements in a program is possible and can be enabled in the »advanced MPEG-H Audio interactivity menu« by enthusiast viewers. All interactivity features offered to the user are strictly defined by the broadcaster during metadata creation. This process of generating metadata is called »**authoring**« and is the most important difference in production of MPEG-H Audio content compared to a legacy production.

In MPEG-H Audio, all metadata is modulated into a »Control Track.« The Control Track (CT) is a timecode-like audio signal and can be treated as a regular audio channel. Typically, the CT is carried on channel 16 within an SDI framework for live broadcast applications or on channel 16 of a multichannel wave-file. This tightly coupled transport of the CT together with the audio channels carrying the audio essence ensures integrity of the transmitted audio scene. Handling the CT as an audio channel ensures that it is always synchronized with its corresponding audio and video signal. It is robust enough to survive A/D and D/A conversions, level changes, sample rate conversions or frame-wise editing. The CT does not force audio equipment to be put into data mode or non-audio mode in order to pass through. In a future IP-based production facility, the audio and metadata would be transmitted in a container over IP according to e.g. SMPTE ST-2110.

In general, an MPEG-H Audio production can be carried out in a traditional fashion (see Figure 8), except for the fact that 3D-Audio signals and audio objects might be produced and processed. In this production scenario, a different handling of audio elements is required; accordingly, a 3D-Audio bus structure needs to be available in the production tools, such as Digital Audio Workstation (DAW), broadcast mixing desk and monitoring paths (see Chapter 5). Using additional audio objects, they must be kept separated from the other components such as »Music & Effects« (M&E) mixes – also called the channel bed – up to the authoring stage. During the authoring process, all audio signals are bundled together and metadata is created that describes the signals and gives further information about the rendering in the end consumer device, e.g. a TV set, a home theater, a soundbar or a mobile phone. Examples for describing metadata are:

- position information about where in space audio objects should be reproduced,
- interactivity limitations for audio objects,
- loudness information about each component and preset,
- text labels for presets and audio objects (including in multiple languages),
- reference and target loudspeaker layouts,
- and many more.

At this stage in the production chain, all audio including metadata is uncompressed PCM. In the next step, the audio and the associated metadata can be fed over existing SDI infrastructure to an MPEG-H Audio-enabled A/V broadcast encoder. In the encoder, the audio signals and metadata are efficiently compressed into an MPEG-H Audio bitstream and muxed together with the video bitstream into a transport stream that can be delivered to the consumer over different distribution paths, such as terrestrial broadcast or streaming. When delivered to the consumer, the MPEG-H Audio bitstream is decoded and the audio signals are rendered based on the metadata. The renderer will create the audio signals (i.e. loudspeaker feeds) for reproduction, depending on what kind of reproduction system is available. In addition, user interactivity can be fed into the renderer, e.g. for selecting a preset or adjusting the level of an audio object.

