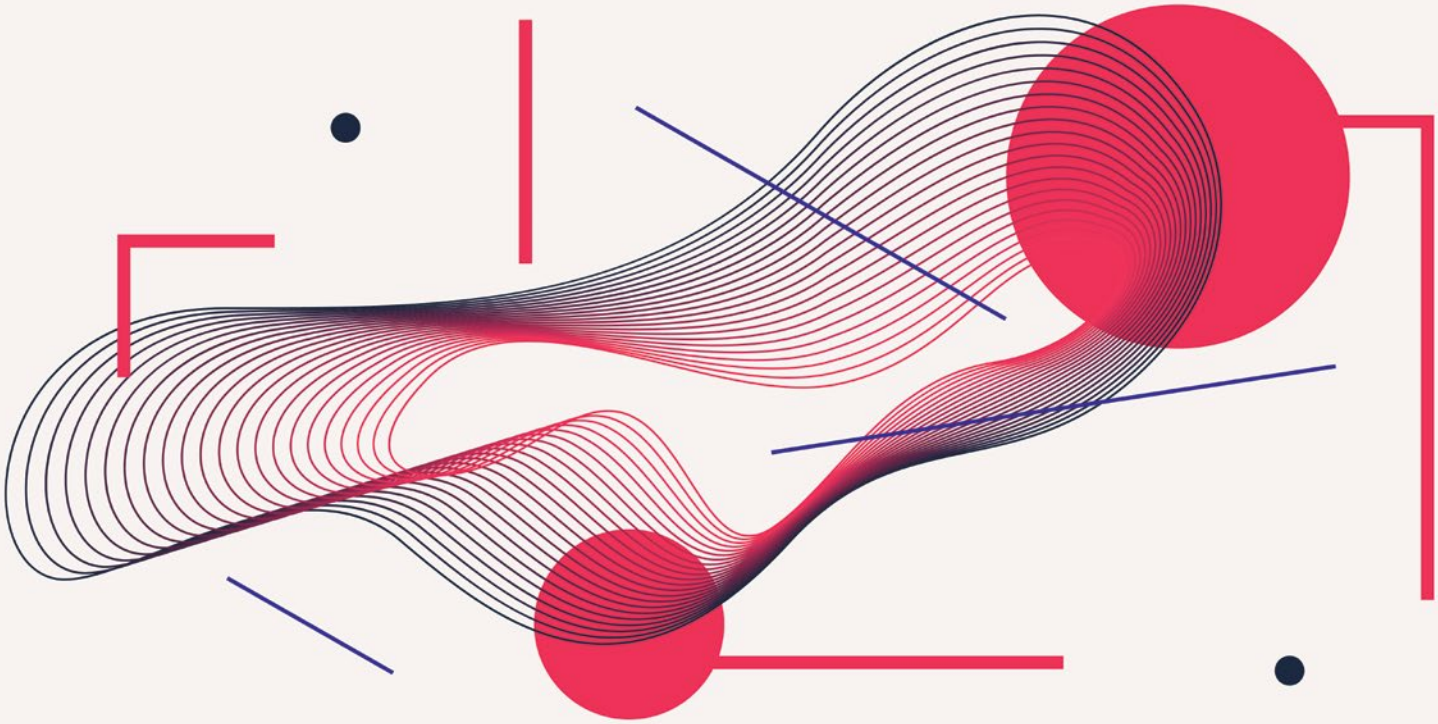
10 things you need to know about

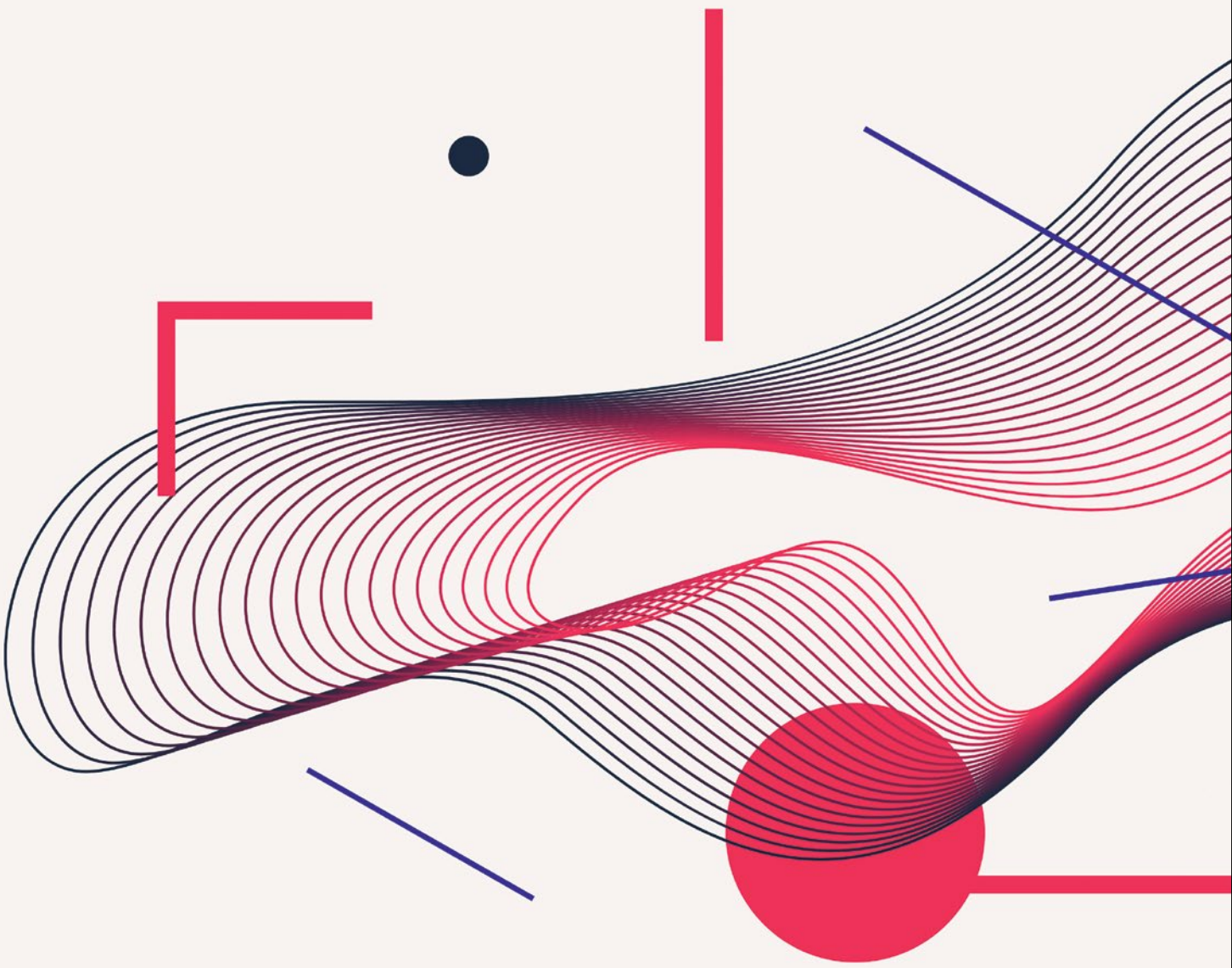
Next Generation Audio



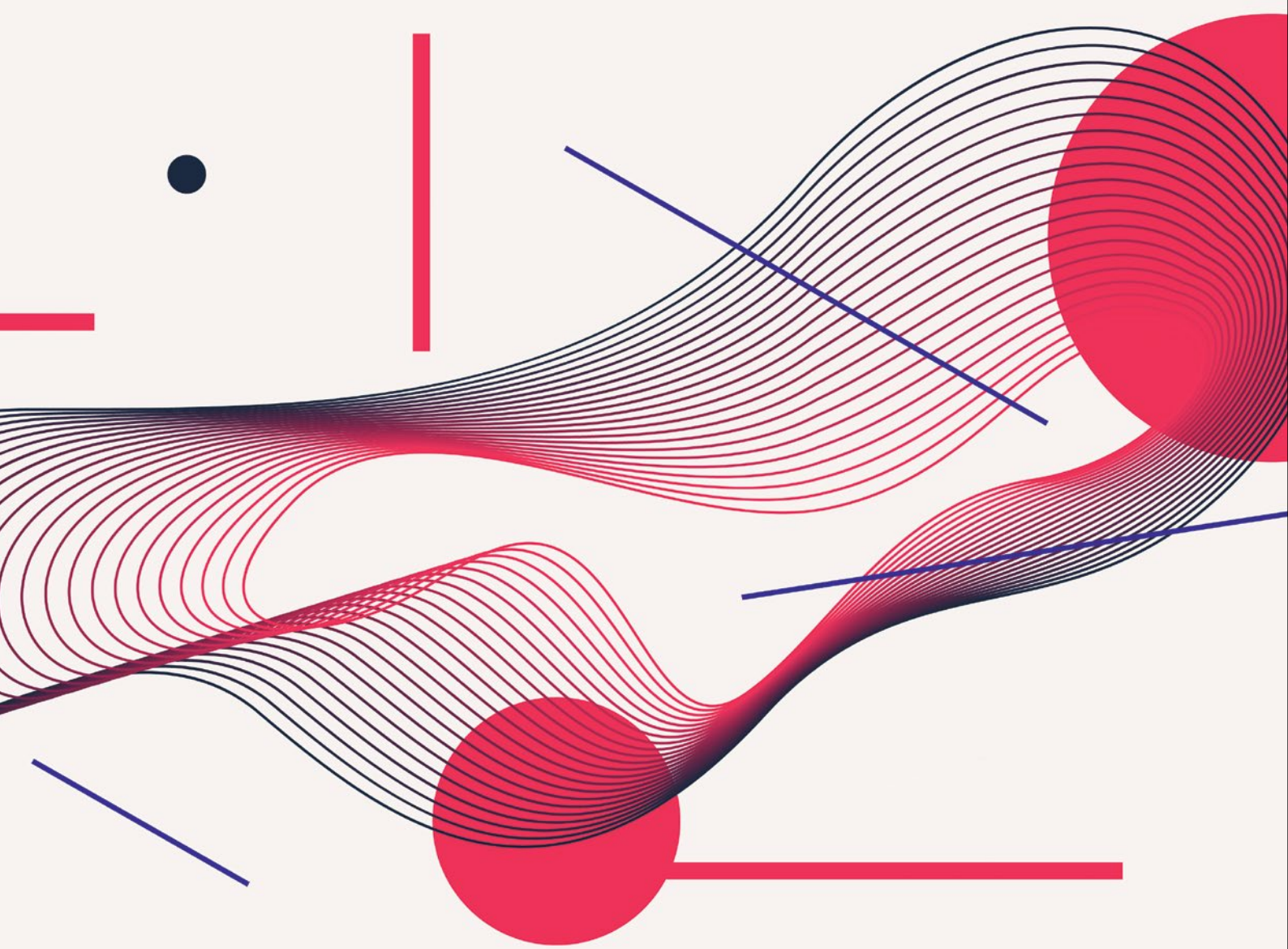
EBU
OPERATING EUROVISION AND EURORADIO

Forum for Advanced Media in Europe

Contents



Introduction



Next Generation Audio (NGA) offers very significant benefits to the audience, the content creator and the broadcaster. Whilst immersive audio is already valued, many in the industry believe that personalisation has the potential to be even more widely appreciated. The ability to adjust dialogue levels relative to other sounds will address one of the most frequent complaints received by broadcasters. NGA also offers potential long-term efficiencies in production and distribution, especially when reversioning is required.

In order to explain the significance of the differences between traditional and next generation audio we will start by briefly looking at how audio technology has developed to the present day.

It was in 1877 that Thomas Edison invented what is generally considered to be the world's first practical audio recording device in the Phonograph [1]. We then had 55 years to wait until 1931 when Alan Blumlein [2] added a second channel, thus creating stereo and a surprisingly immersive experience. We recommend you listen to one of Alan's early test recordings on YouTube, called Walking and Talking [3], as this clip demonstrates just how effective early stereo was. This may in part explain why there hasn't really been much of a need to replace stereo so far. Certainly the fidelity has been improved over the years but the format hasn't really changed much, to the extent that Alan's original patent suggests the addition of extra speaker channels to create surround sound, which is now commonplace in cinemas and on HDTV.
Change is in the air, however. We all know that modern forms of media are becoming increasingly sophisticated, and patterns of consumption are evolving, enabled by better internet connections and advances in Web based technology. It won't come as a surprise to learn that a new audio paradigm is emerging. It is this next generation of audio that this guide explores. The guide has been written by the members of the FAME [4] Audio subgroup to help those working in the broadcasting industry, or those who have an interest in media technology, to understand the potential of these latest developments in audio, which we think could be the first really fundamental change for 80 years or more. We hope to provide a background for managers, context for decision makers, ideas for producers, pointers for commissioners, a vocabulary for skilled operators and a starter for technical staff, whether they are working in television, radio, online,film or publishing.

1. What is the most important development in audio technology for decades?

We know that modern media forms are becoming more sophisticated. The experience offered is of a higher quality and it's increasingly immersive. There are more pixels than ever before in cameras and on screens. The whites are brighter, the blacks are darker, with the option of more frames per second, and yet the latest codecs are more bandwidth efficient. Everyone is talking about 4k, 8k, HDR, HFR and HEVC [H.265] video but what is the sound system that should accompany this superb visual experience?

Well, the next generation of audio technology is exactly that, Next Generation Audio or NGA. If you're now wondering if a better name couldn't be found for what should be the biggest change in audio technology for 80 years, we'd be inclined to agree with you. But this is the name that has been adopted and in its defence, it is exactly what it says on the tin. Unlike the developments in picture quality, NGA isn't primarily about "more" or "better", but instead offers different workflow and distribution options as well as enabling new, more flexible personalised user experiences.

NGA technology combines our current world of mono, stereo and surround sound (the channel-based audio formats) and takes it to the next level to create a powerful new audio toolkit. It does this by adding Object-Based Audio [28], which consists of audio content together with metadata to tell the consumer device how to handle the audio. The metadata also enable personalisation and control the way this is presented to the user. Immersive audio, or "surround sound with height" is also supported along with scene-based [5] audio using Ambisonics [6].These three technologies of channels, objects and scenes (of which there will be more later) are becoming the established partner for UHD in the latest generation of TVs. But they can also work with SDTV, HDTV and VR technologies or, as demonstrated by the Orpheus project [7], as an audio-only radio service.

