APPLICATION BULLETIN

FRAUNHOFER SURROUND CODECS

SURROUND CODEC BACKGROUND, TECHNOLOGY, AND PERFORMANCE

This bulletin provides some background information on Fraunhofer's current surround codecs, part of the AAC codec family. It discusses the evolution of surround sound and typical applications of surround sound. It then presents the two types of surround audio coding used in AAC today, channel-based and spatial coding, and briefly introduces a new method, object-based coding. Loudness metadata, which is a common feature of surround sound systems, is also explained in terms of its AAC implementation.

A discussion of the relative performance of the AAC surround codecs and their strengths is presented, and then upcoming support of 7.1 channel surround is explained. There is some brief information on the wide support for the AAC family in application standards and consumer devices – five billion products at the moment – and then some useful details on surround file formats, audio editing software, and the LFE channel is presented.

The reader is encouraged to visit the Fraunhofer web site at www.iis.fraunhofer.de/amm for detailed information on each of the codecs in the AAC family, as well as our other technologies, and for possible revisions of this bulletin.

RELEASE V0.20 – INITIAL RELEASE WITH CORRECTIONS

Fraunhofer Institute for
Integrated Circuits IIS

Director
Prof. Dr.-Ing. Albert Heuberger
Am Wolfsmantel 33
91058 Erlangen
www.iis.fraunhofer.de

Contact
Matthias Rose
Phone +49 9131 776-6175
matthias.rose@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

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

1. EVOLUTION OF SURROUND SOUND SYSTEMS

The early decades of sound recording used a monaural signal, with sound appearing to emanate from a single point. Stereo improved the quality of reproduction by widening this point to a 40 to 60 degree image or window between the speakers.

This provides a more realistic simulation of a performance where the performers or instruments are arrayed in front of the listener, but is still not representative of real life, where acoustic reflections, audience reaction, or sound effects may arrive from any direction. Surround sound breaks this stereo window, allowing the consumer to move from listening to sound that appears to come from the next room, to being immersed in the performance around them. In addition to the feeling of envelopment that surround provides, there is also the enlargement of the "sweet spot" optimum listening area compared to stereo.

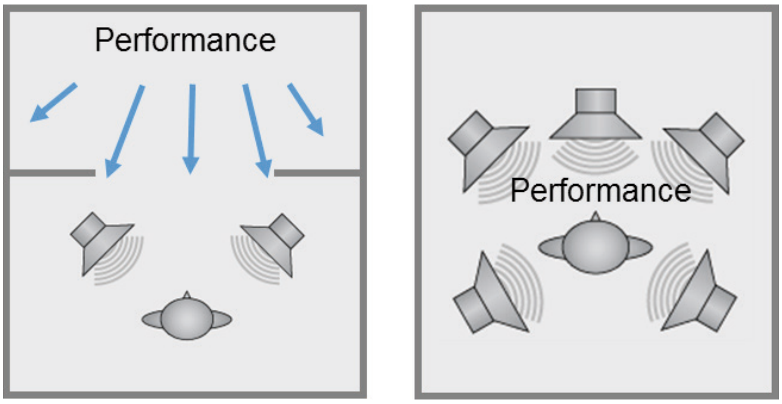

Figure 1: Surround Audio Breaks the "Stereo Window" to Immerse Listener in the Performance.

Surround sound adds additional channels to expand the sound field around the listener and refine the perceived position of sounds. Perhaps the most important is the center channel, which allows the localization of on-screen dialogue (in film or video) and eliminates the movement of the center sound image with head position that a listener experiences with stereo.[1] Adding two surround speakers extends the sound around and behind the listener, forming the common 5.1 system.

The introduction of Blu-ray discs allowed consumers to experience 7.1 surround using additional rear or front speakers. While 5.1 systems primarily use the ITU standard speaker layout, 7.1 and higher systems are not yet standardized, though Blu-ray 7.1 commonly uses an extension of the ITU layout, as shown in Figure 2.

The placement of the surround speakers is a compromise between providing envelopment by providing sound from behind the listener, and from extending the front sound image through placement at the sides of the listener. For 5.1, the ITU standard[2] specifies the surround speakers to be placed at an angle of 100-120 degrees. Common recommendations for 7.1 systems in the Blu-ray configuration call for the side speakers to be placed at 90 to 110 degrees and the rear speakers to be placed at 135 to 150 degrees.

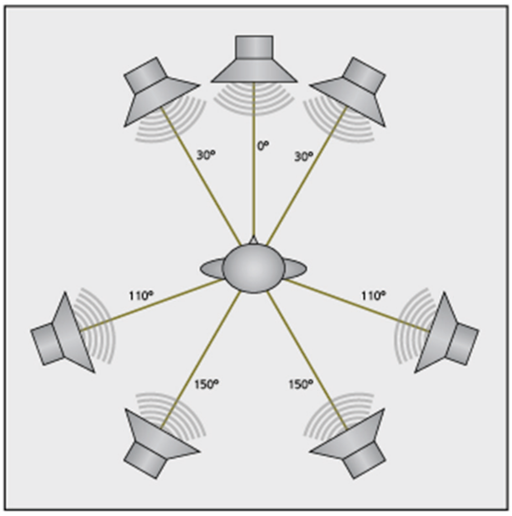

Figure 2: Common position of Blu-ray 7.1 Speaker Arrangement for Surround.
Without The Rear Speakers at 150°, this is the ITU Standard 5.1 Layout [3].

These configurations, mono, stereo, 5.1, and Blu-ray 7.1, are what are available to consumers today. Many other systems for surround have been proposed in the past: Some for theatrical use, such as Cinerama or Dolby Surround, some for consumer use, such as quadraphonic systems in the 1970s, and some mainly of theoretical interest, such as ambisonics.

Future surround systems may offer increased realism through the use of more channels. The benefits are greater envelopment or feeling of being present in the recording location, increased precision of localization of sounds, and increased creative range. Although a 5.1 or 7.1 channel system offers listeners a good sense of envelopment, the sound image is still confined to the horizontal plane, and room reflections from the upper walls and ceiling are not rendered properly.

Systems in development have begun offering height information, with the 22.2 system promoted by NHK including three levels of speakers.[4] ITU-R BS.2159-4[5] offers an overview of several of these systems in development.

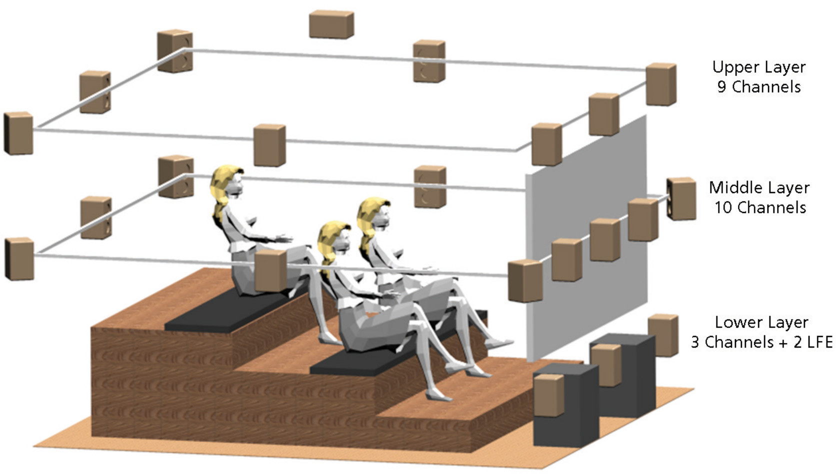

Figure 3: 22.2 Speaker Configuration

Figure 4: Height speakers integrated into Audi SUV passenger A-pillar (view is above dashboard with windshield to left and passenger window to right).

2. SURROUND APPLICATIONS

The primary use of surround sound today, aside from in cinemas, is for experiencing films on Blu-ray or DVD or broadcast TV content in a home living room. Surround sound is also used by many gaming enthusiasts for its increased realism. The AAC codec family is a part of DVB and ISDB standards and is used today in the UK, Japan, Brazil, and other countries for surround TV broadcasting.

As consumers move to consuming content on tablet computers and mobile phones, surround sound must evolve to more flexible rendering techniques to enable consumers to experience cinematic quality in a more personal setting. Fraunhofer is working with device and operating system manufacturers to enable this through two technologies. One is the inclusion of multichannel AAC decoding in the iOS and Android operating systems. This enables content to be consumed on the device while mobile, yet be continued in a living room environment over a large monitor/TV and surround speakers. The other is the provision of virtual rendering technology to enable surround listening over earphones.

An increasing use of surround is in the car. Car listeners are always outside the sweet spot with traditional stereo listening, and cars already include rear and usually center speakers. Surround sound helps to enlarge the sweet spot and increase the feeling of envelopment beyond that of a front/rear speaker fader. Some premium cars have been equipped with discrete surround systems for DVD and DVD-Audio playback, but most cars today create a surround signal by up-mixing stereo content from CD or iPod sources. It is envisioned that more cars will support true surround playback offered as surround radio or internet broadcasting develops.

Since a particular car model offers a challenging, but uniform environment for sound reproduction, Fraunhofer has begun offering 3D surround rendering in some cars through work with automotive OEMs.[6] This includes sophisticated upmixing technology so that consumer's existing content can be experienced in 3D.

3. SURROUND CODING TECHNIQUES

To transmit or record the many channels needed for surround, three techniques have evolved: matrix, discrete, and spatial surround coding. A fourth technique that is beginning initial deployment is object-based coding, which uses elements of discrete or spatial coding.

Matrix surround encoding uses analog-style signal processing to encode surround channels into a stereo signal (usually termed Lt/Rt), basically by applying a 90° phase shift to each surround channel before mixing it with the front channels. Dolby Pro Logic, SRS Circle Surround, and Neural Surround are common examples of matrix encoding. Since they are designed to work with two analog channels, they can be transmitted by any stereo audio codec, including AAC.[7] The two channels are decoded back into four or five channels by adding or subtracting the Lt and Rt signals from each other. The 90° phase shifts of each surround channel result in the surround information being out-of-phase in the Lt/Rt signal, resulting in some cancellation of it when heard through the front speakers.

