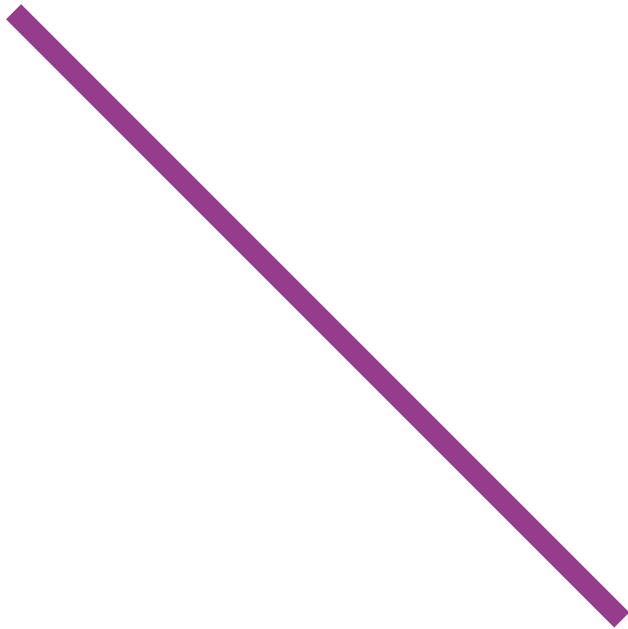


Immersive Audio Series - Part 2



Essential Guide

EG

ESSENTIAL GUIDES

Immersive Audio Essential Guide is a four-part series sponsored by;

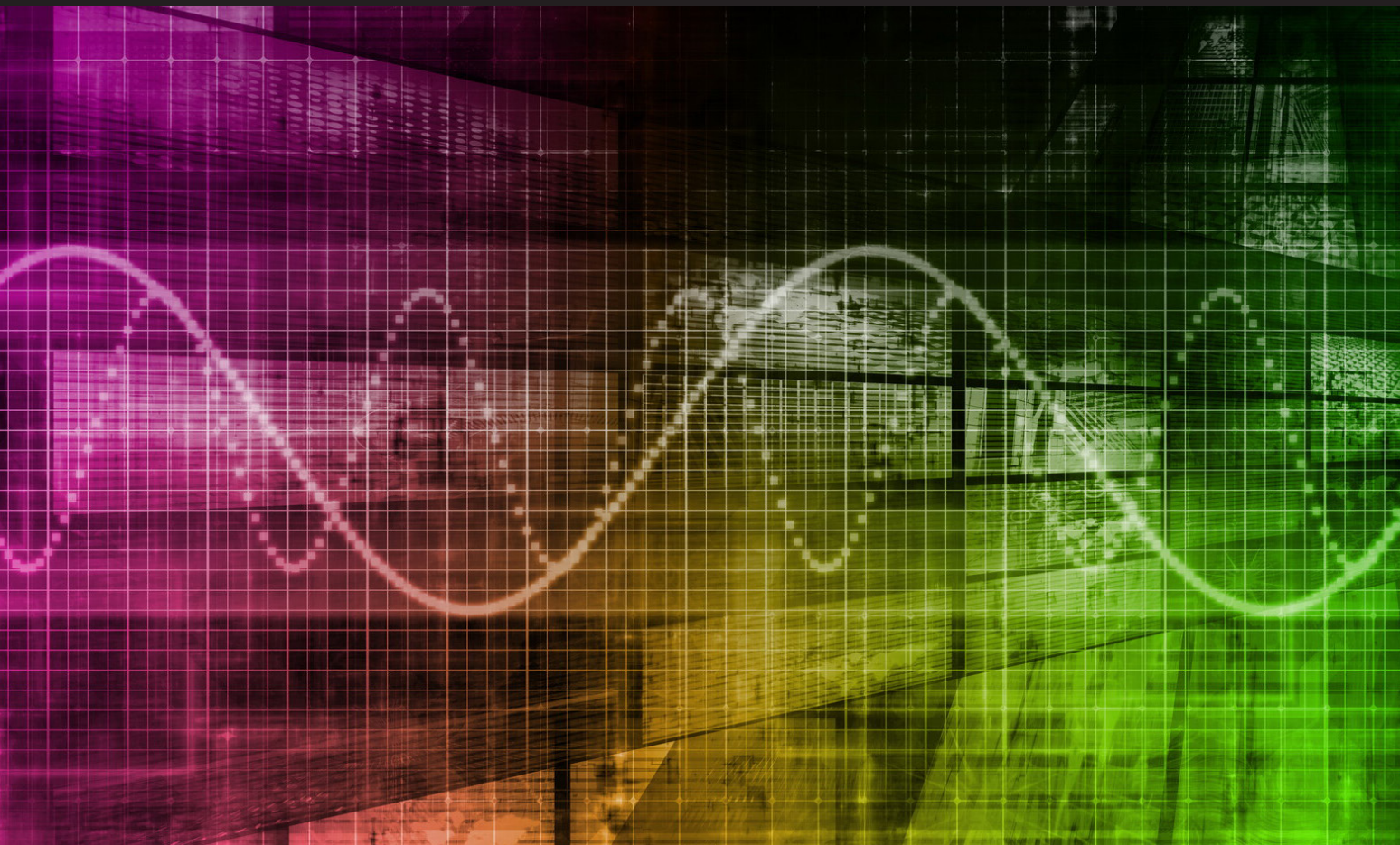


GENELEC®



Register today at TheBroadcastBridge.com to receive your email publication notification for parts three and four.

Immersive Audio Series - Part 2



By Paul Mac – Writer, Professional Broadcast Audio

Listening To The Standards

Immersive audio ‘standards’ and technologies are not so rare these days. Many - the new ones - are part of a new wave of speaker-agnostic, or object-based schemes such as Dolby Atmos, DTS:X, MPEG-H Audio from Fraunhofer, AuroMax from Auro Technologies and Barco in the cinema, broadcast, and home theatre spaces. Live sound has its object-based equivalents too, such as L-Acoustic L-ISA, Meyer Sound’s Constellation, d&b audiotechnik’s Soundscape, Barco’s IOSONO, and several more to-boot.