NGA will benefit content creators and broadcasters by enabling one master copy or stream to provide the optimum experience for a range of platforms and user devices. This means it's not just extremely bandwidth efficient but it should also create the possibility of process efficiencies for broadcasters and producers by simplifying complex workflows. Multiple language versions can be handled simultaneously, potentially removing the need to create multiple different balances for each version. These capabilities combine to allow the content creator to tell the story better, to the benefit of the consumer. Consumers also gain the ability to personalise the audio and experience improved accessibility.

As we have seen in our introduction, the process of acquiring and distributing audio has not changed significantly since the 1930s. A number of sources (voices, effects, music, ambience) are mixed together and the resulting audio is distributed to the consumer. Innovation has been limited to adding more loudspeakers for replay, starting with one (mono) then two (stereo) and finally moving to various surround sound formats such as 5.1 [8]. Each piece of distributed audio was optimised for a specific number and layout of loudspeakers and the viewer only had control of the overall sound level.

This is important because NGA fundamentally changes the way audio is distributed and consumed. It offers broadcasters the opportunity to transfer some of the final mixing and processing into the consumer device, with the following results:

- The viewer could personalise their experience by choosing the language or version of commentary, or electing to hear additional audio such as a director's commentary or referee's mic.
- The viewer might control how prominent the dialogue and other narrative elements are compared with background music and effects. This is beneficial not just for those with hearing impairment but also for viewers in a noisy environment, or for those where the content is not in their native language, or who are not giving the TV their full attention.
- Audio description can be added and moved to a location away from the screen, making it easier for the person who needs it to understand the description, whilst also making it less distracting for others watching with them, who don't need it.
- Content is created in a format which is agnostic to the number of speakers used for replay; one piece of content should work equally well in stereo, surround sound, immersive with soundbars or height speakers [9]. It can even render to binaural if the consumer is using headphones.
- Truly immersive audio with sound positioned in three dimensions is possible in the home, even without speakers in the ceiling.
- All these options are provided under the control of the broadcaster, who decides what level of interaction to allow the viewer and how it is presented to them, which could be as simple presets or as a more sophisticated interactive experience.

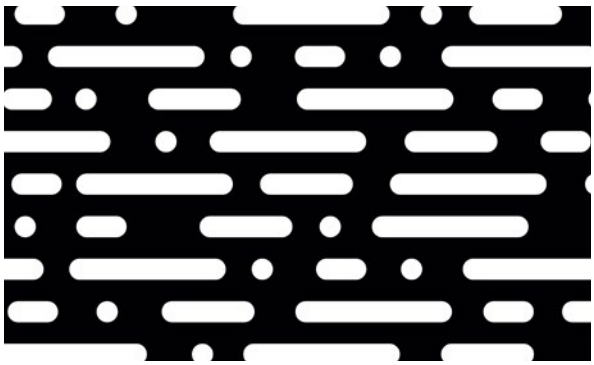
2. How does Next Generation Audio work?

Next Generation Audio is based on 3 different techniques: channel-based audio, object based audio and scene-based audio. These can be combined to improve the experience, to enable many more creative opportunities and to allow interaction and personalisation. We describe these techniques in more detail later but first, why do we need something so complex? Our current audio systems are designed around the channel-based approach. Here, the content is mixed by the broadcaster and the whole audience hears the same sound balance. We know the consumer replay environment is diversifying, for example TV sets are no longer boxes with one or two reasonably large loudspeakers facing the viewer. Instead, a TV may be only a few millimetres thick, with very small speakers facing backwards or down, and less expensive sets can offer poor dialogue clarity. At the other extreme ,high-end home cinema systems offer some viewers a better audio experience in the home than ever before. The diversification of the replay environment coupled with an ageing population combine to create a world in which one sound balance will no longer work for everyone. This change has manifested itself in an increasing number of complaints made to broadcasters about poor dialogue audibility on a range of productions. Current broadcast technologies not only suffer from a single mix having to serve a diverse range of listening situations, they also only support (at best) two dimensional surround sound. Whilst the traditional approach could be extended to deliver immersive audio this would not be bandwidth efficient. NGA solves all these problems using three techniques.

i) Channels

Even with our current channel-based approach we find it increasingly difficult to manage the number of formats and compatibility between them. In a progression from mono, through stereo, to surround sound and onwards to immersive audio, we go from a single audio channel to as many as 22.