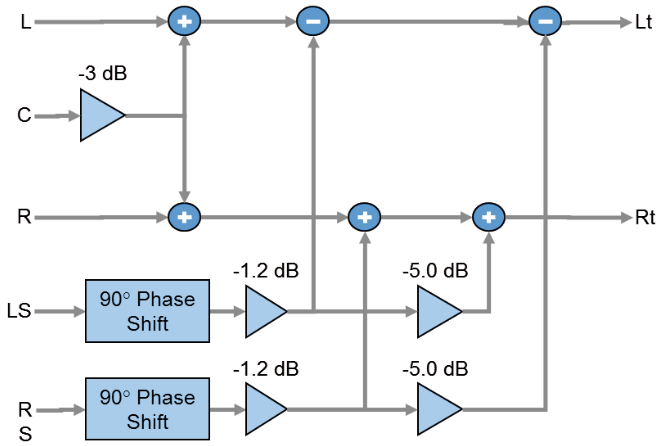

Figure 5: Dolby Pro Logic II Encoder

Matrix codecs are a proven technique and allow playback of the encoded stereo signal without surround decoding, though with possible artifacts. They offer moderate surround performance due to the potential image shifts[8] that can result from "steering" in the decoder to improve channel separation. The channel separation without steering is only about 3dB, while separation up to 30 dB is possible with steering. Another issue is that the downmix is an analog-style sum of broadband channels, so phase cancellation effects can occur when channels are mixed together. Since there are only two signal channels, the audio codec operates in a stereo mode.

The primary uses of matrix encoding have been for theatrical films (Dolby Surround) and broadcasting (Dolby Pro Logic) in the era before digital techniques became feasible.

In contrast to matrix encoding, discrete multichannel transmission encodes each channel (or pair of channels) separately. This offers the highest surround quality and also the highest bitrate. This is the type of surround traditionally used on DVDs with Dolby Digital or DTS codecs, and for TV broadcasting with Dolby Digital or HE-AAC Multichannel.
Since each channel or pair of channels is transmitted separately, there are none of the channel separation or image shift problems encountered with matrix encoding. However, to playback a multichannel signal on stereo or mono devices requires that the channels be down-mixed. This is either done by a fixed downmix in the decoder, or with some codecs through metadata sent with the bitstream. Dolby Digital, AAC-LC, and HE-AAC offer downmix control of the center channel and surround channel levels, and whether to create a Pro Logic-compatible encoding.[9,10]

Spatial audio coding is the newest method for encoding surround. The discrete coding approach makes no assumptions about the signals, while the matrix approach takes advantage of the phase relationships of the direct and reverberant sound in most recordings. In spatial coding, the signals are unconstrained as in discrete coding, but human perception is modeled just as has been done for the perception of mono and stereo signals in codecs such as MP3 and AAC.

Humans seem to use two techniques for perceiving sound direction. One is based on the time difference between sounds arriving at each ear due to the difference in distance between the source and each ear. Another is based on the level differences in sound at each ear caused by the blocking of the sound by the head or outer ear. The time difference is more important at lower frequencies, while the ear uses level differences more at higher frequencies (above 1-2 KHz), where the wavelengths of sound become small in relation to the head.

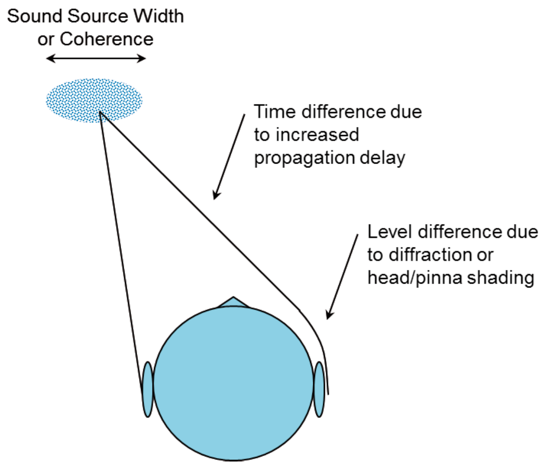

Figure 6: Perception of the Spatial Characteristics of Sound.

These level differences vary with frequency and the height and direction of the sound source, leading to the concept of the Head-Related Transfer Function, or HRTF, which describes the frequency and phase response of the ear as a function of a sound's angle with the listener's head. HRTFs have some variation depending on the shape of the listener's head and outer ear.

These techniques alone do not allow discrimination of sounds that are directly in front or behind the listener, nor much of a sense of the vertical direction. In these cases, the listener usually moves his head slightly to resolve the sound location.

Spatial audio coding separates the coding of the sound from the coding of its spatial direction. This allows the spatial direction of a sound source to be coded based on these perception techniques in terms of level differences between channels and the "width" or coherence of a source. These levels and coherence values, or spatial parameters, can be coded very efficiently into a bitstream of typically 3 to 15 kb/s. The audio information in each of the channels is then combined into a mono or stereo downmix, which can be encoded by a standard audio codec. To take advantage of the ear's limitations, the spatial parameters are calculated separately for each of the ear's critical frequency bands. Note that the downmix is not a simple addition of the channels as occurs in matrix coding, but one that avoids phase cancellation effects by processing the signal in each channel over many frequency bands.

As more envelopment and precision is desired in future surround systems, consideration has been given to an object-based as opposed to a channel-based approach. With object coding, parametric descriptions of a sound source, such as an actor speaking or airplane flying overhead, are sent along with a (usually) monophonic audio signal. The playback device then renders the object with the appropriate direction, distance, and diffuseness on the loudspeakers present. In this manner, very precise localization of sources can be maintained with only a limited number of transmission channels. Practical systems are envisioned as a hybrid of channel-based audio for diffuse environmental sounds and objects for point sources.

An initial object-based transmission system developed by Fraunhofer is Dialog Enhancement, which has been tested in consumer trials by the BBC. In this trial, a sport announcer's commentary is sent as a separate object from the environmental sounds of a sporting event, allowing viewers to adjust the level of the announcer higher (in case of a noisy listening environment or hearing loss) or lower (in case they wish to experience the event with less commentary). This system uses spatial coding for the transmission of objects, so an overall downmix signal is presented to existing receivers without object decoding capability.

4. HE-AAC MULTICHANNEL AND METADATA

Discrete channel coding of surround in the AAC family codecs has been provided since the initial development of AAC. The simplicity and coding efficiency of AAC and its successor HE-AAC make it extremely attractive for both broadcasting and new media application, and today broadcasting in most DVB countries and all ISDB countries use HE-AAC[11], as well as all major PC and mobile operating systems, and the majority of HTML5 web browsers.

The simplicity of the HE-AAC multichannel codec results from its use of copies of the proven stereo HE-AAC codec to carry channels or pairs of channels. A 5.1 surround signal is simply encoded as mono signal for the center channel, and two stereo pairs for the front and rear L-R speakers, plus a reduced-bandwidth mono signal for the LFE signal.

Since HE-AAC is a channel-based codec, a multi-channel signal must be decoded and downmixed for a stereo or mono output. A standard feature of an HE-AAC multichannel decoder is a fixed downmix to create a stereo or mono output:

$$L' = \frac{1}{1 + \frac{1}{\sqrt{2}} + A} \cdot \left[L + \frac{C}{\sqrt{2}} + A \cdot L_S \right]$$

$$R' = \frac{1}{1 + \frac{1}{\sqrt{2}} + A} \cdot \left[R + \frac{C}{\sqrt{2}} + A \cdot R_S \right]$$

Where the gain factor A may be -3, -6, -9 dB or 0 (infinite attenuation) and is specified in the multichannel bitstream.[12]

With multichannel operation comes the expectation of a more engaging experience and one with more artistic control by the creator. Thus, HE-AAC multichannel (and stereo HE-AAC as well) includes optional MPEG and DVB metadata which functions identically to the metadata of the Dolby Digital codec. This allows loudness normalization, downmix gains, and dynamic range control in a manner similar to that used for DVDs.

- ***Dialogue Normalization*** is used to adjust and achieve a constant long-term average level of the main program components across various program materials, e.g., a feature film interspersed by commercials.
- ***Dynamic Range Control (DRC)*** facilitates control of the final dynamic range of the audio and adjusts compression to suit individual listening requirements.
- ***Downmix*** maps the channels of a multi-channel signal to the user's mono or two-channel stereo speaker configuration.

These terms come from the metadata parameters defined for the AC-3 audio codec, used for emission in some digital television systems. They also relate to the Dolby E[13] audio codec, still used in some program broadcast production and contribution chains.
HE-AAC supports these same features with different names. The following table compares the Dolby nomenclature of these parameters with their equivalents specified for the HE-AAC codec.

	HE-AAC	**(E-)AC-3**
Loudness Normalization	Program Reference Level	Dialnorm
Dynamic Range Control		
Light Compression	Dynamic Range Control	Line Mode
Heavy Compression	Compression Value	RF Mode
Downmix	Matrix-mixdown Downmixing Levels	Downmix

Table 1: Nomenclature of HE-AAC metadata in comparison to Dolby metadata.

4.1 Program Reference Level

This parameter may be used to indicate the loudness of a transmitted program, indicating the program loudness relative to full-scale digital signal level in a range between 0 dB FS and -31.75 dB FS. During decoding, this parameter allows the audio output level to be changed to match a given target level. This parameter is expected to remain constant for the duration of each item of program content.

The target level is the intended output level of the decoder. It is a modifiable, external parameter given to the decoder according to the receiver requirement specifications of the related broadcast system. As shown in Figure 7, the audio level is adjusted by the difference between the target level and program reference level so that content in the AAC bitstream at the program reference level will be reproduced at the desired target level.

4.2 Dynamic Range Control

In order to adjust the dynamic range to match listening requirements, HE-AAC is able to convey dynamic range control gains to be applied in the receiver. The application of dynamic range compression is a user-selectable feature of the decoder. However, dynamic range compression is carried out automatically in the case of a downmix in order to avoid overload of the output signal. There are two compression modes available: one for 'light' and one for 'heavy' compression.