Both Dolby AC-4 and MPEG-H Audio - as codecs, and interactive object-based schemes with dynamics and so on - between them are specified in the latest ATSC and DVB specs, and DTS-UHD - the ‘high-efficiency’ implementation of DTS:X is also specified in DVB.

This essential guide focusses on AC-4, MPEG-H, and DTS-UHD as distinct options in the immersive broadcast realm and necessarily, albeit briefly, brings in Dolby Atmos - mostly to differentiate between that and AC-4.

As we work through the next instalments, we'll be looking more closely at both the theory and practice of object-based audio, and at the workflows and production tools required to implement it, and the wider standards.

DTS-UHD

Here we explore the DTS-UHD standard that joins AC-4 and MPEG-H Audio in the DVB latest DVB specs.

DTS Inc, acquired by Xperi Corporation in 2016, is a company that is linked inextricably to cinema and home theatre sound, as well as to broadcast. Its own immersive audio 'ecosystem' comes under the general banner of DTS:X and includes a number of sub-brands and technologies such as DTS:X Headphone, and DTS:X Virtual. DTS:X is an object-based technology that includes channel-based provision. It also provides for both lossy and lossless compression.

While not - at the time of writing - mentioned on the DTS website, and actually quite sparsely mentioned online in general, DTS-UHD is also part of the DTS:X landscape and is the 'high efficiency' version, though still featuring both the ACE 'lossy' compression and the XLL lossless compression technologies.

DTS-UHD objects are channel-based audio, discrete objects, and also up to fourth order Ambisonics, and the format maxes out at 224 discrete objects and 32 object groups per stream. Of course, bandwidth will have a lot to say about that. Like the other formats, DTS-UHD includes the concept of 'presentations', which are combinations of other objects and object groups that can be specified in metadata to avoid repeats and increase efficiency. To that end, the spec for DTS-UHD terms any channel of any presentation's object a 'lane', excepting the LFE channel. Thus, a 5.1 + Stereo presentation would use seven lanes.

Be Brave – Dolby Atmos and AC-4

The two newest technologies from Dolby are examined here, how they have evolved, how they connect, and how they are transforming the production and broadcast of immersive audio.

Dolby Atmos came to the fore with the film *Brave* in 2012 - the first cinema release to use object-based mixing. The basic idea is that rather than printing everything to a finite number of ordinary channels, with position determined by the relative levels of a sound in those channels (as panned), different sounds, channels, or groups of sounds can be designated as objects, with their position in the final mix represented by three-dimensional polar coordinates - metadata that stays with the mix.

When it comes to playback, the system itself can determine the best way to represent that position in space, in anything from a full-on 64-speaker Atmos theatre to a binaural representation of that position on headphones. Atmos of course has a channel-based 'bed' in its make up, but object-based mixing and the height (z) axis are the big headlines.

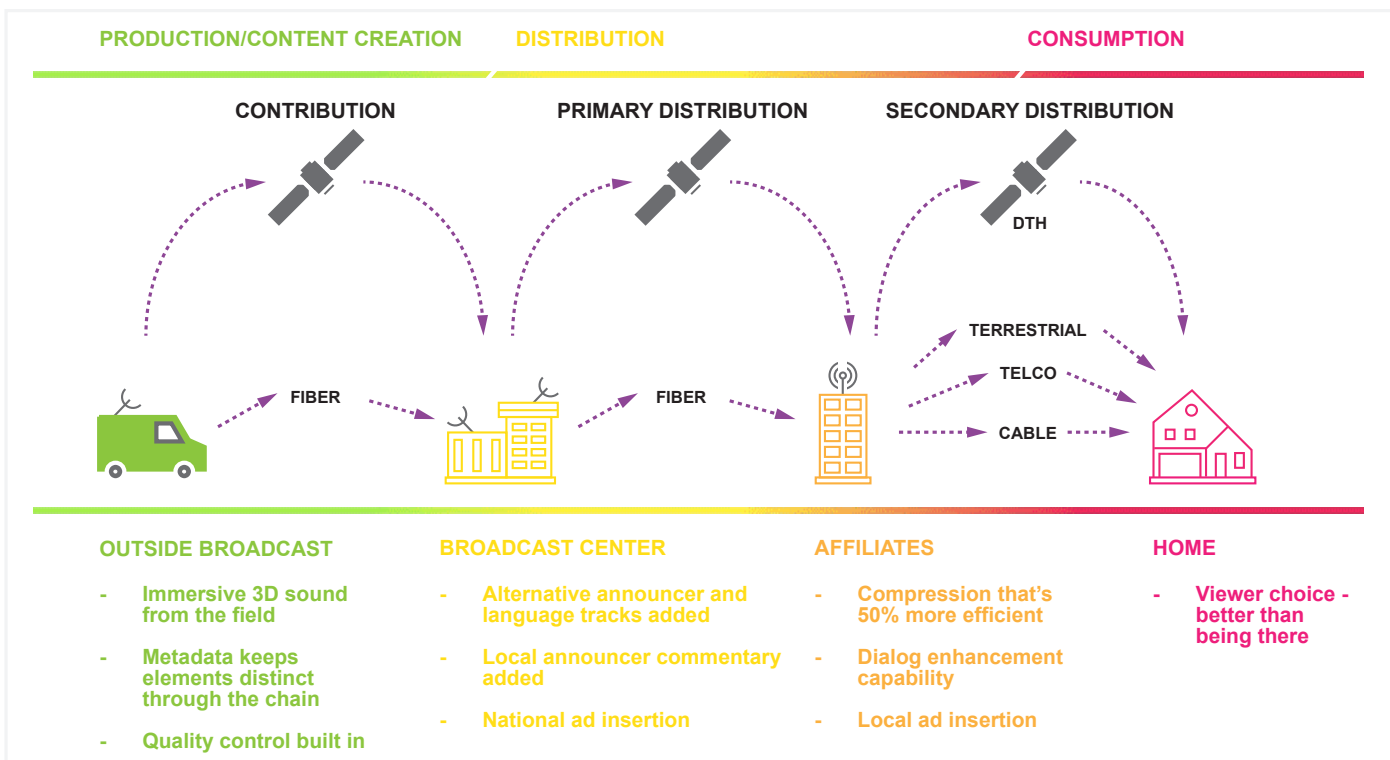
Atmos Here, Atmos There

Dolby Atmos has come a long way since *Brave* was released, and while the cinema is still a core Atmos space, it's the home TV, the computer, the tablet, and the mobile phone that have become the new frontier. A full Dolby Atmos soundtrack for cinema has a lot of data - too much for current domestic delivery formats, so Atmos has evolved to address that and become the catch-all term for Dolby's immersive audio experiences, wherever they are.