NGA provides support for established channel-based approaches, produced from one mix and one bit stream. This ensures some degree of backwards compatibility with the existing catalogue of content, established workflows and current distribution platforms.



ii) Objects

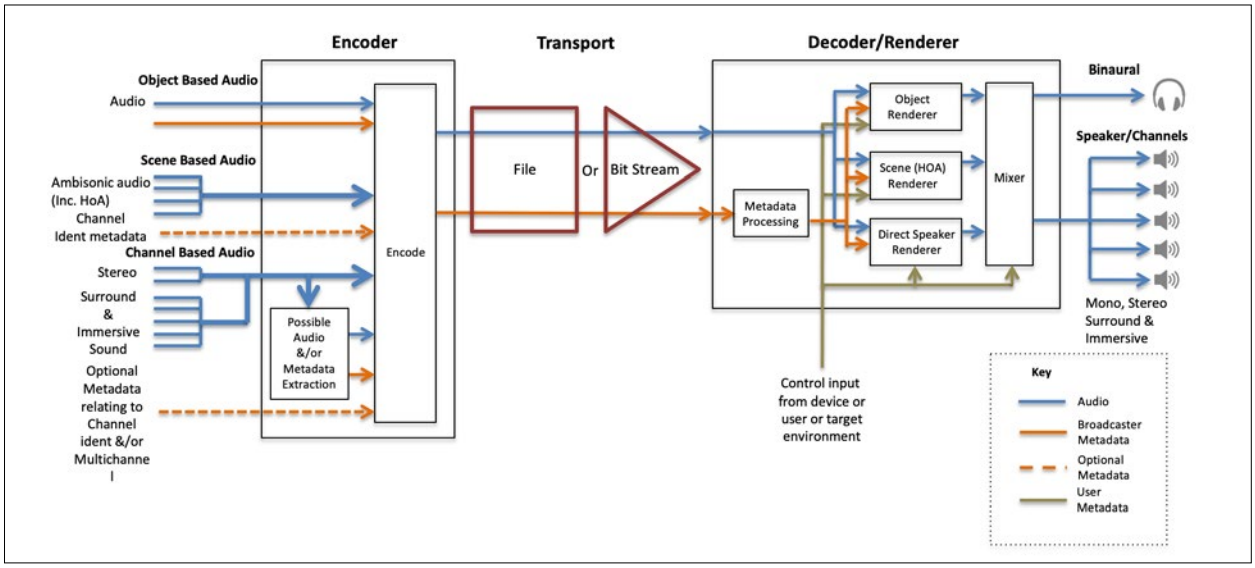
We are all familiar with the traditional audio production process described earlier, where we start with many audio sources (microphones, etc) then mix them down to a number of channels which represent the final speaker layout. For example 100 microphones on an orchestra might get mixed down to two audio channels, one for “left” and one for “right”. With object-based NGA, sources are not necessarily mixed in the usual way but grouped or isolated into audio feeds which constitute separate, logical audio objects. These could be individual voices or instruments, sound effects like a passing vehicle or a group of mics which make up a logical entity like the string section or a drum kit. The key thing is that they are kept separate and are distributed as individual objects. This doesn’t mean that every microphone feed has to be kept separate. We might, for example, mix the sound of a sports arena into a channel-based bed (as at present) and broadcast this alongside a “commentary” object containing all the speech. These objects are made up of not just the audio itself, but descriptive data describing the audio’s position in space, its level etc. The receiver replays the commentary object using the instructions contained in these data. The instructions can (at the broadcaster’s discretion) be modified by the user allowing them, for example, to alter the prominence of the commentary compared with the arena sound. It is here that we begin to see the power unlocked by the object-based approach.

© chanut iamnoy/istockphoto.com

ii) High Order Ambisonics

Ambisonics is another technique which can be used for distributing audio as part of the NGA tool kit although it is so far less important in broadcast than channels and objects. You may recognise the term as it was originally developed by Michael Gerzon, Professor Fellgett and Peter Craven in the 1970's as a method for capturing the detail of three-dimensional sound without the need for a large number of audio signals. It is only in recent years that the technique has become practical, with the emergence of more powerful software based electronics and driven by the popularity of virtual reality. High-Order Ambisonics suitability for virtual reality production and gaming is because of the ease with which an entire sound stage can be rotated in response to the viewer turning their head or moving in a virtual world. It also has the advantage of being inherently replay format agnostic. This means that rather than transmitting a signal for each loudspeaker, a group of carefully interrelated audio signals are used to describe the entire sound field. The accuracy achievable with this technique depends, amongst other factors, on the number of audio signals used. Basic "first order" ambisonics uses just four signals to describe the full scene, one which captures the total sound pressure and a further three difference signals for left/right, front/back and top/bottom at a single point in space. This creates a relatively small "sweet spot" in which good immersion and accurate localisation of sources is experienced. Higher orders of ambisonic sound add more signals to improve the precision of the image and reduce the sensitivity of the listening position at the expense of increased complexity and signal count. Third-order ambisonic sound uses 16 signals, NGA typically supports up to 6th order ambisonic sound. HOA is also thought of as scene-based Audio [10].

NGA Signal Path Overview



NEXT GENERATION EXPERIENCES ARE GENERALLY GROUPED INTO THREE CATEGORIES:

Immersive

The final presentation is "rendered" in the consumer device, to recreate as faithfully as possible the original creative intent given the number and location of loudspeakers available in the home. (Rendering is explained on p12.) This is a very flexible process that does not demand any adaptation of the programme created by the content producer. A single version of the programme is sufficient for distribution over multiple multiple infrastructures and can be assimilated and personalised on multiple platforms including binaural rendering for head-phone listening. Reuse of the programme on catch-up services and repurposing from the archive is also made possible through appropriate adjustment of the objects' metadata.

One of the greatest challenges for immersive audio in the home remains where to put all those loudspeakers required for a good immersive experience. A solution to this problem is now possible by combining NGA with the next generation of advanced soundbars or TVs. These allow viewers to experience a high-quality, room filling, immersive audio experience from a single device positioned under the TV and fed via a single HDMI connector. Many new UHDTVs include the capability to reproduce an immersive experience through their own speakers, without any additional devices at all.This means it is no longer necessary to fill the room with speakers and wires in order to fill the room with sound!

Importantly it is now also possible to create a truly immersive experience for headphone listeners using a technology known as binaural. We've all witnessed the popularity of high quality headphones and although we can't be sure sure if people will wear them whilst sitting in front of their large screen at home, we know they're ideal for TV on the move when binaural should make the small experience much more immersive.

© Pixsooz/istockphoto.com

Interactive

FInteractive mixes using objects can allow for personalisation; a sports fan could perhaps choose a partisan commentary supporting their team, together with stadium sounds from the end of the ground where their team's supporters are sitting. Alternatively viewers could choose to hear a microphone on the referee, whilst a film lover might wish to hear the director's commentary.

Giving control to the viewer means the content creator gives up some control over exactly what the viewer will hear. As previously noted consumers very seldom enjoy an ideal listening environment, so we don't have as much control over what they are hearing as we might like to think. Next Generation Audio allows the broadcaster to set limits on the control the viewer has, for example to prevent them from completely fading out the dialogue.

For the keen student of broadcast technology some of this will sound a little familiar, as the approach to delivering audio description on the UK's DTT platform (Freeview) uses what we would describe as an object based approach. Here the audio stream is supplemented by a dynamic data stream that describes how to mix the audio description with the standard programme sound when the viewer selects that option.

The FAME team is sure there is a huge untapped potential in this new level of granularity and control. Most of the use cases for NGA technology proposed so far have been developed by the engineering teams working on the technology but now the time has come to widen the creative input, the result is bound to be new, exciting and unexpected experiences.

© piranka/istockphoto.com

Intelligible

For the cinema NGA is mostly about the immersive sound experience, whilst for broadcasting it is the personalisation and accessibility aspects that we think will be the most important innovations. Broadcasters could offer audiences a choice of languages, commentary from different points of view, audio description and clear dialogue all via the control of objects.

Because the various objects are transmitted separately to the receiver, the viewer can not only select their preferred objects but also decide how they want the audio to be mixed. This is part of what is called rendering [11] in the NGA world and is explored later. If the viewer has a hearing disability they might decide they want the dialogue to be louder (and everything else quieter), if they have a visual impairment they might wish to turn up an audio description object.

The object-based approach offers some other intriguing possibilities because the audio description is easier to manage in its own right. A choice of different languages can also be offered, and the chosen version could be panned to a position which best suits the listening position of the viewer, whilst protecting the balance in the rest of the room.