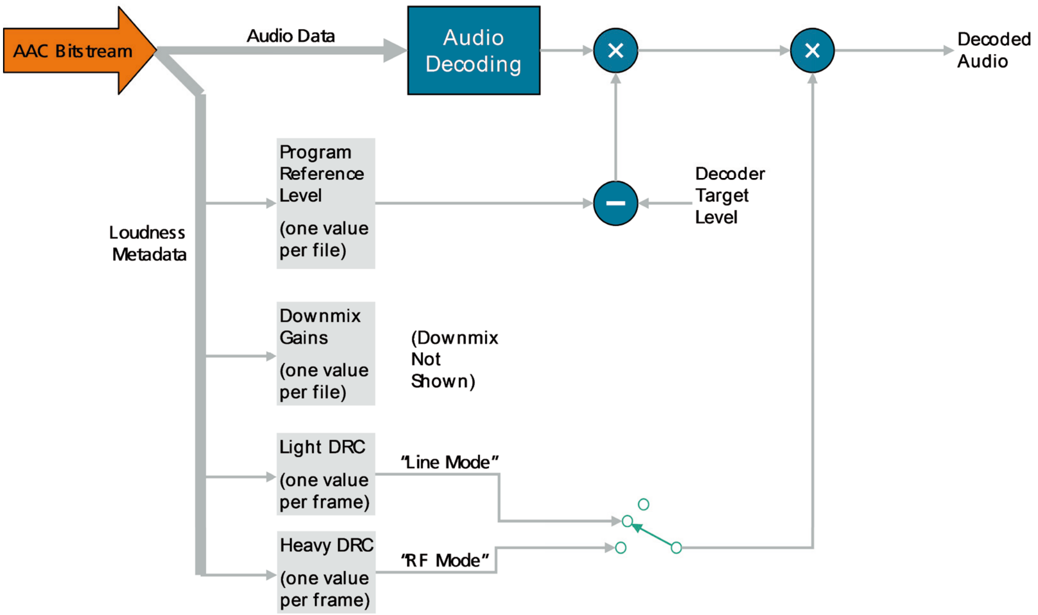

Figure 7: Conceptual Operation of HE-AAC Multichannel and Metadata Decoder.

4.2.1 Light Compression

These values are inserted into the bitstream for each frame according to the desired compression characteristic. The decoder may apply the gains to the audio data in conjunction with the program reference level in order to achieve a reduction of the dynamic range. This parameter is equivalent to AC-3's "Line mode" compression, where the application of dynamic range compression is a user-selectable feature of the decoder.

HE-AAC is capable of applying the compression in a frequency-selective manner for multi-band compression.

4.2.2 Heavy Compression

Some listening environments require a very limited dynamic range. Accordingly, the so-called "midnight mode" of receivers applies a strong compression of the dynamic range using the heavy compression values. This allows, for example, the intelligibility of the movie dialogue at low playback volumes to be improved and simultaneously restricts the volume of special effects such as explosions. This is the equivalent of AC-3's "RF mode" compression.

4.3 Downmix/Matrix-Mixdown

In order to represent a multi-channel signal on a two-channel stereo or mono loudspeaker configuration, the HE-AAC decoder applies a downmix process. Related metadata parameters contain information about the downmix coefficients for the Center channel and Surround Channels and thereby allow a certain level of control of the resulting downmix. The downmix is defined as:

$$L_O = L + center\_mix\_level \cdot C + surround\_mix\_level \cdot L_S$$
$$R_O = R + center\_mix\_level \cdot C + surround\_mix\_level \cdot R_S$$

Where center_mix_level and surround_mix_level may each have the values 0, -1.5, -3, -4.5, -6, -7.5, -9 dB or 0 (infinite attenuation) specified in the metadata. Compression of the downmix to prevent clipping is done by the DRC gain coefficients.

4.4 HE-AAC Metadata in Practice

Depending on the work-flow scenario, a broadcaster or content distributor may choose different methods to obtain the metadata that will be embedded into the emitted HE-AAC bitstream.

One possible route is for metadata to be supplied to the encoder as input parameters, as shown in Figure 8. HE-AAC encoder with metadata fed as encoder input parameters. For example, a broadcaster may select the desired metadata settings individually or choose from a set of pre-defined metadata profiles according to the type of program content, as shown in Table 2. Alternatively, the metadata may originate from earlier stages of the production chain and be carried alongside the audio signal by different interfaces such as Dolby E, HD-SDI[14], SDI[15] or AES-41.

Similarly, file-based production work-flows employ file formats such as the Broadcast Wave File (BWF) or the Material Exchange Format (MXF). Due to the introduction of loudness measurement and normalization, the utilization of loudness-specific metadata based on the ITU or EBU[16] recommendations for loudness increases in importance. Accordingly, these can serve as an alternative basis to generate the metadata for an encoder. For example, the measured program loudness according to ITU-R BS.1770-3[17] is used in DVB based broadcast systems to set the program reference level parameter.

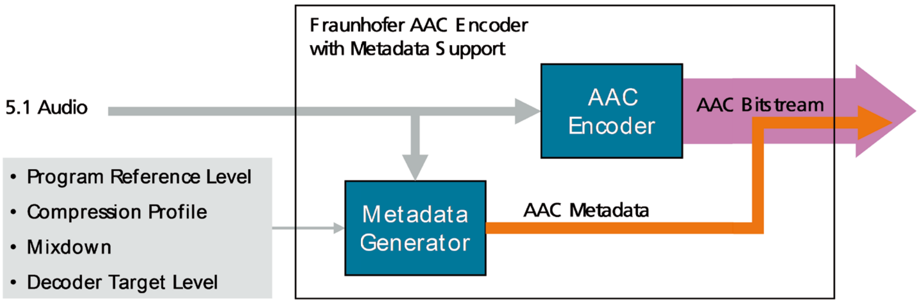

Figure 8: HE-AAC encoder with metadata fed as encoder input parameters.

At the receiver end of the chain, the HE-AAC audio bitstream is decoded to PCM by an HE-AAC decoder. The embedded audio metadata parameters are extracted and applied during decoding. If necessary, the HE-AAC decoder also applies a downmix of the 5.1 surround audio signal depending on the user's audio system and loudspeaker configuration. HE-AAC bitstreams incorporating metadata are compatible with legacy HE-AAC decoders without metadata support. These implementations ignore the metadata information and play back the HE-AAC audio stream without applying the embedded metadata.

Profile	Max Boost	Boost Range	Null Band Width	Early Cut Range	Cut Range	Fast Attack, ms	Fast Decay, ms	Slow Attack, ms	Slow Decay, s	Att Dec Thr, dB
Film Light	6 dB (< -53dB)	-53..-41 dB (2:1 ratio)	20 dB (-41..-21 dB)	-21..-11 dB (2:1 ratio)	-11..+4dB (20:1 ratio)	10	1000	100	3	15/20
Film Std	6 dB (< -43dB)	-43..-31 dB (2:1 ratio)	5 dB (-31..-26 dB)	-26..-16 dB (2:1 ratio)	-16..+4dB (20:1 ratio)	10	1000	100	3	15/20
Music Light	12 dB (< -65 dB)	-65..-41 dB (2:1 ratio)	20 dB -41..-21 dB	none	-21..+9dB (2:1 ratio)	10	1000	100	10	15/20
Music Std	12 dB (< -55 dB)	-55..-31 dB (2:1 ratio)	5 dB (-31..-26 dB)	-26..-16 dB (2:1 ratio)	-16..+4dB (20:1 ratio)	10	1000	100	3	15/20
Speech	15 dB (< -50 dB)	-50..-31 dB (5:1 ratio)	5 dB (-31..-26 dB)	-26..-16 dB (2:1 ratio)	-16..+4dB (20:1 ratio)	10	200	100	1	10/12
None			Full Range							

Table 2: Compression Profiles in Fraunhofer HE-AAC Multichannel and Metadata Encoder.

4.5 *Target Levels, Application Standards, and Presentation Mode Signaling*

Dynamic range control, as currently used in MPEG-DVB metadata, requires the encoder to know the target level to be employed in the decoder. This allows the encoder to control the applied compression to prevent clipping in the decoder on signal peaks. Application standards such as DVB and ATSC define several target levels as shown in Table 3.

dB FS	Standard
-11	Dolby Portable Mode
-14	ATSC Mobile
-20	Dolby RF Mode
-23	EBU
-24	ATSC
-27	THX
-31	AC-3, Dolby Line Mode

Table 3: Decoder Target Levels

The Line Mode and RF Mode were established in the early days of digital audio where compatibility with existing analog equipment was necessary. The Line Mode is intended to provide a dynamic range similar to that of analog audio signals that would appear on the line inputs of an audio amplifier or TV set. The RF Mode is intended to provide a dynamic range similar to that of analog TV signals received by a TV or VCR tuner.

The normal target level when no DRC compression is applied is -31 dB Full Scale. This allows sufficient headroom to reproduce feature films with peaks approaching theatrical levels.

Recently, the EBU initiated a project to harmonize audio levels among broadcasters and consumer devices that resulted in EBU Tech 3344[18], a set of recommendations on operating practices and decoder target levels to insure broadcast content from different channels or sources is reproduced with similar loudness. It currently recommends a decoder target level of -23 dB FS for HE-AAC.

As a result of the implementation of the CALM act in the United States to control the loudness of TV commercials, the ATSC recommendation A/85[19] for loudness measurement has been established. A/85 specifies a target loudness value for content interchange of -24 dB FS.

To account for different receiver designs used in DVB countries, a Presentation Mode bit-field[20] was recently included in the DVB standards. This optionally specifies whether the bit-stream was prepared for either of two receiver types with different downmix strategies.

Presentation Mode Bits	Target Level = -31dB FS, No downmixing	Target Level = -31dB FS, 5.1 to stereo downmix	Target Level = -23 dB FS
0	Not indicated	Not indicated	Not indicated
1 - Mode 1	Light DRC may be scaled, Heavy DRC is allowed	Light DRC is ap-plied, Heavy DRC is allowed	Heavy DRC is applied
2 - Mode 2	Light DRC may be scaled	Light DRC is ap-plied	Light DRC is applied
3	Reserved	Reserved	Reserved

Table 4: Application of DRC values to Multichannel or Stereo Output Signals in DVB IRDs.

In either case, Heavy DRC is applied to a mono downmix. The DVB standard offers a more precise description of each mode, including allowance for scaling of the DRC values (typically a function only of expensive receivers).

4.6 Dolby Pulse

The HE-AAC or HE-AAC v2 codec with MPEG/DVB metadata is occasionally marketed as "Dolby Pulse", as discussed in section 14. The Dolby versions have been tested to be fully interoperable with those of Fraunhofer.