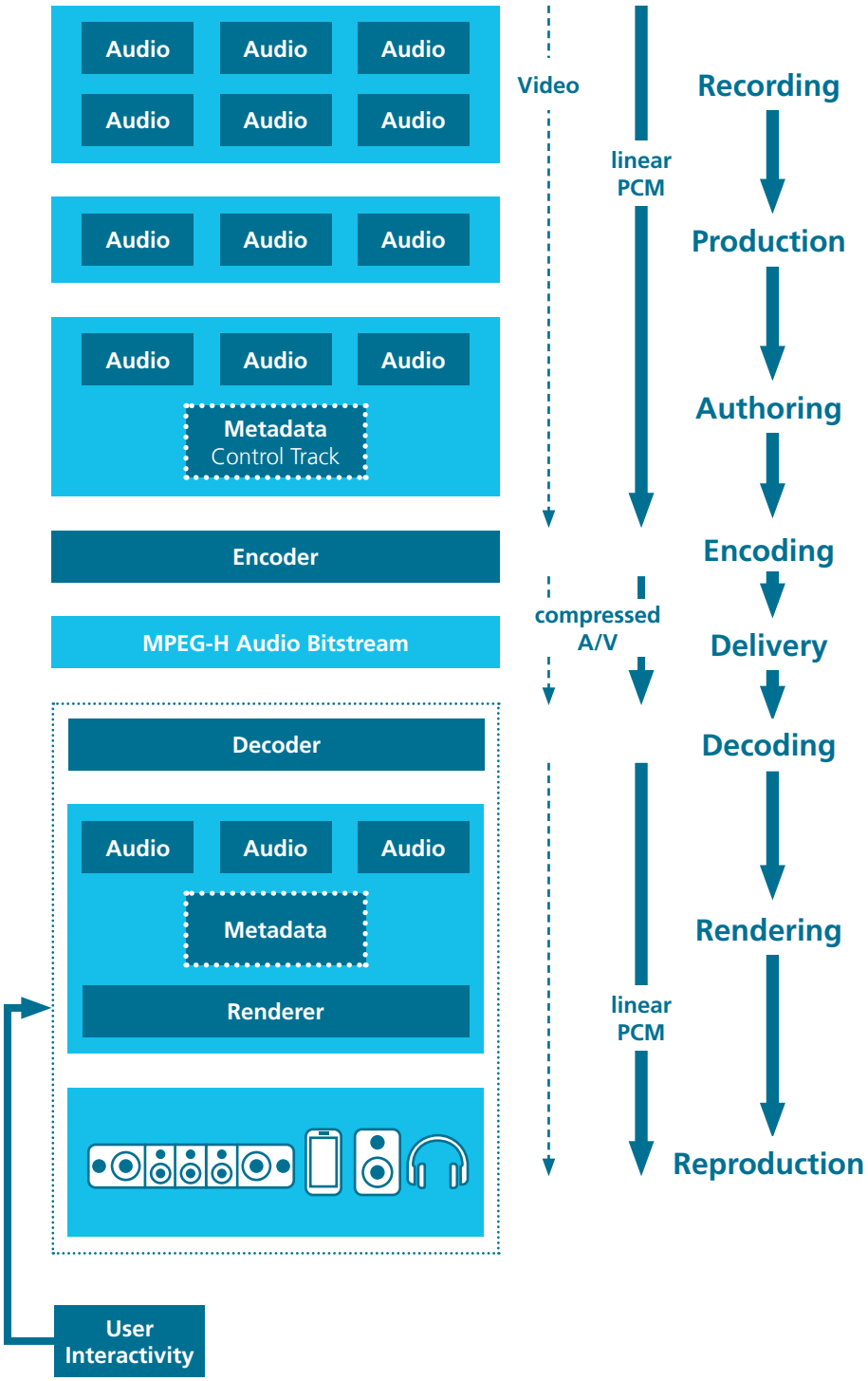

***Figure 8**: Schematic overview of an MPEG-H Audio production workow..*

7 FURTHER RESOURCES AND LITERATURE

7.1 Studio design and reference studios

- Silzle et al. (2009). Vision and Technique behind the New Studios and Listening Rooms of the Fraunhofer IIS Audio Laboratory. In Proceedings of the 126th Convention of the Audio Engineering Society. Munich. [17]

- Nixon, T., Bonney, A., Melchior, F. (2015). A Reference Listening Room for 3D-Audio Research. In International Conference on Spatial Audio. [18]

- EBU Tech. 3276 – 2nd edition (1998). Listening conditions for the assessment of sound programme material: monophonic and two-channel stereophonic. [15] and [16]