© NicoElNino/istockphoto.com

© nadla/istockphoto.com

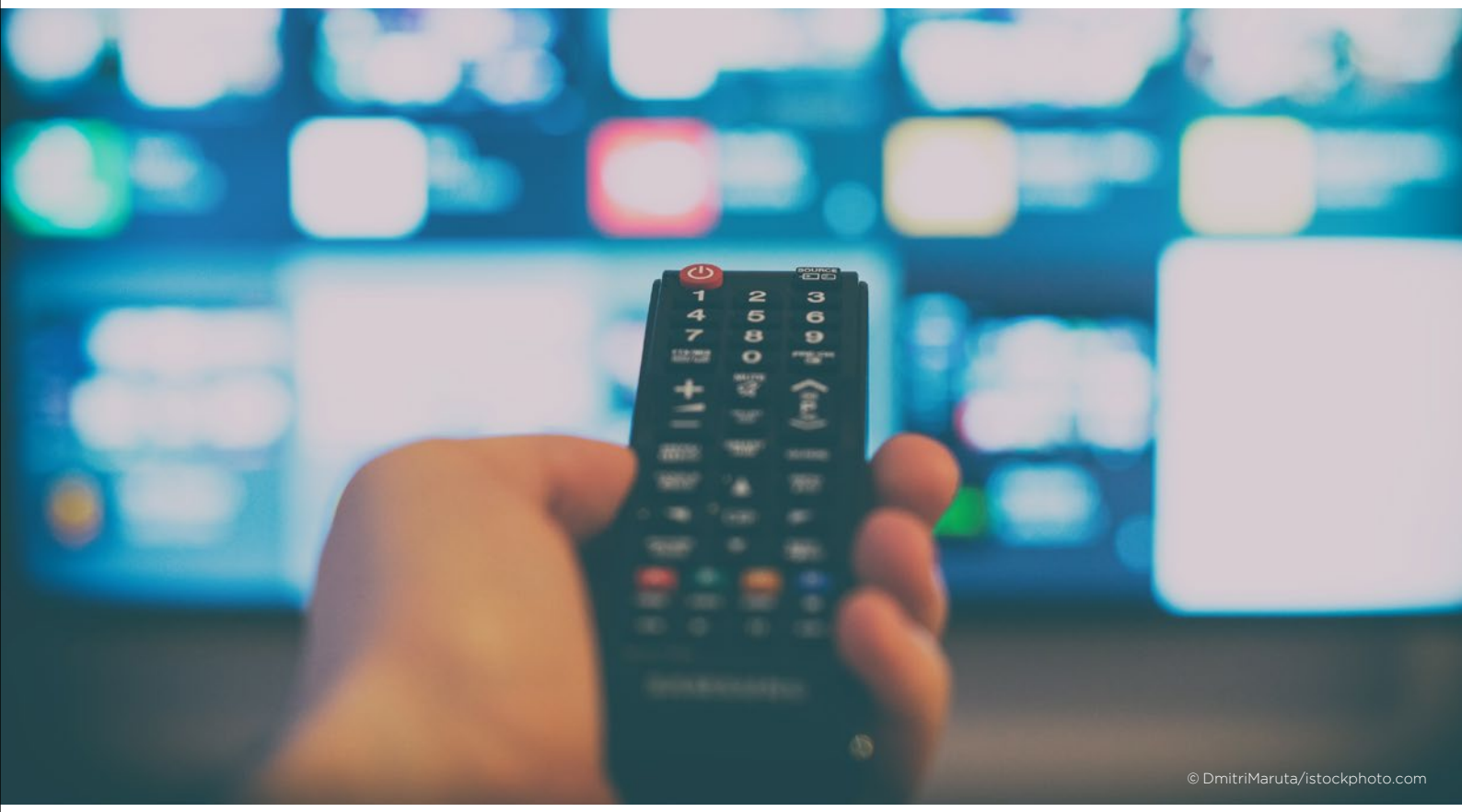
Metadata

You probably spotted the topic of metadata appearing earlier in this section and to be honest it had to happen sooner or later. Yes, metadata has finally become a core component of audio. This shouldn't come as surprise as extracting the full benefit of any technology on the web requires metadata. Neither is it entirely a new concept for audio as anyone who has worked with surround sound for TV will know that technical metadata is necessary to make the system work in a band-width efficient and controllable manner. The experience of implementing audio for HDTV has shown the benefits of including metadata in our system designs but it has also made us aware of the challenges of end to end metadata across broadcasting infrastructures.

The Renderer

The renderer is the heart of the NGA system. It's the component where the audio and its associated metadata are combined to produce the signal that will feed the loudspeakers of your reproduction environment, e.g. your phone, TV, home cinema etc. It is at this point that information about viewer choices and the configuration of their equipment are added to the mix, so that the experience can be optimised. Thus if you only have two speakers you can listen in stereo. Or if those two speakers are in fact headphones you can produce an immersive binaural effect. If you have difficulty hearing what's being said, you can turn up the dialogue or turn down the effects. If you want to pick that authentic international dialogue option, then you can. If your surround sound system is a bit asymmetric because of the living room door, the system can compensate. Each of the emission codecs specified by DVB [12] and discussed in more detail in section 6 has its own renderer and these do not work with content encoded using another technology. For this reason, content creators may choose to employ an open standard, such as the Audio Definition Model (ADM) during production. Content can then be converted using one or more NGA emission codecs at the point of distribution to the consumer. It is also quite possible to include multiple renderers in one receiver or decoder, although this has cost implications of course. So part of the challenge of getting an NGA system to work well is not only in selecting a distribution technology. We must also equip all stages of the production and Quality Control workflows with the ability to create, monitor, pass and store technical metadata. We must choose the right rendering technology to maintain consistency across the broadcasting process. Fortunately our colleagues at the EBU [13] have worked on this, and there is now an internationally-recognised renderer for use in production, QC, subjective evaluation, recommended by the ITU-R.[14]

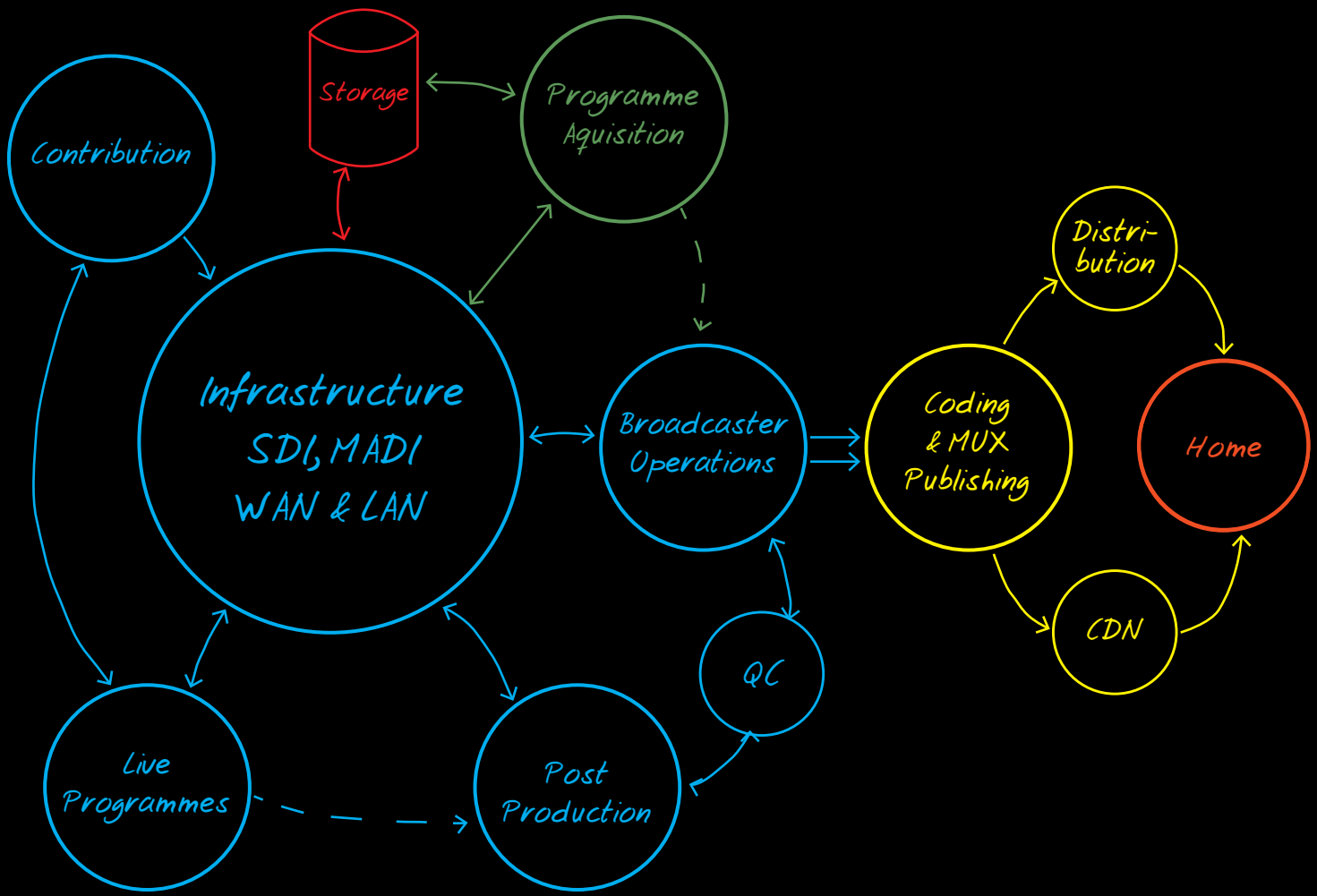
3. How does this relate to UHDTV?



Next Generation Audio is not dependent on UHDTV as it is possible to have one without the other. However in most markets the introduction of UHDTV has been used to introduce NGA. Likewise consumer products supporting NGA generally also support UHDTV since introducing the two technologies together is arguably a more efficient investment for the content creator, broadcaster and consumer. In fact adopting NGA to accompany the introduction of UHDTV may well make sense even if the advanced features are not initially exploited, because its better codecs can improve quality, save bandwidth or enable delivery over IP to hybrid or connected TVs and mobile devices. As we've discussed, NGA builds on established techniques and technology as well as making use of the latest codecs to deliver immersive audio and interactivity using a lower bandwidth than is possible with just our current channel-based approach. There are already a few broadcasters who are covering sport with immersive sound, a choice of three languages for commentary, plus audio description, all within a bandwidth of 380 kb/s.

The DVB project provides the tools to incorporate NGA as UHD technology is implemented and the EBU explores how this might be done in their report R151 [15] which we will visit again later.

4. How does this change my workflow?



It may surprise you to know that in many cases you won't have to make big changes to what you're currently doing. In fact the introduction of NGA should allow you to realise the true value of the effort you're already putting into content creation. In some cases you may find that you're creating something like audio objects already. The stems in your current mix, or your clean international audio feeds, could become the audio objects you deliver to the audience. What you'll need to do however is keep these components clean (avoiding spill), discrete and coupled to their metadata. In fact it's the addition of metadata, much of which is dynamic [16], that will be the greatest challenge in getting NGA to work effectively. Metadata will have to persist from end to end, that is from production to the home, for both live and pre-recorded content. This is part of a much broader challenge which broadcasters must meet if their offering is to be fit for the internet age, when experiences will need data to be discoverable, personalisable or to just work.

So why should I mix with audio objects?

We are already familiar with surround sound based mixing and the addition of height with speakers overhead an immersive soundbar or a TV with immersive audio speakers built in. If you choose to increase the immersion of your audio, it is only a small step up in complexity. In many scenarios, four additional microphones in the right place will allow you to add height to the presentation. There are a number of microphones available which will capture three-dimensional immersive audio from a single unit and whilst some are large and expensive, others like the familiar Soundfield microphone can also be used to create an ambisonic presentation or decoded to provide a "height" component. Treating some of your sources or stems as audio objects will then give you a range of benefits when mixing. You can defer fixing elements of the mix that are best made in the receiver, until the receiver. Such things as where to position a voice which might be best behind you in 5.1 surround for example, or above you in full immersive but to the left in stereo. Finally, NGA allows you to produce a single asset which can be used on any device and can contain several languages, reducing the need to create multiple versions of a content item.