5. THE MPEG SURROUND CODEC

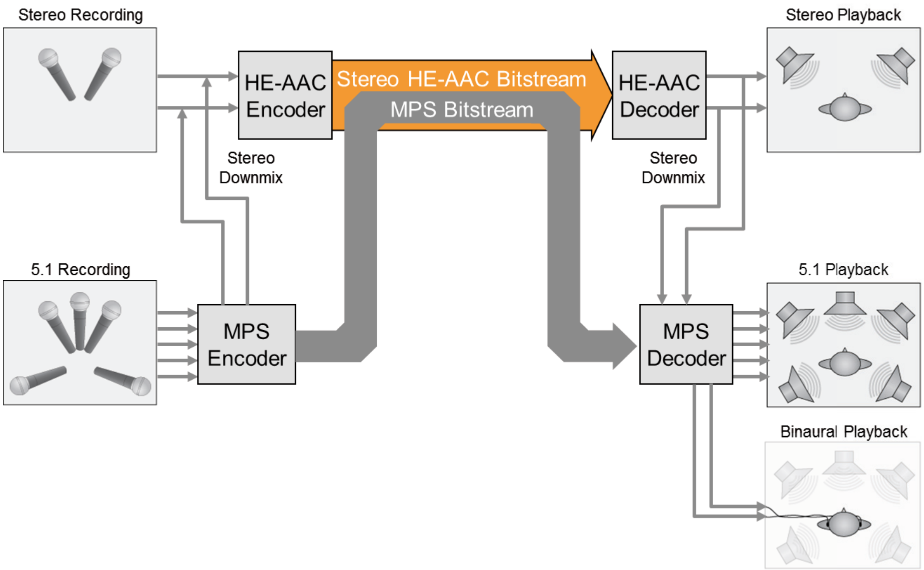

Figure 9: MPEG Surround Encoding and Decoding.

MPEG Surround brings spatial coding to the AAC codec family, as shown in Figure 9. A 5.1 channel signal, for example, is input to the MPEG Surround encoder, which down-mixes the input into a stereo signal that is sent to a stereo AAC encoder and encoded into a normal AAC bitstream. The MPEG Surround encoder also extracts the spatial parameters from the signal and encodes them into a separate MPEG Surround bitstream that may be hidden in the AAC bitstream in a standardized manner. A standard AAC decoder will ignore the MPEG Surround bitstream and decode only the stereo downmix signal. An MPEG Surround decoder will include this downmix decoding, and will then use the spatial parameters to control the remixing of the down-mixed signal back into six channels.

This allows a MPEG Surround bitstream to be used in a compatible way in applications where stereo AAC decoders already exist. Legacy devices will decode only the stereo signal, while new devices decode the full 5.1 surround signal. This combination of stereo compatibility and low additional bitrate have led to the adoption of MPEG Surround for many radio broadcasting standards, include DAB+, DRM+, T-DMB, and ISDB-Tmm.

MPEG Surround also offers a binaural mode, where a realistic simulation of the surround content is rendered over stereo earphones. This is done by using a typical listener's HRTF to simulate the attenuation and delay that a virtual speaker for each of the channels would produce at each ear, as shown in Figure 10. To make a more realistic simulation, room reflections modeled by a simulated listening room's impulse response are included. MPEG Surround allows this to be done in an efficient manner during the decoding process.

Accuracy of the rendering depends on matching the actual listener's HRTF. Often several are provided in an MPEG Surround binaural decoder in order that one close to the listener's may be selected.

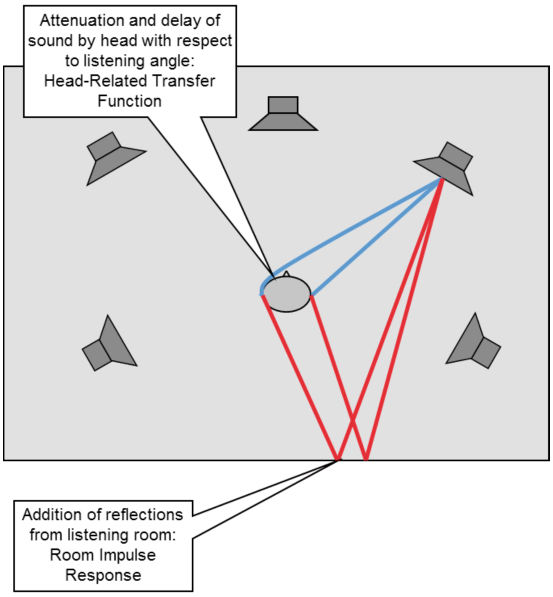

Figure 10: MPEG Surround Binaural Rendering

Just as the sound quality varies with bitrate for audio codecs, the preciseness of the spatial reproduction varies with the spatial parameter bitrate. If a very high spatial quality is needed, residual signals that contain the difference between the input signal and that created by the spatial parameters may also be encoded in the MPEG Surround bitstream.[21] Normally, the spatial bitrate is 5 to 15 percent of the down-mixed audio bitrate, providing "high-quality surround at stereo bitrates".

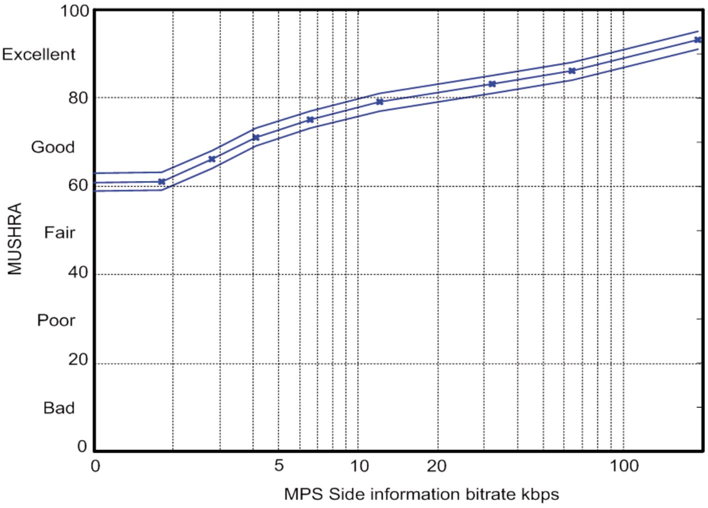

Figure 11: Perceived Surround Image Quality vs. MPEG Surround Spatial Parameter Bitrate with 160 kb/s AAC-LC Core Codec [22].

For cases where surround must be transmitted at low bitrates, for example below 48 kb/s, MPEG Surround may be used with a mono downmix instead of a stereo one. This offers much of the quality possible with the stereo downmix, but cuts the audio codec bitrate in half. The operating mode of MPEG Surround is usually described by the number of input, downmix, and output channels, so for 5.1 sources and a stereo downmix, the codec is said to operate in 5-2-5 mode, while for a mono downmix, 5-1-5 mode is used.

An additional set of tools in MPEG Surround deals with matrix compatibility. For use with legacy equipment where MPEG Surround decoding is not possible, the downmix signal can be encoded to be compatible with the equipment's Pro Logic decoders. During MPEG Surround decoding, the Pro Logic encoding is inverted to eliminate the matrix artifacts.

6. CHOOSING A SURROUND CODEC

Which of the surround coding systems to choose for an application depends on the relative importance of surround quality, legacy compatibility, bitrate, and decoder workload. As Table 5 shows, matrix encoding provides a moderate surround quality that is compatible with existing equipment and requires little processor (computer or DSP chip) workload to implement. If the desire is to just supply a feeling of ambience with content that was authored for matrix encoding, or satisfy the consumer that all speakers are working, matrix encoding is a good choice.

If compatibility with existing receivers and bitrate is less important, discrete coding with HE-AAC or AAC-LC is the proven choice, offering excellent surround quality.

Spatial coding using MPEG Surround offers surround quality approaching discrete coding at higher bitrates, and outperforms matrix encoding at a zero spatial bitrate. Since a stereo downmix is always present, it also offers compatibility with legacy decoders that may already be deployed.

	Matrix	Discrete	Spatial
Applicable AAC Codec	Any	HE-AAC or AAC-LC Multichannel	MPEG Surround
Primary Use	Analog Transmission	TV Broadcasting, New Media Video	Radio Broadcasting
Strength	(Modern coding systems require digital transmission)	Highest quality, wide deployment	Offers compatible downmix to legacy receivers
Surround Image Quality	Moderate	Excellent	Good to Excellent, depending on spatial bitrate
Bitrate (5.1)	Stereo Bitrate	2.5 x Stereo Bitrate	1.1 x Stereo Bitrate
Decoder Workload (5.1)	1 x Stereo Decoder	2.5 x Stereo Decoder	3 x Stereo Decoder

Minimum Bitrate (5.1)	64 kb/s (due to audio codec)	96 kb/s with HE-AAC 384 kb/s with Dolby Digital 200 kb/s with Dolby Digital Plus	48 kb/s with HE-AAC and stereo downmix
Binaural Mode	No*	Yes	Yes
Broadcast Meta-data	No	Yes	No

* External binaural processing could be added after decoding, with additional workload.

Table 5: Comparing Surround Codecs.

7. COMPARING SURROUND CODECS

To confirm the recommendations of Table 5, it is useful to examine the results of listening tests. A test conducted by the European Broadcast Union, summarized in Figure 12, compares many surround codecs.

Like many large listening tests, this one has a few compromises that should be noted. Given the amount of time needed to conduct a large, multi-site test, it usually not practical to repeat the test if a problem is uncovered.

In this test, the HE-AAC codec (not supplied by Fraunhofer) was initially broken and a phase two of the test was conducted to test a corrected version. The Dolby Digital Plus codec was also changed from phase 1 to phase 2, as can be seen in the slight improvement at 200 kb/s in phase 2.

It should also be noted that the anchors were not normalized (such as by scaling the results so the original anchor has a mean score of 100) as commonly done, which limits the ability to compare results from one phase to the other. It is not completely correct to do this anyway, as different test content was used for phase two. However, as shown by the similar scores for the codecs which did not change, AC-3 at 448 kb/s, and AAC-LC at 320 kb/s, the tests show similar results within the usual experimental variance.

The MPEG Surround codec used for this test was a very early version and some quality improvements have been made since that time.