Dolby Atmos, for example, can be encoded into Dolby Digital Plus using the Dolby Digital Plus JOC (Joint Object Coding) codec (DD+ JOC), which adds additional sub streams. Dolby TrueHD - the Dolby format used in Blu-ray has a fourth sub stream for a lossless object-based Atmos mix, and MAT 2.0 technology provides the means to move it over HDMI. For broadcast infrastructure, Dolby ED2 is a backwards compatible extension to Dolby E that provides additional audio channels and metadata space to manage not just immersive audio per se, but personalized audio, loudness, and rendering technologies.



The film that started it all – *Brave* (2012). Credit: Copyright 2012 - Disney/Pixar. All Rights Reserved.



The Dolby AC-4 vision: From production to consumer.

The streaming platforms that currently offer Dolby Atmos content include Netflix, iTunes, Amazon Prime Video, and Vudu on a variety of platforms such as Apple TV 4K, Windows 10, Microsoft's Xbox One. There's a growing list of phones that support Atmos, and another growing list of high-profile games that incorporate Dolby Atmos sound such as *Assassin's Creed Origins*, *Shadow of the Tomb Raider*, and *Tom Clancy's The Division 2*. The Dolby Digital Plus encoding is responsible for much of this, as well as the availability on satellite TV services such as premium sports transmissions and so on.

In practice this means that Dolby Atmos can be rolled out anywhere. It isn't exactly the same thing for every platform everywhere, in part because of the delivery mechanisms, but also because requirements, expectations, and playback environments are different.

Dolby is more of a three-dimensional audio delivery ideology - from production workflow to playback, offering up immersive audio to everywhere.

Next Gen Delivery: AC-4.

Moving forward, the next generation of broadcast standards come with an expectation of even better immersive audio provision, including the option of specific mixes for particular playback scenarios (a stereo version of a Dolby Digital Plus 5.1.4 mix is a straight downmix, which is a significant point at the mix stage) as well as the other object-based advantages of translations, multiple-choice commentaries, audio description, and so on. For Dolby, that's where AC-4 comes in.

Dolby AC-4 is the company's newest codec for next-gen video and audio services including the latest specs for broadcast formats like ATSC 3.0 and the latest DVB specs (DASH, MPEG, SI), and much more, including online and streaming platforms.

The main take-home features for AC-4 include a vastly improved compression efficiency that in turn brings the bitrate cost down, native support for dialogue enhancement, intelligent loudness control, advanced dynamic range control, support for multiple languages and descriptive services, personalised audio streams in the form of 'presentations' that other elements that consumers can choose from, and of course, immersive audio – Dolby Atmos.

Codec Technologies

For encoding audio. AC-4 uses a range of strategies to produce efficiencies in data rates and in the processing required to decode, which is important for devices with limited processing power or battery life.

In the first stage of encoding, AC-4 uses two different strategies for general audio and speech. With a more efficient speech process, much lower data rates can be achieved, so stacking multiple dialogue sub streams for multilingual content and alternative commentary becomes more practical. In addition, Dolby employs special high frequency-specific coding.

For both Object and Channel-based stream types, Dolby has its Advanced Joint Coding technologies. In the object stream, for instance, the objects are combined into a smaller set of object groups using a reversible process. This means that basic playback devices can simply decode the smaller object set, or higher end device such as sound bars and dedicated AV receivers go through the process of decoding the original set of objects using a separate data stream for the highest resolution.

Presenting Options

On top of that, the bitstream design of AC-4 means that programme can be defined in multiple 'presentations' that combine audio elements in predefined ways. For example, it's much more efficient to transmit a single M&E mix along with multiple dialogue streams and define presentations simply by using combinations of those, rather than transmitting multiple whole mixes.

There is much more to Dolby Atmos and AC-4 than this, particularly the practical side of actually producing and encoding content for different platforms and programmes - something we will be looking into soon. In the interim there are some fantastic resources online, particularly if you're interested in live TV production as there's a very informative (and free) online course available through the Conservatory of Recording Arts & Sciences (CRAS) at cras-atmos.com.



Stefan Meltzer, Chief Business Development Manager, Audio & Media Technologies Division, Fraunhofer Institute

MPEG-H in Practice

MPEG-H is, strictly speaking, a group of international standards, called 'parts' that together define a broadcasting encoding standard across many different aspects of audio, video, metadata transmission, and more. In the same way that 'MP3' was actually the 'layer III' audio component of the MPEG-1, MPEG-2, and MPEG 2.5 standards, MPEG-H 3D Audio is 'Part 3' of the MPEG-H standards, and MPEG-H Audio is based on that and developed by the Fraunhofer Institute for Integrated Circuits. As a standard it could be seen as a parallel to Dolby AC-4 and is defined separately in broadcast standards like DVB, ATSC, and so on.

Stefan Meltzer is the Chief Business Development Manager, Audio & Media Technologies Division, at the Fraunhofer Institute for Integrated Circuits IIS, which is hub for all things MPEG-H Audio related. Stefan spoke to audio journalist Paul Mac about the standard, and what it means for broadcasters:

Could you update us on the state of MPEG-H Audio, especially on the immersive side?