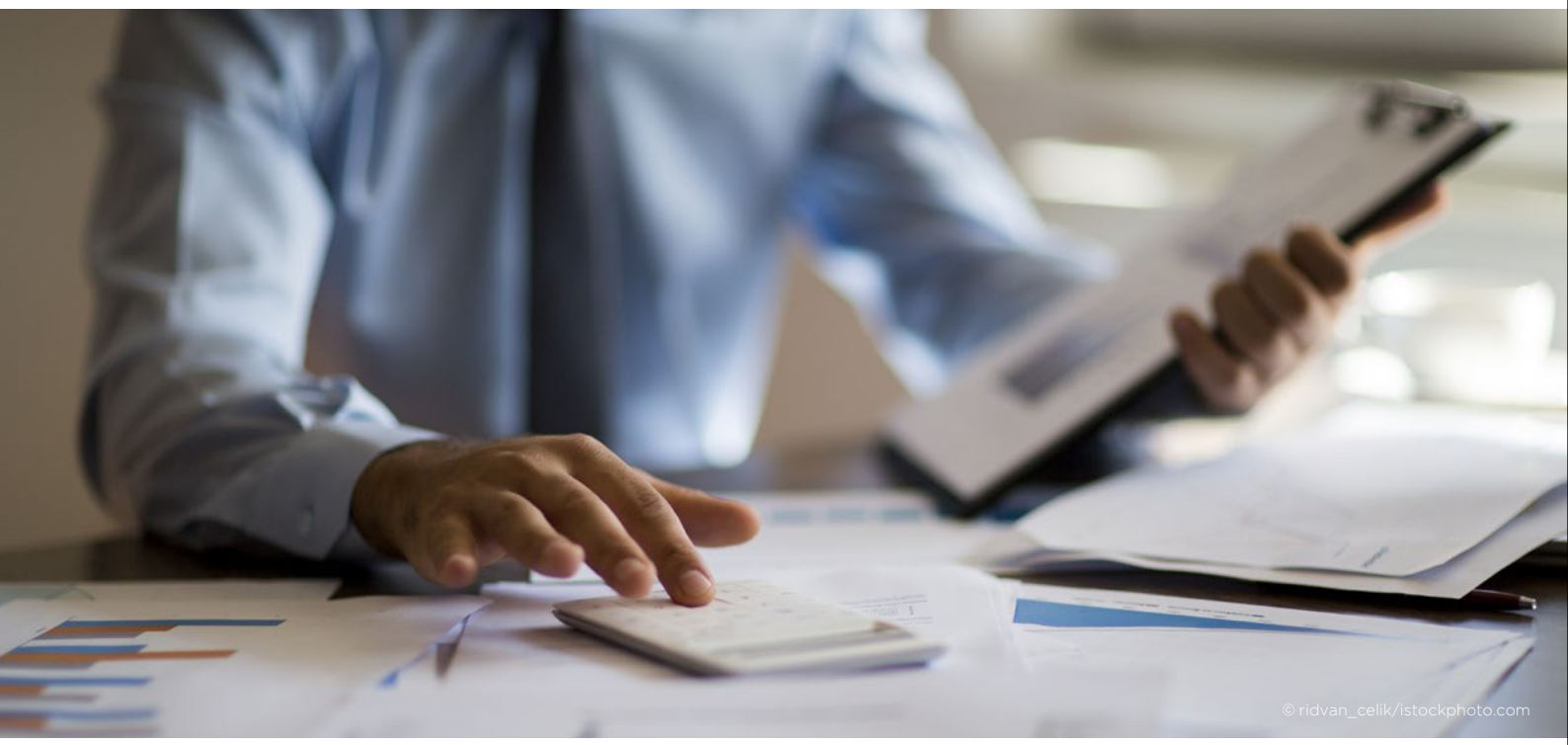
Do I need new technology?

To use NGA to its fullest extent you will need new technology. You'll probably need new encoders and decoders, more loudspeakers if you wish to add immersion, tools to generate, transport, monitor and decode audio associated data. There is much you can do to adapt your current infrastructure by using adaptors to transport objects over SDI connections or using plug-ins with Digital Audio Workstations.

Can I do live programmes using NGA?

Yes, NGA can be used for both live and post-production workflows. Most of the NGA broadcast applications have been for live sport, where a greater sense of "being there" can be created, and extra commentary feeds can be added for different fan groups. Development has also focused on enhancing feature films and is thus well suited to large scale, high value, pre-produced, features. This is not to say that NGA will currently scale up easily across everything broadcasters do. There is still work to do as it hasn't yet been established how we might make news programmes using NGA, or manage fully automated NGA playout, or create large NGA archives. But many of these challenges are being worked on, for example the new file-based interchange format, the IMF [17] (Interoperable Master Format), is being enhanced to include NGA.

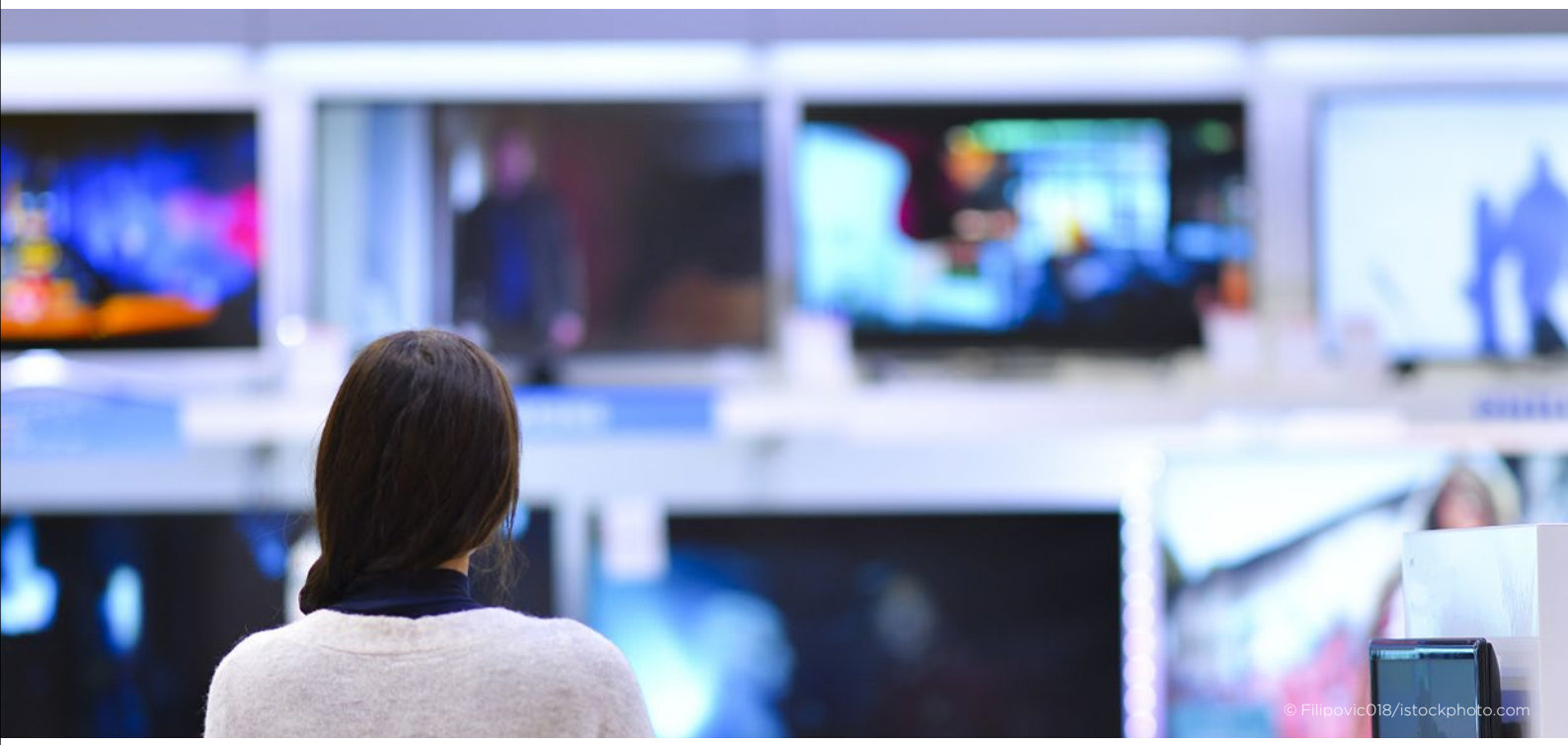
5. So how much is this going to cost?



If we're completely honest we don't know. What has been shown is that you can do entry level object based broadcasts at very little cost, just by adapting current workflows and configurations and adding some well understood encoding technology. The BBC demonstrated a system providing audience dialogue control of their radio commentary from the Wimbledon tennis championships as far back as 2011. Using early OBA technology, the "Netmix" experiment was streamed over the internet and required the a of a browser plugin, but was remarkably cost effective to produce and free to users of the right browser. For audiences this continues to be true, with more and more domestic UHDTV receivers having at least one NGA technology installed at source by the manufacturer. A number of broadcasters are also using immersive audio as part of their premium offering on a regular basis, although it is currently thought that to use NGA across the full scale of a broadcaster's operations would be challenging, mostly because not all the technology or workflows are mature enough. Areas of development include exchange and delivery standards, Automatic Quality Control, and archiving.