In examining Figure 12, one can see that, as expected, the traditional discrete DTS and AC-3 codecs, operating at the usual DVD or Blu-ray bitrates, provide equivalent and basically transparent sound quality, with AC-3 being only a point or so less than DTS. AAC-LC discrete multichannel at 320 kb/s is also in this range, scoring above AC-3 at 448 kb/s in phase 2, and below AC-3 in phase 1. HE-AAC discrete at 160 kb/s also performed well, scoring similar to AC-3 at 448 kb/s in phase 2. Dolby Digital Plus, a later discrete codec than

HE-AAC, did not perform as well, scoring about 20 points lower than HE-AAC at similar (200 vs.160 kb/s) bitrates, and only about 5-10 points above Pro Logic II and the spatial anchor.

MPEG Surround showed its ability to provide an intermediate range of sound quality when operating at very low total bitrates, surpassing the scores for Dolby Digital Plus at 1/3 of the bitrate using HE-AAC. The test of MPEG Surround and Pro Logic II at similar Layer II bitrates also shows the 25-point improvement MPEG Surround offers over matrix surround codecs.

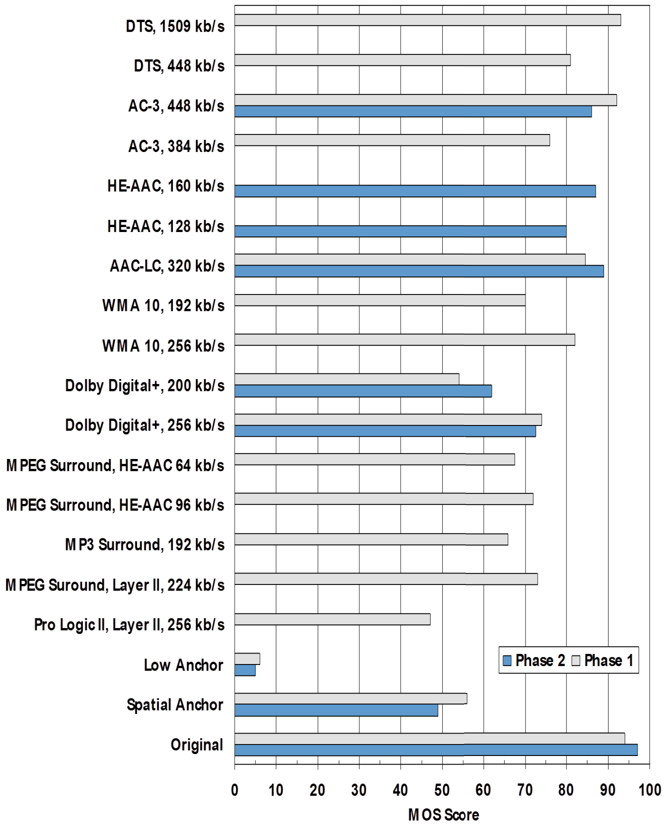

Figure 12: Overall Sound Quality of 5.1 Audio Codecs [23].

8. 7.1 CHANNEL SUPPORT

The AAC codec family has supported up to 48 channels of audio since its initial development through predefined channel configurations and a flexible escape mechanism. The predefined channel configurations from the 2005 version of the AAC standard are:

Channel Configuration Value	Channels	Speaker Mapping	Common Use
0	Escape Code	Escape Code	Escape Code
1	1	Center	Mono
2	2	Left, Right	Stereo
3	3	Center, Left, Right	
4	4	Center, Left, Right, Surround	
5	5	ITU 5.0	
6	5.1	ITU 5.1	5.1 Surround
7	7.1	SDDS	
8-15	Reserved Value	Reserved Value	Reserved Value

Table 6: Channel configurations

This channel configuration value is specified in the AudioSpecificConfig structure of the AAC bitstream. When the channel configuration value is set to 0, the channel configuration is not predefined, but is explicitly described in the Program Config Element structure. This allows arbitrary channel configurations to be used. (For information on these structures, refer to the Fraunhofer Application Bulletin AAC Transport Formats or to the MPEG AAC standard: ISO/IEC 14496-3.)

Supplying the configuration with the PCE escape method is necessary since the only predefined 7.1 configuration is the theatrical SDDS speaker configuration of five front speakers and two surround speakers. Thus, currently the more common Blu-ray 7.1 configuration with three front speakers and four surround speakers is specified using the PCE structure.

An amendment (ISO/IEC 14496-3 2009 PDAM 4) to the AAC standard currently being reviewed will also allow signaling the Blu-ray 7.1 configuration in the channel configuration field by using the value 12 which is currently reserved. However, an AAC decoder will also be required to continue decoding configurations sent in the PCE.

More importantly, AAC currently lacks profiles requiring 7.1 support, with a level 4 or level 5 decoder only required to support 5.1 decoding. The amendment will include level 6 to require 7.1 decoding. Also, the amendment will specify a method for controlling the down-mixing of 7.1 channels to 5.1 channels with controlled gains, much as the existing standard does for 5.1. The amendment will also bring the extra loudness metadata currently specified in DVB into the AAC specification.

For spatial coding, while the MPEG Surround standard is easily extended to support systems above 5.1, Fraunhofer's implementation currently operates with 5.1 channels.

9. PLAYBACK ON CONSUMER DEVICES AND APPLICATIONS[24]

9.1 Native Playback on Windows and Mac

Windows Media Player 12 supports AAC up to 5.1 channels. Windows Media Player 9 and above supports 7.1 playback of WAV files. In each case, it downmixes the audio to match the PC's installed speaker configuration, which may be 1.0 to 7.1. Metadata is not supported.

QuickTime supports AAC up to 7.1 channels and WAV up to 5.1 channels. In each case, it downmixes the audio to match the PC's installed speaker configuration, which may be 1.0 to 7.1. Metadata is not supported.

9.2 Native Playback on Android and iOS

Android 4.1 and above supports 5.1 AAC playback, with downmixing to stereo for local playback and 5.1 output over HDMI. MPEG/DVB metadata is supported. Android's capabilities are provided by the Fraunhofer FDK open-source AAC library.[25]

iOS 5.0 and above supports 5.1 AAC playback with downmixing to stereo for local playback and HDMI output. Metadata is not supported.

9.3 HTML 5 Browser Playback

Popular web browsers (except Firefox, which does not support AAC) offer playback of AAC bitstreams directly in the browser. Current versions of Internet Explorer, Google Chrome, and Safari support 5.1 playback or downmixing to stereo. Metadata is not supported. Fraunhofer maintains a test site at www2.iis.fraunhofer.de/AAC with examples of HTML 5 playback, including setup guides for HDMI playback from PCs.

9.4 Flash

Adobe flash supports AAC up to 5.1 channels. In each case, it downmixes the audio to match the PC's installed speaker configuration, which may be 1.0 to 7.1 Metadata is not supported.

9.5 Television Broadcasting

HE-AAC Multichannel is the choice of preeminent broadcasters around the world, including the BBC, NHK, and TV Globo. It's a part of the DVB standard and is the sole codec of the ISDB and SBTVD standards, and has been adopted by the United Kingdom, France, Spain, Italy, Austria, Nordig (Norway, Sweden, Denmark, Finland and Ireland), Israel, Slovenia, Malaysia, New Zealand, Japan, Brazil, Peru, Argentina, Chile, Venezuela, Ecuador, Costa Rica, Paraguay, Philippines, Bolivia, Nicaragua and Uruguay for television broadcasting.

For more information on the broadcast use of HE-AAC, please refer to the Fraunhofer white paper, "The AAC Audio Coding Family for Broadcast and Cable TV."

10. SURROUND PRODUCTION FILE FORMATS AND TOOLS

10.1 Wave Files, BWF, and RF64

Wave files are the PC's native file format for uncompressed audio, and they work well for stereo and 5.1 content. For 7.1, it is necessary to understand one issue: in a 7.1 wav file, the back channels are ordered before the side channels. This can cause confusion when preparing a 7.1 file in a digital audio workstation.

As explained in the Microsoft documentation[26], a dwChannelMask structure maps channels in the WAV file to speaker locations. It defines locations sufficient for many popular surround formats:

```
#define SPEAKER_FRONT_LEFT              0x1
#define SPEAKER_FRONT_RIGHT             0x2
#define SPEAKER_FRONT_CENTER            0x4
#define SPEAKER_LOW_FREQUENCY           0x8
#define SPEAKER_BACK_LEFT               0x10
#define SPEAKER_BACK_RIGHT              0x20
#define SPEAKER_FRONT_LEFT_OF_CENTER    0x40
#define SPEAKER_FRONT_RIGHT_OF_CENTER   0x80
#define SPEAKER_BACK_CENTER             0x100
#define SPEAKER_SIDE_LEFT               0x200
#define SPEAKER_SIDE_RIGHT              0x400
#define SPEAKER_TOP_CENTER              0x800
#define SPEAKER_TOP_FRONT_LEFT          0x1000
#define SPEAKER_TOP_FRONT_CENTER        0x2000
#define SPEAKER_TOP_FRONT_RIGHT         0x4000
#define SPEAKER_TOP_BACK_LEFT           0x8000
#define SPEAKER_TOP_BACK_CENTER         0x10000
#define SPEAKER_TOP_BACK_RIGHT          0x20000
#define SPEAKER_RESERVED                0x80000000
```

If the dwChannelMask has bits set for speakers, the corresponding channels have to appear in the file in the order listed in dwChannelMask. If no bits are set in dwChannelMask, the operating system sends them to the output hardware with no mapping – channel 0 in the file is sent to the first channel of the sound card hardware.