"MPEG-H 3D Audio is fully specified in Part 3 of the Moving Picture Experts Group (MPEG) MPEG-H standards; it's also fully specified as part of DVB and ATSC in the broadcast. It's already deployed as a broadcast solution in Korea, for example, where it has been used on-air since May 2017 as part of a 4k UHD terrestrial service.

We did a number of trials with streaming at the football World Cup last year, and we have also had trials with EBU and with France TV. In fact, with France TV we did a live production trial at the French Open, transmitting with DVB-T2 during the tournament. We had immersive sound coming from the stadium, the option to switch between English and French commentators, to adjust the level of commentators relative to the court ambience, and even to switch off the commentary completely and bring up the stadium sound.

At the consumer end, we also have lots of devices available, like TV sets from Samsung and LG, and sound bars that support immersive audio, from companies like Samsung and Sennheiser.

Sony announced at CES this year that it is launching a service called 360 Reality Audio - a streaming service for immersive music that will include content from Sony Music and Warner Music, amongst others. That is based on MPEG-H Audio delivery and uses objects for creating the immersive sound."

Can you take us through some of the highlight features of MPEG-H Audio and what's available to the creative and the consumer?

"First, we are not tied to one kind of representation. We allow the transmission of channel-based content, object-based content, and also high-order Ambisonic content - plus a mixture of all those. This opens up many new possibilities. It's no longer simply an audio encoding standard, it's an audio system. You could transmit the ambient sound as a channel bed in 5.1.4 or 7.1.4 and make use of additional object elements for dialogue, audio description, or anything else you need, and ask the user to select which version they want to hear.

The consumer can now also change the volume of the dialogue compared to the ambience, so if you are hard of hearing you can simply raise the dialogue only and improve intelligibility - instead of just raising the overall volume, as consumers tend to do now. The system also has advanced loudness and dynamic range control, which can be adapted to the device and the end user.

Even though we use the immersive audio banner a lot, it's not all about immersion - it's also about convenience. So, the interactivity is an important factor. When we go to consumer electronic shows, people like the interactive elements a lot - most of the complaints that broadcasters get these days are about the audio mix - about dialogue levels and so on. Interactivity solves this problem by giving the user the option to adapt it to personal need.

It's also important because more and more content is being consumed on mobile devices in many different environments - at the park, at home, in a train station. Interactivity means programme can be adjusted to suit that as well.

Things have moved more towards processing on the receiver or consumer end, rather than at the encoding end because we now have a lot more processing capabilities on the receiver end."

The spec itself says it supports up to 128 core channels, what sort of limits are there for broadcasting MPEG-H audio?

"The current DVB spec and what we call the 'level 3' of MPEG-H allows you to transmit up to 12 channels and/or up to 16 objects or up to third order Ambisonics.

You can transmit up to 32 objects, but you can only decode 16 at the same time. If you have multiple language options, for example, you have to transmit all of them, but will only want to decode one."

What do you think drives the choice between using channels, objects, and Ambisonics in production?

"We are seeing high order Ambisonics being used a lot in virtual reality and augmented reality applications, and in the broadcast side we see a combination of channel beds, plus objects for the elements they want to include in the interactivity options.

Of course, interactivity doesn't have to be just about language and accessibility. You can use objects and interactivity to place sound sources creatively. You could, for example, change the listening position from in front of a band to inside the band, or you could take a channel away to allow a karaoke feature.

MPEG-H gives full flexibility to creatives - whatever they want to do. They can even define the limits and scope of interactivity - how much things can be changed."

How do broadcasters implement MPEG-H audio into their workflows?

"We have developed the means to transmit the audio, together with metadata, over SDI. The metadata for MPEG-H audio is modulated on a PCM channel, so it is always in sync with the audio and it doesn't need additional infrastructure because SDI is still the standard.

Of course, it gets even easier when broadcasters move to IP infrastructures, but at the moment it's mostly SDI.

We also have plug-ins for the DAW that allow the creation of the metadata, the positioning of the objects, scene description, and so on; and for the live use case we have what we call the authoring and monitoring units. These are currently available from Linear Acoustic and from Junger Audio. They create the metadata in real time, using scene descriptions entered via the user interfaces.

So, we have full live production and post production workflows in place for MPEG-H audio."