As we are compiling this report we are not aware of any products for TV playout with a built in NGA capability, but we know that some DAWs offer the technologies as a plug-ins and that new tools are being developed and brought to market all the time. There is also the opportunity for broadcasters to introduce efficiency in their own internal processes and workflows using NGA techniques. It should be possible to reduce the complexity of handling multiple languages and clean international feeds using NGA. Better ways to deliver access services are already possible and streamlined workflows that simultaneously support stereo and immersive presentations have been demonstrated. Perhaps these gaps in the marketplace are best viewed as a commercial opportunity for suppliers of technology, equipment and services because as broadcasters move to realise the benefits NGA has to offer, then the demand for a friction free, cost effective, end to end set of solutions will only grow.

6. Is there any working technology?



This section is quite acronym heavy, and for that we apologise, but to help we've included a glossary at the end of the document. Firstly there are currently 4 major next generation audio standards:

Dolby AC-4

Many of you will be familiar with Dolby's AC-3 surround sound technology for broadcasting. AC-4 brings Dolby's object-based audio technology to the TV, which includes the Dolby Atmos immersive experience as well as personalisation possibilities. Dolby also have a range of associated technologies for producing immersive broadcast experiences, including Dolby Digital Plus (also known as E-AC-3 with Joint Object Coding or JOC). AC-4 is standardised by ETSI, and included in the ATSC 3.0, DVB and HbbTV [19] distribution specifications

DTS:X

DTS:X provides immersive object-based audio technology. The standard utilises the DTS-UHD codecs and is also backwards compatible with DTS-HD Master Audio technology developed by DTS which is part of the Xperi group. This technology is standardised by ETSI [20] and has been added to the DVB specification.

MPEG-H Audio

With core technologies developed by Fraunhofer IIS, including their AAC codec family and enhanced with technologies like HOA, MPEG-H Audio is standardised by ISO/IEC [21] and included in ATSC3, DVB and HbbTV distribution specifications.

ADM[22]

The Audio Definition Model is the open standard for the professional production and interchange of NGA files and streams. A file based version of ADM is standardised by the ITU [23] and also described in an EBU Technical Recommendation and implemented in the ITU ADM Render [14]. A serial version of the ADM standard (S-ADM [24]) was published by the ITU in July 2019.

You'll have also noticed that there are three main broadcast specifications, and the good news is that NGA technology has been included in the ATSC 3.0 specification since 2015 and the DVB broadcast specification since 2017.

HbbTV version 2.0.2, the specification for Over-the-Top (OTT) TV services, now supports NGA decoding using profiles specified by DVB. The missing link is the definition of a unified user interface to be implemented via the HbbTV NGA API to ease the control of NGA presets or options for users.

In 2020, 80% of TVs sold in Europe are capable of decoding at least one NGA standard.

Which companies are making NGA technology?

The different technologies available have come from different backgrounds, with Atmos and DTS:X starting in the cinema, whilst AC-4 and MPEG-H Audio were developed for broadcasting and streaming. Availability of products to support these technologies reflects where they started from. Large numbers of TVs are shipping in Europe with AC-4 from a wide range of manufacturers. MPEG-H Audio is incorporated in a large number of TVs for the Asian market. It is important to note that support for a given codec in a consumer product does not automatically mean that all NGA features are supported or that a user interface is provided for personalisation. Production tools also reflect the origins of each technology. For MPEG-H Audio there is a choice of broadcast-specific authoring hardware and software from a number of manufacturers and VST plug-ins for DAWs along with native support in video editing software. Dolby Atmos and DTS:X are widely supported in workstations for video and audio. It is worth noting that future production may be based on ADM and the FAME audio group believe support for this format will, in the long term, be more important than support of any specific native NGA codec within production tools. A range of authoring and monitoring solutions are also now available from a number of technology companies. At the time of writing it is fair to say however that many manufacturers of broadcast technology, such as playout systems, do not yet natively support any of the NGA formats and support in Digital Audio Workstations is incomplete.

It is likely that ADM will be adopted by many broadcasters as the production standard for NGA metadata. Under this scenario, content will be created in ADM and converted to one or more of the other formats for distribution and a pan-broadcaster EBU team are currently considering the best approach for a public service broadcaster architecture along these lines [25]. The SMPTE and DPP are looking at the inclusion of ADM in the emerging IMF interchange specifications and the EBU have reported on its use in general for broadcast application [26]. It is of course possible to deliver personalised audio without use of any of these NGA codecs. DASH, HLS and other IP based technologies make it possible to create content and distribute it to custom applications in the consumer device, or use HTML5 to create personalised experiences in the browser. FAME Audio Group believes such solutions would be likely to be replaced by NGA standards in time.

7. When will this become a reality for broadcast?

© FAME

It already is! In fact there are a growing number of broadcast trials and specification adoptions. In Europe, for example, BT Sport, Sky, Orange and Canal+ use E-AC-3 + JOC to add height to some content delivered over IPTV or satellite. Whilst this doesn't offer personalisation, it demonstrates that these companies see value in services offering an enhanced audio experience. At the time of writing in 2020, AC-4 has been adopted for the receiver specification in Italy, Poland and the NorDig group in Europe, using the DVB specification, and the first broadcasts are now live. AC-4 broadcasts are also live in the USA as the ATSC 3.0 rollout gathers pace. In South Korea the ATSC 3.0 specification is also used but this time with the MPEG-H Audio format. Continuous services were launched for the 2017 Winter Olympic games and a Chinese TV operator has also just selected MPEG-H Audio as the NGA codec for their UHDTV service. As we are writing we are aware of other national platforms and specification bodies that are considering including NGA technology. The limited availability of terrestrial TV broadcasting bandwidth may mean UHDTV is implemented first on IP platforms by PSBs, and NGA can be used in this way too. To help European broadcasters increase their understanding of UHDTV , the EBU has conducted a number of trials. These have included the European Athletics Championship in Berlin in the summer of 2018. This trial involved HFR & HDR 4k TV including NGA using fully immersive audio with both commentary and audio description available in two different languages, distributed in both AC-4 and MPEG-H Audio. In conclusion we see all of this creating a demand for both production tools and consumer products.

8. How do I listen to Next Generation Audio in the UK?

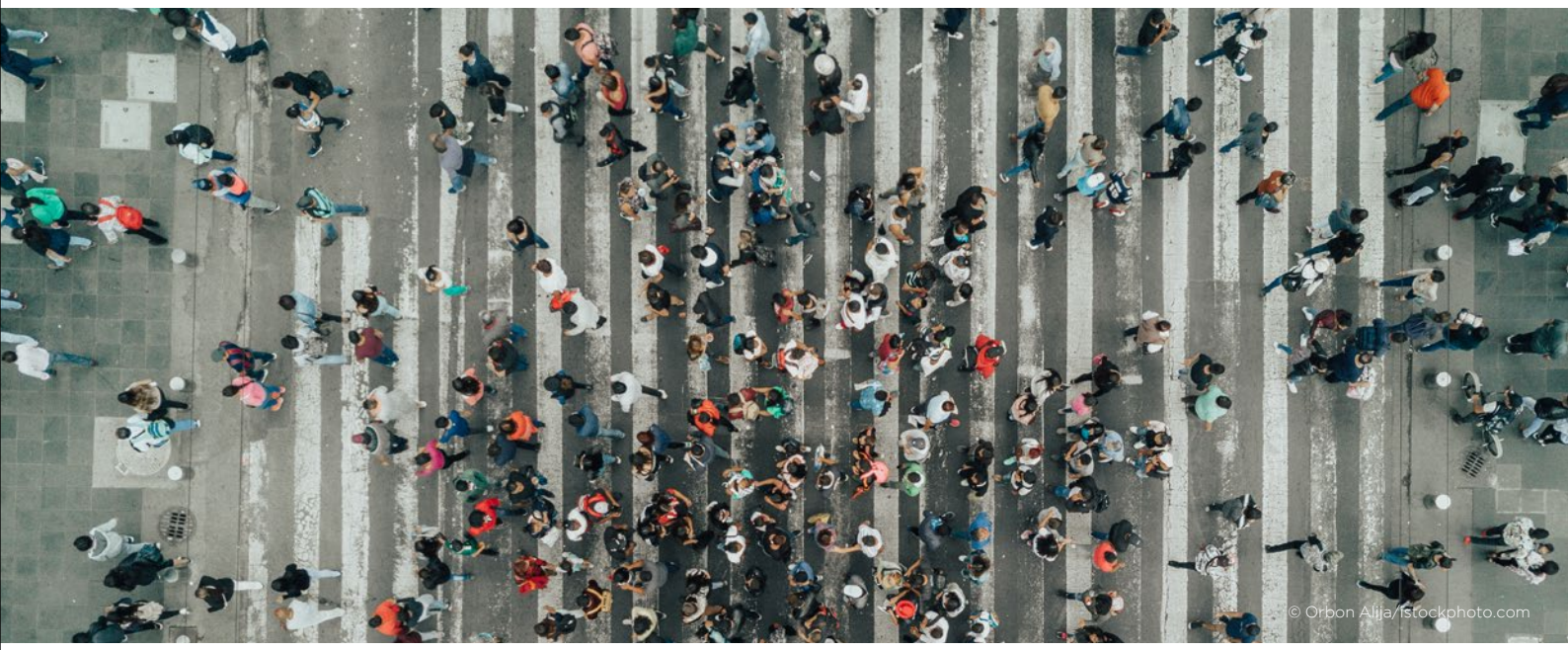
Illustration of an immersive NGA Sound field using a sound bar
© Image courtesy of Dolby

In the UK, Sky and BT Sport along with many streaming services including Netflix and Amazon Prime produce and distribute some UHD content with immersive audio, adding height channels to the conventional surround sound broadcast, although full NGA is not currently offered. It is of course possible to hear immersive audio in high-end cinemas and home cinema systems with height channels. However, without travelling to South Korea, the best way to experience personalised NGA is to arrange a demonstration with one of the technology companies providing solutions.