This latter case is common with stereo applications that do not consider this. Using the InfoAudio tool described in section 11.4, we can examine a stereo wav file ripped from a CD:

```
WAVE file: C:\FHG\Test Signals\ChannelIDs\27.wav
   Number of samples : 348540 (7.903 s)  2005-03-01 02:48:24 UTC
   Sampling frequency: 44100 Hz
   Number of channels: 2 (16-bit integer)
File name: 27.wav
Header length: 44
Sampling frequency: 44100
No. samples: 348540
No. channels: 2
Data type: integer16
File byte order: little-endian
Host byte order: little-endian
```

The highlighted line contains no channel mapping information. Contrast this to a properly prepared surround file:

```
WAVE file: C:\FHG\Test Signals\ChannelIDs\ChannelIDNarrati-
on16b441.wav
   Number of samples : 381205 (8.644 s)  2012-07-27T17:56:44-
07:00
   Sampling frequency: 44100 Hz
   Number of channels: 6 (16-bit integer) [FL FR FC LF BL BR]
File name: ChannelIDNarration16b441.wav
Header length: 68
Sampling frequency: 44100
No. samples: 381205
No. channels: 6
Data type: integer16
File byte order: little-endian
Host byte order: little-endian
--Information records--
creation_date: 2012-07-27T17:56:44-07:00
software: Adobe Audition CS6 (Windows)
```

or this excerpt from a 7.1 file:

```
WAVE file: C:\FHG\Test Signals\ChannelIDs\7.1\7.1.wav
   Number of samples : 244863 (5.101 s)  2013-01-17 21:21:45 UTC
   Sampling frequency: 48000 Hz
   Number of channels: 8 (16-bit integer) [FL FR FC LF BL BR SL
SR]
```

A file that lacks dwChannelMask speaker positions is probably not a concern for stereo or perhaps even 5.1, but in the 7.1 case, some content could be prepared for the SDDS or front height speaker layouts. This could be encoded or downmixed improperly if there is no channel mask to specify the layout.

A more interesting problem is that Microsoft has defined the surround channels of a 5.1 file as Back Left and Back Right. On a Windows computer configured for 7.1 speakers, if a 5.1 wav file is played, the surround channels will be played on the back speakers.

Wav files today are constructed with 32-bit pointers, limiting their file size to 4 GB. This limit is removed by the extensions standardized by the EBU in the BWF (EBU tech 3285) and RF64 (EBU tech 3306) formats. Most DAW software supports these extensions.

10.2 Apple Files

The Apple CAF file format, successor to the AIFF format, provides similar support for WAV files with a channel map defined as follows:

```
enum {
    kCAFChannelBit_Left                  = (1<<0),
    kCAFChannelBit_Right                 = (1<<1),
    kCAFChannelBit_Center                = (1<<2),
    kCAFChannelBit_LFEScreen             = (1<<3),
    kCAFChannelBit_LeftSurround          = (1<<4),   // WAVE:
„Back Left"
    kCAFChannelBit_RightSurround         = (1<<5),   // WAVE:
„Back Right"
    kCAFChannelBit_LeftCenter            = (1<<6),
    kCAFChannelBit_RightCenter           = (1<<7),
    kCAFChannelBit_CenterSurround        = (1<<8),   // WAVE:
„Back Center"
    kCAFChannelBit_LeftSurroundDirect    = (1<<9),   // WAVE:
„Side Left"
    kCAFChannelBit_RightSurroundDirect   = (1<<10),  // WAVE:
„Side Right"
    kCAFChannelBit_TopCenterSurround     = (1<<11),
    kCAFChannelBit_VerticalHeightLeft    = (1<<12),  // WAVE:
„Top Front Left"
    kCAFChannelBit_VerticalHeightCenter = (1<<13),  // WAVE:
„Top Front Center"
    kCAFChannelBit_VerticalHeightRight   = (1<<14),  // WAVE:
„Top Front Right"
    kCAFChannelBit_TopBackLeft           = (1<<15),
    kCAFChannelBit_TopBackCenter         = (1<<16),
    kCAFChannelBit_TopBackRight          = (1<<17)

};
```

10.3 Digital Audio Workstation Sessions or Projects

Many Digital Audio Workstation applications have their own file formats for storing the state of an entire session or project. Often this is a master file that contains references to other track files containing the source audio. Storing content as a project works well for work on a specific system, but may lead to problems in exchanging content with others using the same DAW software. The channel mapping within the application must be set properly so that the channel plays from the same speaker on both workstations. A foolproof but tedious way to solve this is to export individual channels of a mix in separate, labeled wav files.

One DAW application used within Fraunhofer is Adobe Audition. The current version, CS6, directly supports 5.1 panning and mixing and reads and writes 5.1 WAV files. An undocumented but intentional feature allows it to support reading, writing, and simple editing of any wav file configurations supported by the dwChannelMask explained in section 11.1. Audition will edit and write existing WAV files in the format they are read. So, to create a 7.1 file, one may read an existing file, delete the contents, if any, add the desired content, and write the new file.

To do this, an existing file must be obtained in the 7.1 format. This can be created by use of the CopyAudio tool described in section 11.4 to copy a mono file into 7.1. For example:

```
c:\util\AFSP\CopyAudio.exe -S „FL FR FC LF BL BR SL SR“
--chanA=A --chanB=A --chanC=A --chanD=A --chanE=A --chanF=A
--chanG=A --chanH=A  „silence.wav“ „7.1proto.wav“
```

where silence.wav is an input mono file, containing silence or any content, and 7.1proto.wav is the file to be read into Audition. Individual channels can be edited by disabling all channels and then enabling the desired channel.

Editing in this way has some limitations:

1. There is no Audition surround panner for 7.1 or above. An external plug-in may be used.
2. The channel order is that specified by dwChannelMask and the CopyAudio documentation. Rear surround channels will always be before side channels.
3. The names of the channels in Audition will not always be those of dwChannelMask, a mixture of abbreviations from several standards is used. Back Left will appear as Ls and Side Left as Lsd, for example.
4. There is no mixdown or downmix operation. However, you can prepare a manual downmix in a 7.1 file, select only the 5.1 channels, and then copy that to a 5.1 file.

There are more elaborate DAW products (such as Nuendo) that support 7.1 production directly. Audition is interesting since it is a simpler product that can also support 3D audio formats using this feature. For example, to create a 7.1 + 4 height channels file, CopyAudio may be used with the parameter `-S „FL FR FC LF BL BR SL SR TFL TFR TBL TBR“`.

10.4 CopyAudio and InfoAudio

Several programs from the McGill University AFsp audio file utilities[27] are commonly used in audio coding research. One flexible utility is CopyAudio, which can do format conversions while copying audio content from one file to another. CopyAudio supports WAV and AIFF files, as well as other uncompressed formats.

CopyAudio is a command-line program that is controlled by several arguments, explained in the AFsp documentation. It allows audio files in WAV or AIFF to be combined or split, to have the gains of channels changed, and to change the sampling frequency or sample width.

The documentation for CopyAudio is fairly complete, but constructing a command line that it will accept sometimes takes some trial and error. Several points to consider:

1. The speaker parameter `–S` expects a following single string containing the codes for each speaker location. For stereo, this would be `–S "FL FR"`
2. There is only one output file. Thus, all other files on the command line except the last are input files.
3. To copy channels in a special manner from one file to another, the channel gains are used. Each channel in the output file is listed and the gains of the input channels that will be mixed to it are then specified. Both the input and output channels are labeled `A, B, C.`. For example, to copy a mono file to a stereo file, the parameters would be `--chanA=A --chanB=A` , since we have one input channel `A`, and two output channels `A` and `B`.
4. All the parameters need to precede the filenames.

InfoAudio is a similar command-line program that outputs a description of an audio file's parameters without modifying it.

10.5 Dolby E

Dolby E is a surround codec used in television broadcast production. Fraunhofer offers utilities to convert the Dolby E or AC-3 metadata to MPEG/DVB metadata for use with AAC.

11. CHANNEL NOMENCLATURE

Several sets of channel names exist for surround systems in different standards. In general, the names fall in two families, the ones specified in IEC 62574 and those used in MPEG. Primarily this is an issue for systems beyond 5.1. In the 5.1 case, the intent of the names is fairly clear and is set by historical precedent:

IEC 62574, (WAV File)		MPEG (ISO/IEC 23001-8), (AAC Bitstream)	
FL	Front Left	L	Left front
FR	Front Right	R	Right front
FC	Front Center	C	Centre front
LFE1	LFE-1	LFE	Low frequency enhancement
BL	Back Left	Ls	Left surround
BR	Back Right	Rs	Right surround

Table 7: 5.1 Channel Nomenclature

The two proprietary standards, WAV and CAF, use the same bit mask values to specify active channels. WAV uses the IEC names. CAF uses the MPEG names. They reference the same subset of channels, and provide all the channels needed for common surround formats up to 11.1, except for the second LFE channel needed in the 10.2 system.

SMPTE 2036-2 is a subset of IEC 62574, using only the names needed for the NHK 22.2 system.

CEA-861-E defines 46 channel configurations mapping to the 8 PCM channels carried by HDMI. The first six PCM channels are statically allocated to the standard 5.1 channels. The two remaining PCM channels carry several different signals to allow use of many 7.1 configurations.

Table 12 shows the names used by many of these standards, listed in channel order, as these standards map a channel to a number or bit field position. The blue shaded area shows the correspondence of the Fraunhofer Level-Azimuth naming convention we use internally to the IEC channel names. MPEG does not use a channel number concept, but states the channels should appear in order from front to back.

It is particularly important to consider the notes in Table 12, as the channel naming used by each standard for the 7.1 case is potentially confusing due to several factors as explained below:

In IEC 62574, LS or Left Surround is defined as an “array of loudspeakers” or diffuse source as in cinema use, while LSd or Left Surround direct is defined as a loudspeaker position for “localization as opposed to the diffuse array” or a point source as in consumer use. It appears the intent of the standard for playback in consumer use is to use LSd as opposed to LS. There are no angular positions specified in this standard, only a conceptual drawing that shows relative positions.

In ISO/IEC 23001-8, table 8 lists the output channel position correspondence between itself and IEC 62574. Notably, the following channel relationships are defined:

MPEG (ISO/IEC 23001-8) (AAC Bitstream)		IEC 62574 (WAV File)	
Ls	Left surround	LS	Left Surround
Rs	Right surround	RS	Right Surround
Lsd	Left surround direct	LSd	Left surround direct
Rsd	Right surround direct	RSd	Right surround direct
Lss	Left side surround	SL	Side left
Rss	Right side surround	SR	Side right
Lsr	Rear surround left	BL	Back left
Rsr	Rear surround right	BR	Back right

Table 8: Correspondence between loudspeaker positions per ISO/IEC 23001-8.

It also has a conceptual drawing that shows general relationships, but no angular positions of the channels.