Sennheiser

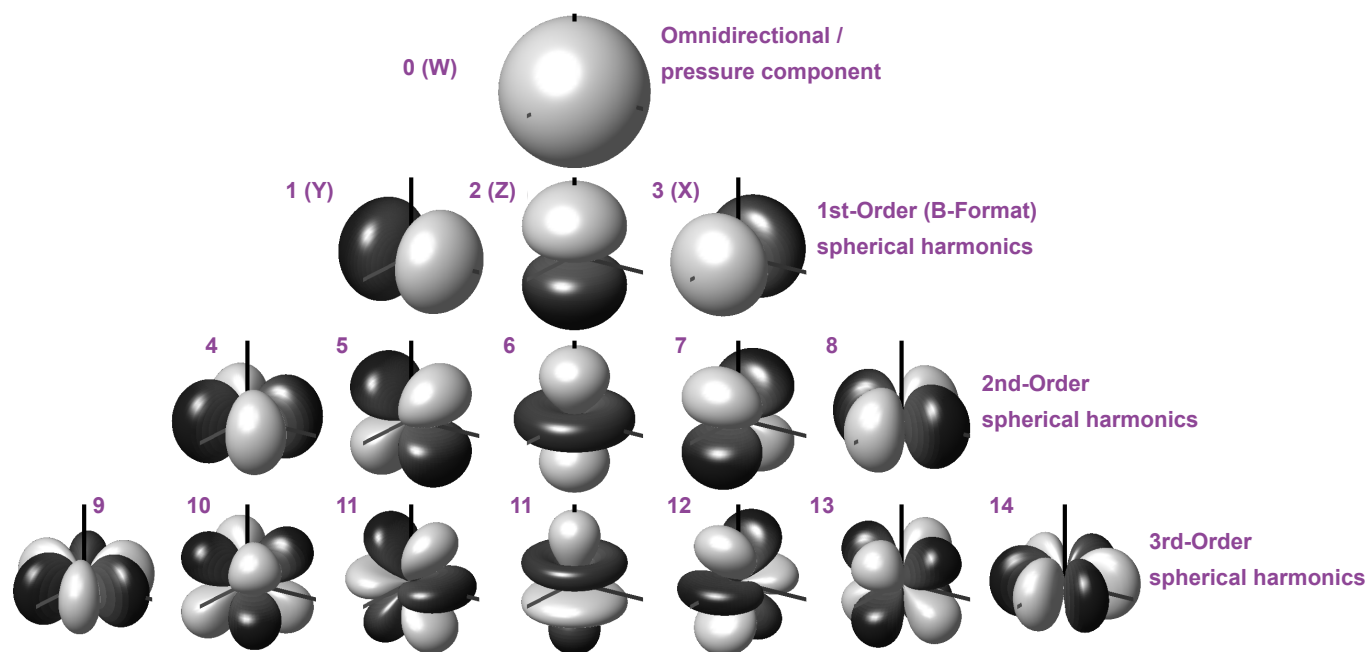
The Sponsors Perspective

An Ambisonics Primer

Sennheiser examines the theory, implementation, and uses of the Ambisonic soundfield, and its important role in the immersive audio world.

Ambisonics is probably the original speaker-agnostic immersive format, and it's been waiting a while for everyone to catch up. If you're familiar with the Mid-Side microphone technique, that gives you an idea of how this format works - in those terms, 'first-order' Ambisonics is essentially a central omni-directional 'mid' or pressure component (W), plus three different 'side' figure-of-eights: Back-front (X), orientated left-right (Y), and up-down (Z). These four signals make up the so-called 'B-Format' first-order Ambisonic format.

This is not an object-based format like Atmos. In fact, if you tried to split Ambisonics up into individual objects with position you would defeat one of its most useful features. All components, together, form the entire soundfield and are, as such, inseparable. However, it is a speaker-agnostic immersive format as it does describe a full 360-degree sound field without referencing speaker positions.



The spherical harmonic components of ambisonics in ACN order, starting with the omnidirectional / pressure component. The two colours represent positive and negative phase - easy to see in the first line made up of what are effectively figure-of-eight polar patterns.



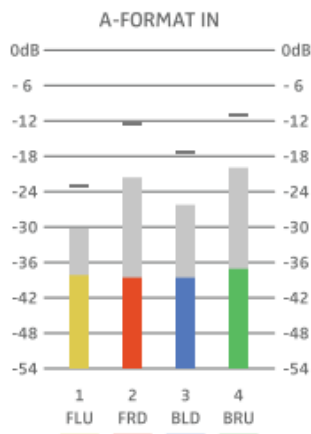
AMBEO

3D AUDIO TECHNOLOGY BY SENNHEISER

AMBISONICS CORRECTION FILTER

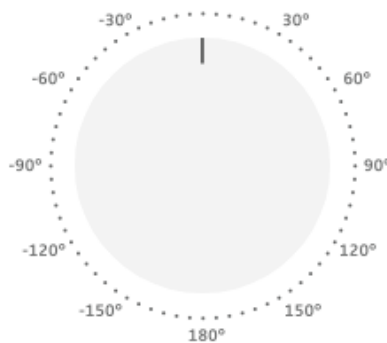
ON OFF

LOW CUT FILTER



MICROPHONE ROTATION

0°



POSITION



UPRIGHT



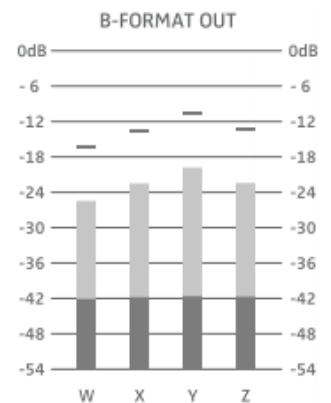
UPSIDE DOWN



ENDFIRE

OUTPUT FORMAT

CLASSIC FuMa ambiX



The Sennheiser AMBEO A-B converter for transforming A-Format channels into B-Format Ambisonics components.

Because of the way this format stores the soundfield, it can easily be ‘decoded’ into any type of speaker set-up or number of speakers and panning and effects can be implemented directly in B-format, which maintains that speaker-agnostic status and explains its starring role in the upcoming 360-degree video boom - especially with live head tracking, which enables audio sources to effectively remain static in the space, while the video reflects the viewing angle. It can also be relatively easily encoded with environment and/or HRTF at the replay end if required to enhance the soundfield for headphones (see Essential Guide “Immersive Audio – Part 1” on binaural audio and the personalised HRTF)

Getting High

Higher Order Ambisonics (HOA) - anything above 1st order - effectively increases the number of ‘sides’ in our virtual Ambisonic microphone (except they’re no-longer figure-of-eights) - a mathematical idea termed Spherical harmonics. As you work your way up the ‘orders’ of Ambisonics, effective resolution of the sound field increases, the sweet-spot gets bigger, and the number of channels required goes up too: For second order Ambisonics you need nine channels, for third-order you need 16.

You don’t need a microphone to create and work with an Ambisonic sound field. There are plenty of Ambisonic panning and processing tools available for different platforms, DAWs, phones, and so on, including headphone encoders for working on Ambisonics when you don’t have the luxury of lots of speakers, along with head tracking options so you can effectively monitor your head-tracking-enabled VR 360 mixes.

Ambisonic audio is specified as an option for both MPEG-H Audio and for DTS-UHD, and therefore can be part of DVB-MPEG/UHD or ATSC 3.0 broadcasts.