Doesn't this just mean I'll have to buy lots more speakers?

When you are using the immersive properties of NGA, then for professional production facilities you're probably going to need more loudspeakers in order to maintain high quality, especially for height channels. However in the home additional speakers are not essential; the consumer device should do a reasonable job of rendering NGA to the number an location of speakers available. In fact immersive audio can be replayed using a three-dimensional soundbar positioned under the TV and fed with a single HDMI lead, or via upward pointing speakers added to an existing surround sound system. TVs are now available with immersive speakers built in, too. It's also now possible to create a compelling immersive audio experience on headphones using binaural technology, and this can also be particularly effective on mobile platforms. For broadcasting, as opposed to cinema, we think that personalisation, especially dialogue enhancement, will be really important and this doesn't require any extra speakers at all. Building on what we discussed in section 5, by early 2020 at least 80% of UHTVs sold in Europe had at least one NGA technology inside, so should you launch a NGA service then you'll have an audience equipped to listen. However, Having the codec in the TV and having a user interface allowing personalisation are different matters, but several major manufacturers starting to introduce support for the DVB-specified personalisation control during 2020 in light of industry demand.

9. Who's going to benefit?



The big winner from this significant experience upgrade is the audience. Whilst immersive audio is already valued, many in the industry believe that personalisation can be an even more broadly appreciated feature of NGA. The ability to adjust dialogue levels relative to other sounds will address one of the most frequent complaints received by broadcasters. Presenting the best possible experience on any device will also be a key benefit, driven by an increase in watching on phones and tablets.

The way we hear is very different from the way we see. Human hearing has an amazing dynamic range, covers a large frequency range, in a continual 360°soundscape. So far what we can deliver falls short of what our senses are capable of experiencing. As we discussed earlier, stereo has been with us since the 1930s but has only been broadcast on TV at scale since the 1980s. Thus NGA is the first significant step in meeting our innate listening capabilities in over 30 years!

Not only does NGA improve the quality of the experience but perhaps more importantly it allows our established production processes to deliver content fit for an audience that is used to more interaction and personalisation offered by virtual and augmented reality (VR & AR). The accessibility features allow these benefits to be enjoyed by a wide range of viewers including some who are currently under served.

The benefits of NGA should go beyond those for the audience. We also predict new creative opportunities during production, as well as the possibility of workflow efficiencies.

- It will be possible to publish a single asset that will deliver the best possible audience experience in a wide variety of environments, from high-end home cinemas to mobile consumption on earphones.

- Enabling the audience to control the dialogue when they need to, will reduce a frequent source of complaints and enable greater creative freedom when mixing.

- It will be possible to offer alternative audio, different languages or alternative commentaries, adding cost-effective value to a TV broadcast. In some cases these options could be presented to the viewer as relevant pre-sets (e.g. a simple "clearer dialogue" button) or it may be appropriate to provide a more comprehensive set of user controls for more sophisticated experiences. Either way it is important to support the work already underway to develop and improve the user interface that will allow audiences to unlock NGA's full potential.

- Content created using objects can be easier to repurpose because key audio elements, such as dialogue, can be kept separate from the rest of the mix whilst still maintaining the correct levels.

You'll also retain creative control of your programme because NGA allows you to choose what options are made available and fixes the degree to which features can be adjusted.

10. NGA for advanced users

The use of enterprise architecture principles are becoming increasingly common in large organisations, and broadcasters are no exception. These principles provide a framework for choosing and configuring technology, standards, layers and interfaces in a way that will meet your business strategy. So whilst it isn't a prerequisite to adopt an architectural approach to planning your technology we think this technique gives us some simple principles for anyone to consider if, when and how they might use NGA:

- We'd suggest you think long term when it comes to deploying NGA. As we saw in our introduction, audio technologies have a history of a long life once they become embedded, so it's worth reflecting on where you think you and your audiences might be in 10 years or even longer. These deliberations should consider the effects of the ownership and the licensing of intellectual property as well as the life cycle of the technology itself. When launching a new NGA service it is always worth considering the technology currently in the shops and already in people's homes, however this short term factor needs to be balanced with your long term vision as we know that the consumer electronics life cycle is reducing whilst the rate of innovation increases.

- Be clear about the benefits you want to deliver, both to your organisation and to your audiences, and resist the temptation to be influenced by impressive demonstrations. Remember that vendors will optimise their demonstration content to flatter their technology so always use your people, workflows and material to test the solutions first.

- Consult your technical teams to work out how you can test the technology in order to be sure it performs as you require. Remember that some of this technology is new and we'd expect most technology providers to welcome the chance to work with users to iron out the wrinkles. The EBU and a group of leading commercial technology providers formed a close collaboration in the summer of 2018 at the European Athletics Championships in Berlin in order to undertake realistic tests of a full range of UHD technologies including several different versions of NGA. The EBU report on NGA makes interesting reading and is available to broadcast members. It's also worth watching what the DTG in the UK and its partners in the German Digital TV platform do in testing at their biannual plugfests for UHDTV. We hope to see NGA technologies becoming a regular feature of these events.

- It is to be expected any organisation adopting NGA will start small and build up. As with the launch of HD you cannot expect all content to be made with NGA from day one, neither is it necessary. In fact full scale NGA production for a broadcaster probably isn't possible yet. The FAME audio group has been tracking the development of NGA technology for the last 4 years or so using a "heat map" and although some workflows are now becoming proven there are parts of our business that don't have NGA ready technology yet. For example, we haven't seen an NGA-equipped asset management system so far. That should be a short-term problem as we see signs that the gaps are becoming filled in over time. It also supports the idea of a long term architectural approach. Adding NGA to your workflows should start to future-proof your content, increasing its quality and the value of your content library and archive, in the same way that many people are already shooting 4k or better video.

- Buying a technology solution from one supplier can lead to better integration and support, and most early NGA deployments have adopted a vertically integrated approach using one technology type from microphone, through production, via distribution, to the home.

- An open set of standards is however now available, to broadcasters to adopt a far more interoperable approach for production, programme exchange, playout archive etc., only converting to the branded distribution standards when appropriate. This open standards approach, based on the Audio Definition Model [ADM] [22], will increase interoperability for the broadcaster, with the result of reduced infrastructure costs in the longer term. It will also lead to increased flexibility in distribution, by allowing the optimum technology to be used for the platform, programme offerings, receiver base and audience. Once again the EBU has led the coordination of a group of European broadcasters to produce a technical report that explores these issues and opportunities. The architectural thinking they have developed is explored in recommendation R151 [15].

- So far we've looked at how flexible NGA is, and how broad a range of use cases and workflows it can support. In fact it's all a bit daunting. To help break the implementation down into more manageable chunks the team working on ADM have come up with the ideal of NGA "profiles". These complement the architectural approach of designing to support the broadcasters business strategy. The first profile definition has been published for production by the EBU in T3392 [27] and helps to define which components of NGA are needed for everyday production. It also starts to help us understand and define the hand-off points in programme commissioning, delivery and quality check. These are explored in Technical Report TR045[25].

Key Insights

Hopefully this guide has been an honest description of the current state of NGA. It has not tried to describe a bed of roses when one doesn't yet exist. There are some challenges to address as some parts of the system are still immature. This is a technology that should be with us for the long term, so it's important to start considering how you're going to make use of NGA in the future. There are a number of reasons for this:

- NGA isn't fully mature; it's important for broadcasters to get involved, so that as it evolves it does so in a way that delivers benefits for our industry as well as the film industry, VR and AR applications and advertising etc.
- If the media experience we know as "TV" has a future as a natural inhabitant of the internet, it will need NGA in order to adapt to new content types and maintain the quality of the audience experiences across devices and platforms. People increasingly expect content to deliver the best experience on any device, and they expect to personalise that experience.
- In a world where the processes and products of modern media are becoming more complex, this is a technology that can help streamline workflows and manage costs. There will be a need for investment in training, equipment and new workflows but in return, there will be efficiencies in revesrioning for different markets and platforms. Many of the workflow changes will, regardless of NGA, be required in order to deliver end-to-end metadata.
- NGA address some real audience needs, especially those concerned with the clarity and versions of dialogue.
- NGA has huge creative possibilities. It will be a significant factor in making better programmes and telling stories in new and exciting ways.

Using NGA does not mean implementing all of the features at once, it is possible to use the technology for stereo content. This still delivers production and reversioning benefits along with a more efficient emission codec. Features such as personalisation and immersive audio can then be introduced as and when required.