7.2 Links and literature about MPEG-H Audio

- www.iis.fraunhofer.de/tvaudio
- www.mpegh.com
- J. Herre et al. MPEGH Audio. The New Standard for Universal Spatial/3D Audio Coding [19]
- H. Stenzel, U. Scuda. Producing Interactive Immersive Sound for MPEG-H: A Field Test for Sports Broadcasting [20]
- U. Scuda. Comparison of Multichannel Surround Speaker Setups in 2D and 3D [2]
- U. Scuda, H. Stenzel, D. Baxter. Using Audio Objects and Spatial Audio in Sports Broadcasting [21]
- R. Bleidt et al. Development of the MPEG-H TV Audio System for ATSC 3.0 [22]
- Y. Grewe, C. Simon, U. Scuda. Producing Next Generation Audio Using the MPEG-H TV Audio System [23]
- C. Simon et al. Field Tests for Immersive and Interactive Broadcast Audio Production using MPEG-H 3D Audio [24]
- S. Meltzer, A. Murtaza. First Experiences with the MPEG-H TV Audio System in Broadcast [25]
- ISO/IEC 23008-3:2019, Information technology High efficiency coding and media delivery in heterogeneous environments Part 3: 3D audio [26]
- ISO/IEC 23091-3:2018, Information technology Coding-independent code points Part 3: Audio. [3]
- A. Murtaza, H. Fuchs, S. Meltzer. MPEG-H TV Audio System for Cable Applications [27]

8 REFERENCES

[1] ITU-R BS.21598: Multichannel sound technology in home and broadcasting applications. International Telecommunication Union, Geneva, 2019.

[2] Ulli Scuda: Comparison of Multichannel Surround Speaker Setups in 2D and 3D. In Proceedings of the International Conference on Spatial Audio, Erlangen, 2014. Verband Deutscher Tonmeister.

[3] ISO/IEC 23091-3:2018, Information technology Coding-independent code points. Part 3: Audio. ISO/IEC, Geneva, 2018.

[4] ITU-R BS.7753: Multichannel Stereophonic Sound System with and without Accompanying Picture. International Telecommunication Union, Geneva, 2012.

[5] ITU-R BS.1116: Methods for the subjective Assessment of small Impairments in Audio Systems including Multichannel Sound Systems. International Telecommunication Union, Geneva, 2014.

[6] ITU-R BS.20510: Advanced Sound System for Programme Production. International Telecommunication Union, Geneva, 2014.

[7] Floyd E. Toole: Sound Reproduction. Loudspeakers and Rooms. Focal Press, Amsterdam, 2008.

[8] Trevor Cox and Peter D'Antonio. Acoustic Absorbers and Diffusers: Theory, design and application, 3rd volume CRC Press, London, 2017.

[9] ISO 19961: Acoustics Description, measurement and assessment of environmental noise Part 1: Basic quantities and assessment procedures. International Organization for Standardization, Geneva, 2003.

[10] ISO 19962: Acoustics - Description, measurement and assessment of environmental noise. Part 2: Determination of environmental noise levels. International Organization for Standardization, Geneva, 2007.

[11] ITU-R BS.20512: Advanced sound system for programme production. International Telecommunication Union, Geneva, 2018.

[12] https://www.genelec.com/monitor-setup

[13] EBU Tech 3343. Guidelines for Production of Programmes in Accordance With EBU R 128: European Broadcasting Union, Geneva, 2016.

[14] EBU Tech. 3276. Listening conditions for the assessment of sound programme material: monophonic and two ñ channel stereophonic. European Broadcasting Union, Geneva, 1998.

[15] EBU Tech 3276E Supplement 1. Listening conditions for the assessment of sound programme material: Multichannel Sound. European Broadcasting Union, Geneva, 2004.

[16] SMPTE ST 222:1994 SMPTE Standard For Television Control and Review Rooms Monitor System Electroacoustic Response. SMPTE, 1994.

[17] Andreas Silzle, Stefan Geyersberger, Gerd Brohasga, Dieter Weninger, and Michael Leistner. Vision and Technique behind the New Studios and Listening Rooms of the Fraunhofer IIS Audio Laboratory. In Proceedings of the 126th Convention of the Audio Engineering Society, Munich, 2009.