Slightly confusingly, standard formats and files for higher order Ambisonics are rather fraught with variations, mainly because there are different options for the derivation and ordering of the spherical harmonic components. The main sequences are ACN and Furse-Malham (FuMa). ACN starts with WXYZ for 1st order while FuMa starts with WXYZ. It’s important to be aware that of the potential for mixing up the order, which will definitely lead to a disappointing, or disorientating, Ambisonic experience. There are also different options for the normalisation of those components such as maxN (for FuMa ordering), SN3D, N3D. and more. Of the proposed file formats, AmbiX seems to be the most popular option and is scalable to any order. It uses ACN ordering, SN3D normalisation, and the core audio format (.caf) container.



SENNHEISER

YouTube and Facebook now support 360 video and Ambisonic audio and in fact there is a free software suite called Facebook 360 Spatial Workstation available for designing spatial audio for Facebook, also compatible with YouTube 360 spatial audio metadata. YouTube's encoding process specifies the Spatial Media Metadata Injector.

Ambisonic Microphones

The standard way of recording Ambisonics has always been a tetrahedral array or cardioid capsules. This was first seen in the Soundfield Microphone, brought to market in the 70s by Calrec. More recently, a good number of tetrahedral array mics have come to market, made economically viable by the upsurge of interest in immersive audio and probably, in particular, the 360 video trend.

The raw audio from a tetrahedral array of cardioid microphones is normally termed 'A-format'. This can then be transformed into the B-Format 1st-order Ambisonic components of W, X, Y, and Z.

The Sennheiser AMBEO VR microphone is one such product and fits into the Sennheiser AMBEO immersive technology landscape along with products like the free AMBEO Orbit plug-in for mixing various sources into binaural audio, plug-ins from it's partner in VR, Dear Reality, the Neumann KU 100 dummy head microphone, and - for the end-user - the high-end Sennheiser AMBEO Soundbar.

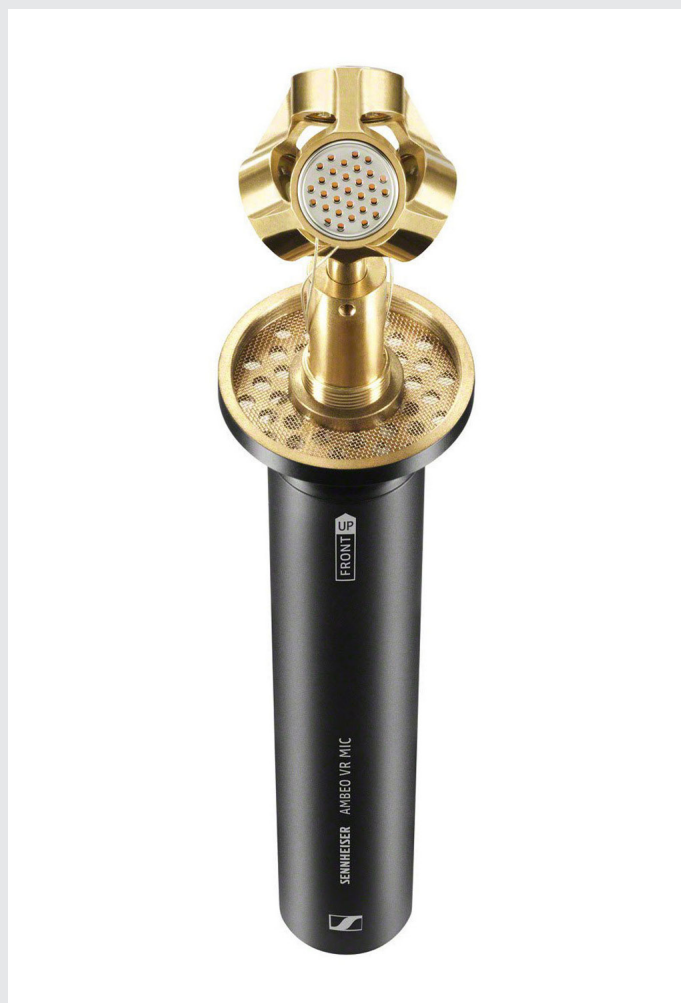
The AMBEO VR microphone uses four matched KE 14 capsules and outputs four corresponding audio channels for the A-Format feed. It also comes with the A-B converter tool for getting the A-Format signal into a DAW in B-Format with various adjustments, such as FuMa or AmbiX ordering / normalisation, microphone position, and filters.

Ambisonic Potential

The rise of Ambisonics has been a long-time coming. The very fact that people are waking up to the advantages of speaker-agnostic immersive audio, and that the consumer now has the technology and every opportunity to experience it in many convenient forms, is driving this boost.

It fits very nicely into the grand immersive scheme along with object -based audio, channel-based beds with height, and with binaural audio for headphones, which is why it's included in the MPEG-H Audio and DTS-UHD specs. A-format capture is well-suited to encoding into channel-based bed as well, so even if you didn't want to include the raw Ambisonic channels, the techniques and technology can be the basis of a high-quality ambience feed for sports broadcast and so on.

Ambisonics should be a valuable part of your immersive audio toolbox.



The Sennheiser AMBEO VR microphone for Ambisonic recording.

Genelec

The Sponsors Perspective

Sound All Around At The Farm

The Farm is a leading full-service post production group that started in Soho, London, subsequently opening facilities in Manchester and Bristol in the UK, and West Hollywood in Los Angeles. The Farm is well known for its ground-breaking creative approaches and its willingness to embrace all that technology has to offer to further that cause. The Farm team has been responsible for some of the most popular television projects of the last 20 years and has an array of awards to prove it.

As part of its continuing mission to stay at the forefront of post production, The Farm has invested wisely in the studios and expertise necessary to satisfy an accelerating demand for immersive audio. Leading that push is The Farm's Head of Audio, Nick Fry. "We have three Home Atmos studios in London," he notes. "We have another in Manchester, and one more in Los Angeles. They're all fully certified Dolby Home Atmos rooms with 7.1.4 Genelec monitoring systems and full Atmos workflow, including Dolby HE-RMU rendering and mastering units." The London and Los Angeles spaces were upgraded from 7.1 to 7.1.4, while the Manchester rooms were built from scratch, with Atmos in mind.