Although the reader should refer to the actual standards for authority, a conceptual schematic of the channel positions is illustrated below for the left channels, as shown in figure 9 in the ISO standard and figure 3 in the IEC standard:

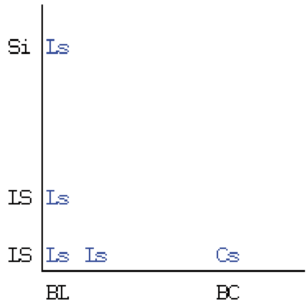

Figure 13: Channel Positions IEC 62574 in black, ISO/IEC 23001-8 in blue.

For the common Blu-ray configuration, it is important to note this relationship:

IEC 62574, (WAV File)		MPEG (ISO/IEC 23001-8) (AAC Bitstream)		CEA-856-E (HDMI Cable)	
BL	Back Left	Lsr	Rear surround left	RLC	Rear Left Center
BR	Back Right	Rsr	Rear surround right	RRC	Rear Right Center
SiL	Side Left	Ls	Left surround	RL	Rear Left
SiR	Side Right	Rs	Right surround	RR	Rear Right

Table 9: 7.1 Channel Nomenclature

Although BL maps to Ls in the 5.1 case, for 7.1 BL maps to Lsr. Additionally, although 23001-8 specifies that SL is equivalent to Lss, for 7.1 SL maps to Ls. Also, although the intent of the standard is to use LSd for point sources, the mapping is to Ls. This scheme has the advantage that Ls and Rs are used in both the 5.1 and 7.1 configurations.

	MPEG (ISO/IEC 23001-8)
	No Channel Numbering, Not in WAV Order
L	Left front
R	Right front
C	Center front
LFE	Low frequency enhancement
Ls	Left surround
Rs	Right surround
Lc	Left front center
Rc	Right front center
Lsr	Rear surround left
Rsr	Rear surround right
Cs	Rear center
Lsd	Left surround direct
Rsd	Right surround direct
Lss	Left side surround
Rss	Right side surround
Lw	Left wide front
Rw	Right wide front
Lv	Left front vertical height
Rv	Right front vertical height

Cv	Center front vertical height
Lvr	Left surround vertical height rear
Rvr	Right surround vertical height rear
Cvr	Center vertical height rear
Lvss	Left vertical height side surround
Rvss	Right vertical height side surround
Ts	Top center surround
LFE2	Low frequency enhancement 2
Lb	Left front vertical bottom
Rb	Right front vertical bottom
Cb	Center front vertical bottom
Lvs	Left vertical height surround
Rvs	Right vertical height surround

Table 10: MPEG Channel Names

12. BASS MANAGEMENT AND THE LFE CHANNEL

The ".1" part of a surround system is often incorrectly thought of as the sub-woofer channel, as it is loosely used to describe the number of sub-woofers in a system. Actually, this is a separate channel with 120 Hz bandwidth used to convey "low-frequency effects." All of the main channels carry full-bandwidth signals.[28]

In a playback system, bass management circuits may extract the low-frequencies that would be lost by or overdrive/intermodulate small speakers and feed them to the subwoofer along with the LFE channel. This is commonly done at 80 (preferred) or 120 Hz in such systems. Actually, the LFE is unneeded in most work, as the main channels can carry all the low-frequency information. Bass management may also feed bass content from all channels to the front left and right speakers if they are full-range and there is no subwoofer. If a discrete surround recording is down-mixed for stereo playback, the LFE channel is usually ignored.

A further point of confusion around the LFE channel is that for playback the LFE channel is boosted by 10 dB relative to the playback level of the other channels. This is based on early theatrical practice, and all surround systems use this +10 dB boost. Unless a mix is being made for theatrical release or special use where separate LFE speakers are used, the LFE channel is best left unused.

In a professional studio, the LFE channel should be calibrated to produce a sound pressure level 10 dB greater than the other channels for the same input signal in the low frequency range. This duplicates the +10 dB gain in the bass management circuits of a consumer receiver or theater sound system. Most professional monitors or monitor controllers have a switchable +10 dB gain switch for this purpose. This should only be turned off if there is already a compensating +10 dB gain elsewhere in the console or monitoring system.

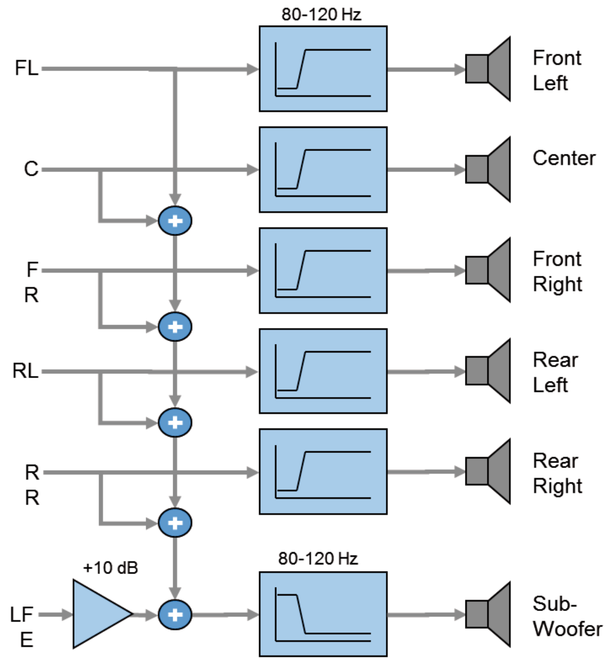

Figure 14: Bass Management in Typical Satellite and Sub-Woofer Loudspeaker System.

13. MONITORING SURROUND SOUND

In a consumer system, the audio-video receiver usually drives the surround sound speakers and provides controls for adjusting the listening environment. Many AVRs now include measurement microphones and software for automatic setup of level, delay, and equalization for each speaker.

In a professional environment, more care is needed than an automatic setup. The sound engineer or tonmeister should consider:

- Overall acoustic response of the listening room, including reverberation decay and early reflections, as well as room size relative to the intended listening environment (consumer living room or cinema).
- Loudspeaker placement at proper listening angles relative to the mix position, considering reflections from walls and equipment.
- Adjustment of loudspeaker levels relative to each other and compensation for unequal loudspeaker distances, if any.
- Placement of the subwoofer to excite the room's low frequency acoustic modes in the optimum manner to provide the best perceived frequency response.
- Adjustment of the LFE channel and subwoofer level. If a consumer AVR is included in the system for listening to prepared content, consideration should be given to management of the +10dB level boost of the LFE channel in the playback from a console or DAW and from the AVR, particularly if bass management in a professional monitor controller or subwoofer is used.
- Normalization of the system gain such that a known reference level, such as -18 dB FS, produces the correct monitoring level, typically 78-85 dBA SPL depending on the intended use and production standard.

These requirements are well-specified by existing standards, papers, and books. A thorough explanation is provided in chapter 2 of Surround Sound by Holman.[29] Requirements for the design of high-quality listening rooms are specified in ITU BS.1116.[30] An introduction to the design techniques of such rooms with modest resources is presented in Everest[31] and for more demanding use in Newell.[32]

Although the design of a listening room fully compliant with BS.1116 can be challenging in some aspects, many existing rooms can be improved with adjustment, equalization and perhaps minor acoustic treatment. Inexpensive measurement microphones and software produced for the audio enthusiast community are readily available and adequate for this purpose.

In any case, monitoring and mixing of content produced for AAC surround encoding is similar to that for other surround codecs or uncompressed audio.

14. EQUIVALENT TRADE NAMES

In general, “AAC” usually means AAC-LC. The Main Profile or Long Term Prediction type of AAC are not used in practice. AAC-LC was standardized in MPEG-2 part 7[33] and extended in MPEG-4 part 3.[34] AAC is an acronym for Advanced Audio Coding.

Some companies have used trade names for their implementations of AAC, as shown in Table 11. All the names refer to the same MPEG standard and are interoperable. Unlike the video part of MPEG, profile names are rarely used (at least correctly) in describing the functions an audio codec implements. Instead, the Audio Object Type ID numbers are used.

	AAC-LC	**HE-AAC**	**HE-AAC v2**
Fraunhofer	AAC-LC	HE-AAC	HE-AAC v2
MPEG	AAC LC	HE AAC	HE AAC v2
3GPP		aacPlus	Enhanced aacPlus
Coding Technologies		aacPlus	aacPlus v2
Dolby			Dolby Pulse
MPEG Audio Object Type	2	5	29
MPEG Profile (rarely used)	AAC	High Efficiency AAC	High Efficiency AAC v2

Table 11: Equivalent Trade Names for MPEG Codecs.