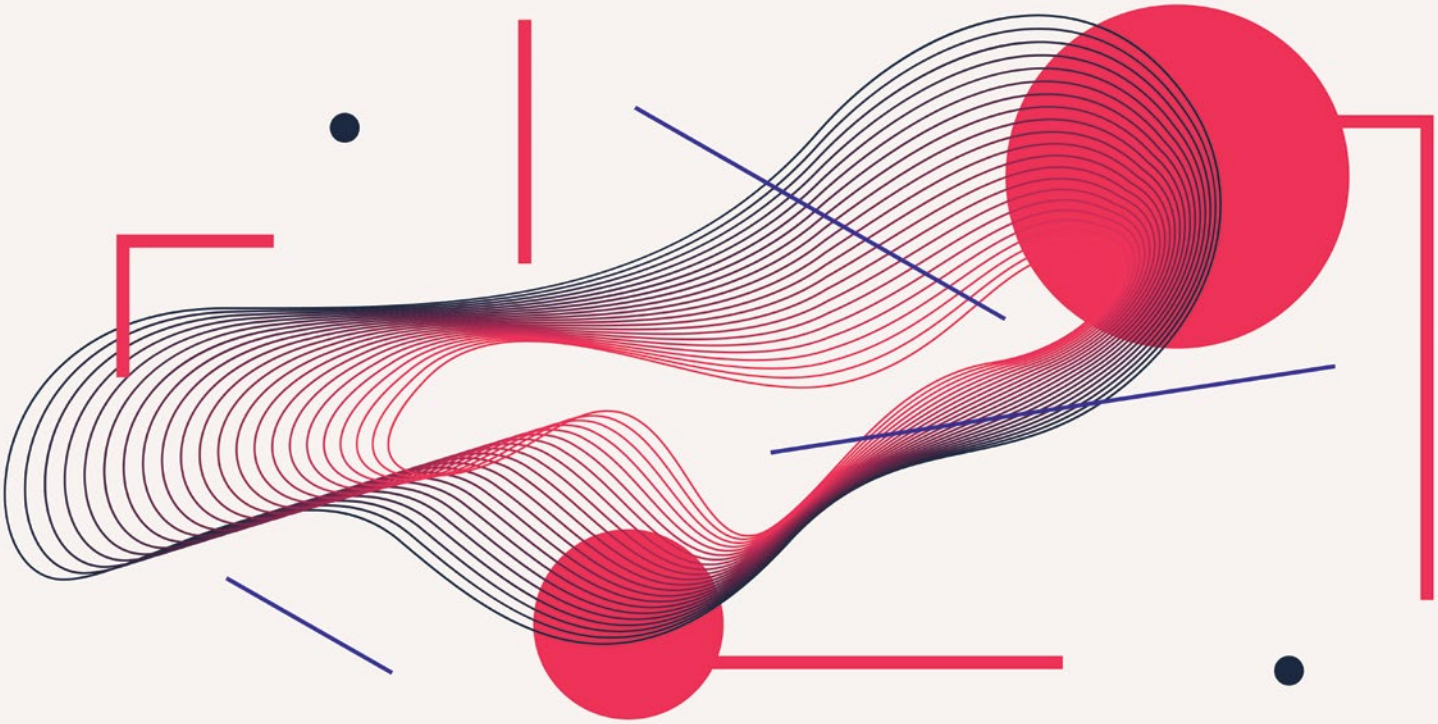
References

Ref.	Term	Link	
1	Phonograph	https://en.wikipedia.org/wiki/phonograph	Thomas Edison invented the Phonograph in 1877 and patented it the following year in 1878. "Mary had a little lamb" etc.
2	Alan Blumlein	https://en.wikipedia.org/wiki/Alan_Blumlein	Alan's patent for stereo (or Binaural as he called it) was accepted in 1933 and one of 128 he was granted during his short life. His work was also fundamental to the development of electronic television and the ground scanning H2S radar system. He's a bit of an engineering hero for the FAME Audio team.
3	Walking and Talking	https://youtu.be/rqaMiDqE6QQ	Recorded at the Abbey Road studios, the Walking and Talking film was shot to demonstrate Alan Blumlein's work on stereo in the year EMI was created by the merger of the Columbia and Gramophone companies. This binaural experience (as Alan called it) is still impressive today.
4	FAME	https://tech.ebu.ch/news/fame-for-uhdtv-03jul13	The Forum for Advanced Media in Europe [FAME] was formed from the original European HD Forum, which in turn was set up under the sponsorship of the European Commission, and the European Broadcasting Union [EBU], The main group deals with video aspects of UHDTV whereas the FAME Ad-hoc Audio Group (to give it its full title) deals with audio. Participation in both groups is from a wide assortment of leading broadcasters and broadcast industries, but there is still a prominent role for the European Commission, the EBU and International consumer electronics industry.
5	Scene		Another term for presentation of an entire sound sphere using ambisonic recording techniques.
6	Ambisonics	https://en.wikipedia.org/wiki/Ambisonics	A way of presenting an entire sound sphere which is not dependent on the number and location of loudspeakers for replay. First order ambisonic sound consists of four signals, sound pressure and three difference signals to represent left right, front/back and up/down. This configuration creates only a small "sweet spot" where reproduction is effective, and the sound locations are not precise. Adding more audio signals to create a High Order Ambisonic format (see below) results in more accurate localisation and a larger sweet spot on replay.

Ref.	Term	Link	
7	Orpheus	https://orpheus-audio.eu/	The Orpheus project was funded by the EU Horizons 2020 programme to experiment with the next generation of audio technology by building an end to end, object based, radio broadcasting system. Over 3 years, from 2015, the 10 main european partners made a valuable contribution to both the technologies behind NGA and to exploring the ways they could be used. EBU TEchnical report TR 042 explores this in more detail. https://tech.ebu.ch/docs/techreports/tr042.pdf
8	Speaker formats	https://www.itu.int/rec/R-REC-BS.2051/en	The format commonly described as 5.1 surround sound has 5 speakers around the listener and a subwoofer. [Note: the " .1' represents the LFE (Low frequency effects) channel, which is not necessarily what is reproduced by the subwoofer since the latter is determined by the 'bass management' of the replay system. Under ITU_R BS2051 this is now refereed to as "0+5+0 ". 4+7+0 (previously known as 7.1.4) increases apparent immersion by add four overhead speakers to 7 speakers around the listener and subwoofer. For the the exact, standardised description of speaker layouts and their correct naming consult ITU-R BS2051.
9	Immersive/ Immersion		Audio which is presented in three dimensions, left/right, front/back and up/down. Delivery can be through a number of discrete speakers such as 7.1 + 4 Height, or a 3D sound bar, or binaural rendering to earphones.
10	HOA	https://en.wikipedia.org/wiki/Ambisonics	High Order Ambisonics, $(l+1)^2 = S$ where l is the order of the system and S the number of audio signals required to create that order.
11	Renderer/ Rendering	https://tech.ebu.ch/publications/tech3388	Because NGA is distributed in a format that is agnostic to the number and placement of loudspeakers used for replay, the consumer device has to "render" the NGA stream to the loudspeakers available. Content may also be rendered to binaural presentation for viewers using earphones, as is likely to be the case for mobile platforms. Rendering is also necessary in the production process in order to allow monitoring of the audio being created. The renderer can also apply the personalisation and interactivity features selected by the user.

Ref.	Term	Link	
12	DVB	https://www.dvb.org	The DVB Project is an Alliance of about 200 companies, originally of European origin but now worldwide. Its objective is to agree specifications for digital media delivery systems, including broadcasting. Its an open, private sector initiative with an annual membership fee.
13	EBU	https://tech.ebu.ch/home	The European Broadcast Union is the European Trade Body for public service broadcasters and is based in Geneva, Switzerland. Most famous for being the organisation behind the European Song Contest the EBU has a very active Technology and Innovation department which is supporting lots of NGA activity.
14	ITU ADM Renderer, formerly EBI ABM Renderer EAR	https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.2127-0-201906-I!!PDF-E.pdf	In 2018 the EBU published a specification for a native ADM Renderer: EBU tech.3388. This has now been incorporated in and superseded by the IAR ITU ADM Renderer.
15	EBU NGA Architecture	https://tech.ebu.ch/docs/r/r151.pdf	This technical recommendation R151 explains a strategy for the adoption of NGA.
16	Dynamic Metadata		Most metadata today is fairly static, which means it doesn't change at all (e.g. its different between channels, but each channel stays the same) or it changes relatively slowly, may be on a daily basis or with each programme. However dynamic metadata could be changing with every TV frame or even faster, meaning some data could be updated 50 times a second.
17	IMF	https://tech.ebu.ch/imf	The interoperable Mastering Format defined by the SMPTE in ST 2067 and is designed to enable the delivery of complex and adaptable packages of media, allowing multiple presentations or combinations to be made available for different uses, based on the one set of reference files, thus removing the need for much duplication and reversioning.
18	ATSC	https://www.atsc.org	ATSC (the Advanced Television Systems Committee), with headquarters in Washington DC, develops voluntary international standards for digital television.Its member organizations represent the broadcast, broadcast equipment, motion picture, consumer electronics,computer, cable, satellite,and semiconductor industries.

Ref.	Term	Link	
19	HbbTV	https://www.hbbtv.org/overview	Hybrid Broadcast Broadband TV is both an industry standard and promotional initiative for hybrid digital TV. It aims to harmonise the broadcast, IPTV and broadband delivery of entertainment to the consumer through connected TVs (smart TVs) and set-top boxes.
20	ETSI	https://www.etsi.org	ETSIis a European Standards Organization (ESO),dealing with telecommunications and broadcasting services, amongst others. Its HQ is located in Sophia Antipolis in France.
21	ISO	https://www.iso.org/home.html	ISO is an international Standards Organization dealing with a vast range of industries, and with its central secretariat located in Geneva.
22	ADM	https://www.itu.int/rec/r-rec-bs.2076/en	The Audio Definition Model describes the structure of a XML metadata model to define the content and format of tracks in a (Next Generation) Audio file. It's an open specification originally published in 2014 by the EBU as tech.3364, It became an ITU Recommendation (ITU-R BS.2076-1) in 2017.
23	ITU	https://www.itu.int	The International Telecommunications Union is a UN standards body based in Geneva.The ITU-R sector provides radio communications standards including those for broadcasting.
24	S-ADM		SerialADM (S-ADM) is an adaption of ADM to enable its use in live streaming. Rec. BS.2125 was first published in 2019.
25	EBU codec independant NGA workflows	https://tech.ebu.ch/docs/techreports/tr045.pdf	A Technical Report on why broadcasters need an open, codec-independant workflow for NGA production.
26	EBU Evaluation of IMF for Broadcasters	https://tech.ebu.ch/docs/techreports/tr040.pdf	EBU Technical Report (TR 040) 2017 presents an attempt to evaluate IMF for use in broadcasting.
27	Profile	https://tech-edu.ch/docs/tech/tech3392.pdf	The EBU technical document Tech. 3392 describes the ADM broadcast production profile. This profile constrains the ADM to simplify implementations and to prevent interoperability problems in the production of Next Generation Audio broadcast programmes.



EBU

OPERATING EUROVISION AND EURORADIO

f@m€

Forum for Advanced Media in Europe