Fry: "Dolby goes through a lengthy process of specifying the rooms. They've been great - helping us, going through the plans, working out where to put the speakers and how we put them up. In most of the rooms we have Genelec 1037C speakers at the front, 8040Bs for surrounds and height, and the 7071A sub. One studio had a lower ceiling height and for that we found the Genelec AIW26B in-wall speaker. It's fantastic, and only about 10cm in depth, which that helped us get the height requirement for Dolby certification... And they look really good!"

On With The Mix

These studios represent a big investment for The Farm, so it must be sure that the work will follow. So far, the immersive clients have been predominantly streaming services and savvy broadcasters like Netflix and Sky. "We've just delivered two particularly interesting projects to Netflix," says Fry. "One is the *Formula One: Drive to Survive* series, and the other is *Turn Up Charlie*, a comedy series based on the experiences of a DJ, played by Idris Elba. *Drive to Survive* was the first unscripted documentary series in Dolby Atmos for Netflix in the UK."



Netflix series *Turn Up Charlie*, available with Dolby Atmos sound.

Both series were mixed fully natively in Atmos - not up-mixed - and both have been fundamental in enabling The Farm to push the boundaries further in delivering immersive audio that brings viewers back for episode 2. "It's about picking the right sounds - the objects - that you want to move around the room," notes Fry. "Otherwise it just becomes a mush. And it does also depend on the programme. We used a lot more objects in *Drive to Survive* than we did in *Turn Up Charlie* due to the content, though the club scenes in *Turn Up Charlie* do sound amazing."

Workflow Considerations

In starting out with Atmos, Fry had to consider lots of workflow options and possibilities, particularly around track-lay, sound design, and creating deliverables. "It's been a big learning curve," he says. "But it has also been a very collaborative road. Netflix has been a great help, as has Dolby, and the London based systems reseller, Scrub."



Netflix series *Formula One: Drive to Survive*, available with Dolby Atmos sound.

“Typically, I do a 7.1.2 mix pass to start with but with the music as objects, the music is no longer channel based then, and the bulk of the mix goes through as normal. Of course, the dialogues tend to be in the centre anyway, and they don’t move on the whole. I’ll get the mix sounding as great as I can, as a foundation, EQ’d and balanced, then I will go back and think about what I can do for the immersive part - the dusting on top of the cake, the fun bit.”

When asked about particularly stand-out examples of immersive audio from the TV series’ he has completed, Fry cites an episode of *Drive to Survive*, focussed on an intense trip around a city F1 track: “At one circuit you’re hemmed in on all sides with the barriers and bridges going past. We put effects on all of those and made them objects- whizzing past overhead and to the side. We also put the sound of the car and the radio communications between the driver and the pit crew in the centre of the room, again as objects. It does sound like you’re in the car.

“And then when you cut back outside the car, all the sounds change direction again. The work we put into that was worth it. It does sound amazing.”

Up until now, Fry has worked with standard sources from both production sound and Foley, though he does have a project coming up that will embrace Ambisonic sources: “The project has no dialogue,” he explains. “Unusually, I had the chance to meet with the sound recordist, who is going to record as much as possible of the programme with an Ambisonic microphone, plus a directional microphone aimed at the action. Then we’re going to recreate that in Atmos. I’ve also met with the Editor and gone through workflows.”

It does seem though as if this approach might remain an infrequent adventure, with much of the work on immersive audio for TV remaining in the mix room.

Watch *Formula One* and *Turn Up Charlie* now streaming on Netflix.

Immersed

An encouraging aspect of The Farm’s move to immersive and object-based mixing has been a willingness to embrace its advantages as a mix medium for all-occasions. “I would happily mix most things in Atmos,” says Fry. “I think it’s a natural transition from the days when we were mixing everything in stereo and then decided to mix everything in 5.1. The extra time required for Atmos is going down all the time as I get more proficient.

“We’ve got a good Atmos workflow now using templates and we’re getting more and more mixers trained up in Atmos as we go along, and it really does add another dimension to the mix for most shows. The way I explain it to clients is that Atmos is the top of the pyramid. From that point we can create any mix version - even play it in a cinema. “And of course, once binaural mixes start to get broadcast as well - that will be really exciting. It’s definitely being considered... It’ll work out what device you’re listening on and if it’s possible, it’ll pick the binaural version for use on headphones.”

Fry considers immersive audio to be a valuable creative tool, and object-based mixing as an important step in transportable content: “It’s about learning that you can put things around you and above you that you can’t necessarily see. You can use Atmos as an effect in its own right, or use it to rebuild an entire soundscape... Immersive audio adds another level of storytelling. It makes a huge difference.



Netflix series *Turn Up Charlie*, available with Dolby Atmos sound.

Lawo

The Sponsors Perspective

Immersive Audio from The A1's Perspective

Felix Krückels is a certified audio engineer who graduated from the Detmold University of Music and has been involved in immersive audio since 2012. He was there when NHK launched its Super Hi-Vision project with the help of Lawo.

In 2018, he became professor for Broadcast Production and System Design at the University of Applied Sciences in Darmstadt (Germany) where he is conducting research into new dimensions of immersive audio. A major highlight of his career was as A1 at the soccer world cup 2018 in Russia, which was produced in Dolby Atmos.

Complexity, Abstraction

Asked whether mixing immersive audio is more demanding than working in 5.1 or stereo Krückels states that the most important challenges come with the 'bonus features' afforded by immersive audio broadcasts. Thanks to object-based audio, consumers can individualize the streams they receive changing ambience and commentator levels, and so on, so the engineer always has to bear that in mind.

On the other hand, he notes that with Next Generation Audio (NGA) object-based production, it's no longer necessary to focus on several discrete listening scenarios: The engineer creates the 3D space from objects and the end user is at liberty to decide what to listen to. Whether it's binaural headphones, two speakers, ten speakers, or a sound bar, the consumer decoder mixes in tandem with the A1.