[18] T. Nixon, A. Bonney, and F. Melchior. A Reference Listening Room for 3D Audio Research. In International Conference on Spatial Audio, 2015.

[19] Jürgen Herre, Johannes Hilpert, Achim Kuntz, and Jan Plogsties. MPEG-H Audio The New Standard for Universal Spatial/3D Audio Coding. In Proceedings of the 137th Convention of the AES, Los Angeles, 2014. Audio Engineering Society.

[20] Hanne Stenzel and Ulli Scuda. Producing Interactive Immersive Sound for MPEG-H: A Field Test for Sports Broadcasting. In Proceedings of the 137th Convention of the AES, Los Angeles, 2014. Audio Engineering Society.

[21] Ulli Scuda, Hanne Stenzel, and Dennis Baxter. Using Audio Objects and Spatial Audio in Sports Broadcasting. In Proceedings of the 57th International Conference of the AES, Hollywood, 2015. Audio Engineering Society.

[22] Robert L. Bleidt, Deep Sen, Andreas Niedermeier, Bernd Czelhan, Simone Fug, Sascha Disch, Jurgen Herre, Johannes Hilpert, Max Neuendorf, Harald Fuchs, Jochen Issing, Adrian Murtaza, Achim Kuntz, Michael Kratschmer, Fabian Kuch, Richard Fug, Benjamin Schubert, Sascha Dick, Guillaume Fuchs, Florian Schuh, Elena Burdiel, Nils Peters, and MooYoung Kim. Development of the MPEG-H TV Audio System for ATSC 3.0. IEEE Transactions on Broadcasting, 63(1):202 – 236, 2017.

[23] Yannik Grewe, Christian Simon, and Ulli Scuda. Producing Next Generation Audio Using the MPEG-H TV Audio System. In Broadcast Engineering and Information Technology Conference, Las Vegas, 2018. NAB.

[24] Christian Simon, Yannik Grewe, Nicolas Faecks, and Ulli Scuda. Field Tests for Immersive and Interactive Broadcast Audio Production using MPEG-H 3D Audio. SET International Journal of Broadcast Engineering, 2018.

[25] Stefan Meltzer and Adrian Murtaza. First Experiences with the MPEG-H TV Audio System in Broadcast. SET EXPO, 2018.

[26] ISO/IEC 23008-3:2019, Information technology - High efficiency coding and me diadelivery in heterogeneous environments - Part 3: 3D audio, including ISO/IEC 23008-3:2019/Amd 1: 2019, Audio metadata enhancements and ISO/IEC 23008-3:2019/Amd 2, 3D Audio Baseline profile, Corrections and Improvements. ISO/IEC, Geneva, 2019.

[27] Adrian Murtaza, Harald Fuchs, and Stefan Meltzer. Journal of Digital Video: MPEG-H TV Audio System for Cable Applications. Society of Cable Telecommunications Engineers and International Society of Broadband Experts, Exton, 2017.

9 NOTICE AND DISCLAIMER

ABOUT FRAUNHOFER IIS

For over 30 years, the institute's Audio and Media Technologies division has been shaping the globally deployed standards and technologies in the fields of audio and moving picture production. Starting with the creation of mp3 and continuing with the co-development of AAC and the Digital Cinema Initiative test plan, almost all consumer electronic devices, computers and mobile phones are equipped with systems and technologies from Erlangen today. Meanwhile, a new generation of best-in-class media technologies – such as MPEG-H Audio, xHE-AAC, EVS, LC3/LC3plus, Symphoria, Sonamic and up-Hear – is elevating the user experience to new heights. Always taking into account the demands of the market, Fraunhofer IIS develops technology that makes memorable moments.

Fraunhofer IIS, based in Erlangen, Germany, is one of 72 divisions of Fraunhofer-Gesellschaft, Europe's largest application-oriented research organization.

For more information, contact amm-info@iis.fraunhofer.de, or visit www.iis.fraunhofer.de/amm