dwChannel Mask	WAV (Microsoft)	CAF (Apple)	SMPTE 2036-2	Fraunhofer Label	Fraunhofer Az °	Fraunhofer El. °	channel number		IEC 62574	CEA-861-F	
0x1	SPEAKER_FRONT_LEFT (FL)	kCAFChannelBit_Left	Same as IEC 62574	M 45	45	0	1	FL	Front left	FL	
0x2	SPEAKER_FRONT_RIGHT (FR)	kCAFChannelBit_Right		M-45	-45	0	2	FR	Front right	FR	
0x4	SPEAKER_FRONT_CENTER (FC)	kCAFChannelBit_Center		M 0	0	0	3	FC	Front center	LFE	
0x8	SPEAKER_LOW_FREQUENCY (LF)	kCAFChannelBit_LFEScreen		Sub L	50	-15	4	LFE1	LFE-1	FC	
0x10	SPEAKER_BACK_LEFT (BL)	kCAFChannelBit_LeftSurround		M 135	135	0	5	BL	Back left	RL	
0x20	SPEAKER_BACK_RIGHT (BR)	kCAFChannelBit_RightSurround		M-135	-135	0	6	BR	Back right	RR	
0x40	SPEAKER_FRONT_LEFT_OF_CENTER (FLC)	kCAFChannelBit_LeftCenter		M 30	30	0	7	FLc	Front left center	RLC|FLC|FLH|FCH|FLH|FLW|RC	
0x80	SPEAKER_FRONT_RIGHT_OF_CENTER (FRC)	kCAFChannelBit_RightCenter		M-30	-30	0	8	FRc	Front right center	RRC|FRC|FRH|TC|FRH|FRW|FCH	
0x100	SPEAKER_BACK_CENTER (BC)	kCAFChannelBit_CenterSurround		M 180	180	0	9	BC	Back center		
0x200	SPEAKER_SIDE_LEFT (SL)	kCAFChannelBit_LeftSurroundDirect		Sub R	-50	-15	10	LFE2	LFE-2	CEA-861 Channel Names:	
0x400	SPEAKER_SIDE_RIGHT (SR)	kCAFChannelBit_RightSurroundDirect		M 90	90	0	11	SiL	Side left	FLC	Front Left Center
0x800	SPEAKER_TOP_CENTER (TC)	kCAFChannelBit_TopCenterSurround		M-90	-90	0	12	SiR	Side right	FLW	Front Left Wide
0x1000	SPEAKER_TOP_FRONT_LEFT (TFL)	kCAFChannelBit_VerticalHeightLeft		U 45	45	33	13	TpFL	Top front left	FLH	Front Left High
0x2000	SPEAKER_TOP_FRONT_CENTER (TFC)	kCAFChannelBit_VerticalHeightCenter		U-45	-45	33	14	TpFR	Top front right	FRC	Front Right Center
0x4000	SPEAKER_TOP_FRONT_RIGHT (TFR)	kCAFChannelBit_VerticalHeightRight		U 0	0	33	15	TpFC	Top front center	FRW	Front Right Wide
0x8000	SPEAKER_TOP_BACK_LEFT (TBL)	kCAFChannelBit_TopBackLeft		T 0	0	90	16	TpC	Top center	FRH	Front Right High
0x10000	SPEAKER_TOP_BACK_CENTER (TBC)	kCAFChannelBit_TopBackCenter		U 135	135	33	17	TpBL	Top back left	FCH	Front Center High
0x20000	SPEAKER_TOP_BACK_RIGHT (TBR)	kCAFChannelBit_TopBackRight		U-135	-135	33	18	TpBR	Top back right	RC	Rear Center
0x80000000	SPEAKER_RESERVED			U 90	90	33	19	TpSiL	Top side left	RLC	Rear Left Center
				U-90	-90	33	20	TpSiR	Top side right	RRC	Rear Right Center
				U 180	180	33	21	TpBC	Top back center	TC	Top Center
				L 0w	0	-15	22	BtFC	Bottom front center		
				L 45	45	-10	23	BtFL	Bottom front left		
				L-45	-45	-10	24	BtFR	Bottom front right		
				M 60	60	0	25	FLw	Front Left wide		
				M-60	-60	0	26	FRw	Front Right wide		
							27	LS	Left Surround		
							28	RS	Right Surround		
				M 110	110	0	29	LSd	Left Surround direct		
				M-110	-110	0	30	RSd	Right Surround direct		
				U 110	110	33	31	TpLS	Top Left Surround		
				U-110	-110	33	32	TpRS	Top Right Surround		
				U 30	30	33			not in WAV		
				U-30	-30	33			not in WAV order	speakerPositions v6 RB	

The names are listed in channel order, as most of the standards map a channel to a number or bit field position. MPEG does not follow this convention, but states the channels should appear in order from front to back. The blue shaded area shows the correspondance of the Fraunhofer Level-Azimuth naming convention we use internally to the IEC channel names.

Fraunhofer channel labels and speaker positions refer to the those for audio systems with more than 7.1 channels. For 7.1 channel systems and below, FL corresponds to the M30 position; FLc, if present, is equally spaced between FC and FL. These positions are often adjusted to place FL at the edge of the display screen. For 5.1 channel systems and below, BL corresponds to the M110 position as specified in the ITU BS.775 standard. In all cases, the position of the right speakers is symmetrical in the horizontal plane to the left speakers.

Table 12: Channel Names in Various Standards.

REFERENCES

1. Audio-only surround mixes can also benefit from use of the center channel to anchor a centered soloist or vocalist, though in current practice, the center channel is either not used or is combined with a center-panned signal in the left and right speakers. This may be due to concerns about consumers not connecting the center speaker, to timbre differences between a panned and direct signal, or to concerns about downmixed outputs for stereo playback.
2. International Telecommunication Union, Recommendation ITU-R BS.775-1: Multichannel Stereophonic Sound System With and Without Accompanying Picture. www.itu.int.
3. One advantage of this specific layout for professional use is it offers direct compatibility with the ITU 5.1 layout. Operation in a strictly 7.1 environment may benefit from reducing the surround speaker positions to 90 and 135 degrees.
4. An alternative approach which does not involve height effects, but improves the accuracy in the horizontal listening plane, is wave-field synthesis. It relies on an acoustic theorem that states that if the acoustic pressure at all of the room boundaries is reproduced, the sound field in the room is recreated exactly. This is true not just for a sweet spot, but everywhere in the room, allowing virtual sound sources to appear to come from a given location that can be walked around as though there was a speaker there. Unfortunately, implementing this just in the horizontal plane requires speakers to be placed every 10-20 inches around the periphery of the room, requiring hundreds of channels. This technique has been used in entertainment venues and high-end automotive experiments.
5. International Telecommunication Union, Report ITU-R BS.2159-4: Multichannel sound technology in home and broadcasting applications. www.itu.int.
6. http://ces.cnet.com/8301-34438_1-57563135/listening-in-on-audi-3d-audio/
7. For codecs operating at low bitrates, the codecs may interfere with the matrix encoding if they use their M/S or PS tools to encode the signal.
8. Several of these effects are discussed in Dolby Laboratories, Inc., Dolby DP 563 Dolby Surround and Pro Logic II Encoder User's Manual, 2003, p. A-7. www.dolby.com
9. Dolby Laboratories, Inc., Dolby Metadata Guide, 2005, pp. 17-18, 21-24. www.dolby.com
10. Fraunhofer IIS, White Paper: Fraunhofer IIS Audio Metadata Technology for AAC Audio Codecs, 2010.
11. The original legacy ISDB system in Japan uses AAC-LC multichannel.
12. Note that this standard dowmix is not loudness-preserving – the dowmixed channels are scaled to prevent possible clipping when the channels are combined. The downmix gains transmitted in metadata are loudness-preserving, as no scaling is applied and potential clipping is prevented by DRC compression.
13. Dolby and Dolby E are registered trademarks of Dolby Laboratories.
14. SMPTE 292M with SMPTE 2020.
15. SMPTE 259M with SMPTE 2020.
16. European Broadcast Union, EBU Recommendation R 128: Loudness normalisation and permitted maximum level of audio signals. www.ebu.ch.
17. International Telecommunication Union, Recommendation ITU-R BS.1770-3: Algorithms to measure audio programme loudness and true-peak audio level. www.itu.int.
18. European Broadcast Union, EBU Tech 3344: Practical guidelines for distribution systems in accordance with EBU R 128. www.ebu.ch.

19. Advanced Television Systems Committee, Inc., A/85: ATSC Recommended Practice: Techniques for Establishing and Maintaining Audio Loudness for Digital Television. www.atsc.org.
20. European Telecommunications Standards Institute, ETSI TS 101 154, v1.11.1 Technical Specification: Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in Broadcasting Applications based on the MPEG-2 Transport Stream, www.etsi.org, section C.5.4.
21. Residual signals are sent in Fraunhofer MPEG Surround encoders when the spatial bitrate is above 12 kb/s. This usually occurs at 96 kb/s with a HE-AAC stereo codec and at 160 kb/s for an AAC-LC stereo codec for the downmix.
22. MPEG Surround Verification Test, ISO/IEC JTC 1/SC 29/WG 11N8851, January 2007, Marrakech, Morocco.
23. From: European Broadcast Union, EBU Evaluations of Multichannel Audio Codecs, EBU Tech Doc 3324, September 2007. www.ebu.ch. The actual test scores were not published in the document, and the values shown here are estimated from the printed graphs.
24. Support in consumer applications and products changes frequently. Fraunhofer's AAC test website at www2.iis.fraunhofer.de/AAC can be used to check the level of AAC support in the latest versions these products.
25. http://www.iis.fraunhofer.de/en/bf/amm/implementierungen/fdkaaccodec.html
26. http://msdn.microsoft.com/en-us/windows/hardware/gg463006.aspx#ECMAC
27. http://www-mmsp.ece.mcgill.ca/documents/Software/Packages/AFsp/AFsp.html
28. For a good discussion of bass management and other surround issues, see Tomlinson Holman, Surround Sound: Up and Running, 2nd ed., Focal Press.
29. Ibid.
30. International Telecommunication Union, Recommendation ITU-R BS.1116 : Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems, www.itu.int.
31. F. Alton Everest, Master Handbook of Acoustics, 4th edition, McGraw Hill.
32. Phillip Newell, Recording Studio Design, 2nd edition, Focal Press.
33. ISO/IEC 13818-7: International Standard 13818-7 - Information technology -- Generic coding of moving pictures and associated audio information - Part 7: Advanced Audio Coding (AAC). www.iso.org.
34. ISO/IEC 14496-3: International Standard 14496-3 - Information technology - Coding of audio-visual objects – Part 3: Audio. www.iso.org.

ABOUT FRAUNHOFER IIS

When it comes to advanced audio technologies for the rapidly evolving media world, Fraunhofer IIS stands alone. For more than 25 years, digital audio technology has been the principle focus of the Audio and Multimedia division of Fraunhofer Institute for Integrated Circuits (IIS). From the creation of mp3 and the co-development of AAC to the future of audio entertainment for broadcast, Fraunhofer IIS brings innovations in sound to reality. Today, technologies such as Fraunhofer Cingo for virtual surround sound, Fraunhofer Symphoria for automotive 3D audio, AAC-ELD for telephone calls with CD-like audio quality, and Dialogue Enhancement that allows television viewers to adjust dialogue volume to suit their personal preferences are among the division's most compelling new developments.

Fraunhofer IIS technologies enable more than 6 billion devices worldwide. The audio codec software and application-specific customizations are licensed to more than 1,000 companies. The division's mp3 and AAC audio codecs are now ubiquitous in mobile multimedia systems.

Fraunhofer IIS is based in Erlangen, Germany and is an institute of Fraunhofer-Gesellschaft. With 23,000 employees worldwide, Fraunhofer-Gesellschaft is comprised of 67 institutes making it Europe's largest research organization.

For more information, contact Matthias Rose, matthias.rose@iis.fraunhofer.de, or visit www.iis.fraunhofer.de/audio.