Of course, an audio engineer always needs to check typical reference scenarios and for Krückels those are typically 5.1.4, 5.1, stereo, and binaural, which should be checked with different NGA 'presentations' afforded by the interactive side of NGA.

Pub-Sound

A good example of a useful presentation for sports is a creation by the Dolby team and Felix Krückels called the 'Pub Presentation'. When a football match is shown at a public bar, the actual crowd at the sports event becomes barely audible because the cheers and boos of the audience in the bar are the dominant crowd noise. Therefore, field-of-play audio details, like ball kicks, tackles, whistle and moaning noises, need to be more prominent.



"The ultimate goal for an A1 in TV live coverage," says Krückels, "Is to exaggerate all relevant noises to such an extent that viewers at home have the illusion of being at the venue where the soccer game, boxing fight, etc., takes place."

The interactive features of 3D/immersive audio can be a blessing in disguise: Any added flexibility afforded to the consumer might attract tweaks and changes to the point where the audio content and the experience is ruined. This explains why sound supervisors usually favor a slightly conservative approach, with relatively few bells and whistles. They know that they are unable to control what viewers at home do with the presentations, and so limit the options.



The actual audio production as such, is relatively straightforward and very similar to 5.1 production. The added height dimension, from an operator’s point of view, can be managed quite easily. The big new considerations include how the additional options are presented to end consumers, how to monitor the presentations, and which individualization options to make available to the general public in the first place.

Doing It

So, how does one approach a 3D/immersive audio mix for sports broadcasting?

The first step, says Krückels, is to look for a venue’s ‘sweet spot’: The position where one can hear everything. That is where a 3D microphone should be mounted. It turns out that this position is usually located close to camera 1.

The 3D microphone is suspended from the roof, at a suitable distance from the crowd to avoid too much interference from drums, vuvuzelas, offensive language, and so on. The 3D microphone thus serves the same purpose as a suspended microphone used to capture the overall sound of a symphonic orchestra.

For reasons of intelligibility and flexibility, spot microphones are positioned close to all important sound sources. The resulting signals are combined in such a way that they make acoustic sense coming out of nine speakers.

Felix Krückels likes to work with three ‘planes’ for his mixes: a mono, a stereo and a surround/3D plane, created from several aspects, or source types.

For him, the surround/3D information usually only concerns the ambience (crowd, city noises, etc.). He hardly ever uses novelty effects or dramatic pans, though some broadcasters use the occasional ‘whoosh’ to announce slo-mo instant replays and the like. To this he adds - usually in mono - typical field-of-play noises - the signals captured by or directly associated with microphones close to the cameras. And finally, there’s the narrator or commentator. Krückels takes great care to maintain the separation among these three aspects to leave sufficient room for artistic license and alternatives.

He considers it important to have a stationary position for the ambience mix, even though visual switching among cameras might suggest otherwise. Applying audio-follows-video to the ambience signal, he says, would quickly lead to listening fatigue and discomfort, because a rapid succession of audio perspectives triggers innate human reflexes of insecurity.

If the cameras do a proper job, the viewer will easily realize that the action is on one side of the field even though the audio information seems to suggest otherwise. This also explains why ball kick noises are always at the center (mono), irrespective of whether they occur on the left or right side of the field.

This might sound like a big compromise, but it is a good one, says Krückels - especially since the spill of background noises into the field-of-play noises is such that moving the those between left and right would cause serious imbalances to the ambience.

Are there different philosophies regarding how surround/immersive audio should be mixed? There do seem to be European and American preferences: Europeans pay more attention to a convincing crowd sound, while American productions often favor an 'in your face' aesthetic - placing players at the center of the audio image and using heavy compression ratios for 'ultra-realism' - complete with the audible artefacts of that compression.

Krückels himself subscribes to a compromise between these two—paying attention to details (FOP noises) while maintaining a truly immersive ambience where viewers have the impression of being at the venue. He does not want to be immersive at all costs, though. A silent crowd, he says, does not sound more animated when captured in 3D/immersive audio. For Mexican waves, on the other hand, he likes it when the sound travels from speaker to speaker.

Dynamics

Dynamics effects are extremely important in a surround/immersive audio scenario. In the broadcast world, no audio engineer can afford to do without them, says Krückels. A dynamic range of 30 to 40dB simply would not work. For key signals such as speech, music, and FOP noises, the human ear prefers to stay in a +7 to -10dB LUFS range. This is even more important when you consider that the level is usually much higher in an arena (110dBA, for instance) than the rendition consumed at home (maybe around 68dBA). That doesn't leave much headroom for the audio engineer anyway. Add to that that most people's preference to stay within a dynamic range of around 15dB, and the necessity of dynamics processing becomes obvious.

One solution that puts this principle to clever use, says Krückels, is Lawo's KICK software. According to him, it is the only solution that manages to keep the kick and spill noises at a constant level, thus avoiding sometimes brutal level jumps and artefacts.

Will it Take Off?

Most sound engineers are confident that 3D/immersive audio will establish itself much faster than 5.1, not least thanks to important side technologies like Virtual and Augmented Reality. Many people are already familiar with binaural listening and even head-tracking in VR, which is available on most gaming consoles.

Audio engineers can easily create binaural mixes that serve as immersive sound renditions—and most people will be hooked almost instantly and never want to return to a stereo mix. Krückels therefore believes that headphones will play an important part in establishing immersive audio.

However, he is not so fond of sound bars, which he considers a compromise that might just fall short of accurately rendering the added value.

In any case, the consumers' ready acceptance of all these technologies, are playing a major part in the success of immersive audio broadcasts - particularly in sports - and with lots of standards development and interactive functionality to come, it looks like Next Gen Audio is headed for success. It's important that every A1 is well prepared for it.



Felix Krückels.

Find Out More

For more information and access to white papers, case studies and essential guides please visit:

thebroadcastbridge.com

WP

WHITE PAPERS

EG

ESSENTIAL GUIDES



MEDIA

CS

CASE STUDIES

05/2019

Sponsored by Lawo, Genelec and Sennheiser



GENELEC®

