

# SCENE BASED AUDIO: A NOVEL PARADIGM FOR IMMERSIVE AND INTERACTIVE AUDIO USER EXPERIENCE



At the heart of devices you love

# Qualcomm Technologies, Inc.

Qualcomm Snapdragon is a product of Qualcomm Technologies, Inc.

Qualcomm and Snapdragon are trademarks of Qualcomm Incorporated, registered in the United States and other countries. All Qualcomm Incorporated trademarks are used with permission. Other products and brand names may be trademarks or registered trademarks of their respective owners.

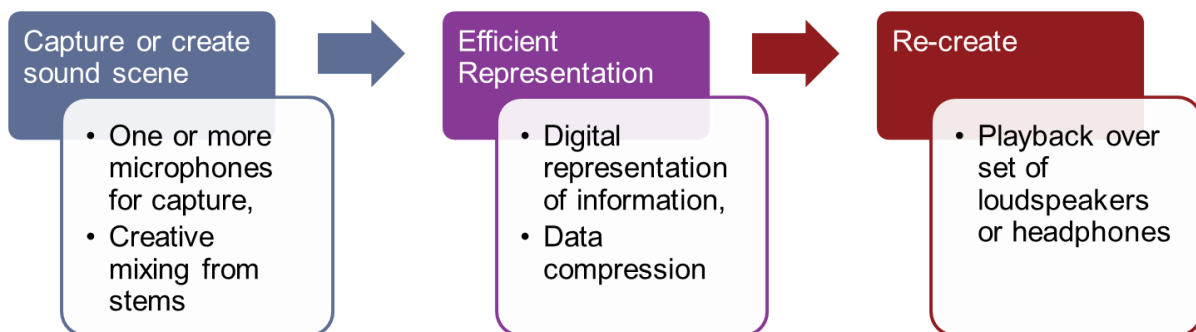
Qualcomm Technologies, Inc.  
5775 Morehouse Drive  
San Diego, CA 92121  
U.S.A.

© 2015 Qualcomm Technologies, Inc. All Rights Reserved.

# 1 Immersive audio – the experience we seek


Our ears and our brain bestow upon us the ability to identify the position of various sources of sound around us. While our ears receive sound waves from various sources in our environment, our brain looks for cues in the received waves that can tell, to a great degree of accuracy, the positions of the object emitting the sound. In some ways, we can imagine ourselves as being immersed or enveloped in constantly evolving sound scene. One may even say that our perception of our environment is closely tied to how we perceive the sound we are immersed in. Unlike our visual perception that is limited by our field of view, our auditory perception is 3D all the time – we can easily estimate the location and proximity of audio sources that are behind us even though we may not be able to see them.

As modern day consumer electronics devices attempt to create a sense of audio visual envelopment while entertaining its customer base, several technical challenge lie in capturing, representing and rendering the visual and sound scenes as authentically as it happens at the capture location or as it is intended by creative artists. As human perception of sound is truly 3D, a key technical challenge is in accurately representing the time varying sound scene while living with the limitations of various capture mechanisms, transmission bandwidths and rendering devices.



**Figure 1: Spatial Audio Coding Framework**

Increasingly viewing movies, television broadcasts and interesting user generated content is moving from traditional to mobile devices. Content hosting services such as Netflix, iTunes etc. offer viewers the choice of viewing their favorite audio and video content anywhere from home theaters to mobile devices and in cars. As a result, streaming services are becoming a popular



choice for content creators for reaching out to a growing demographic of viewers with a preference for viewing content on the go over mobile devices such as tablets and smartphones. Recent technical advances in mobile multimedia such as 4K and High Dynamic Range (HDR) video coding have made the visual experience so compelling that content providers are more excited to deliver content over mobile devices than ever before. Spatial audio is an integral part of this multimedia experience and significant opportunities exist for providing a far more immersive audio experience for consumers of content over mobile devices.

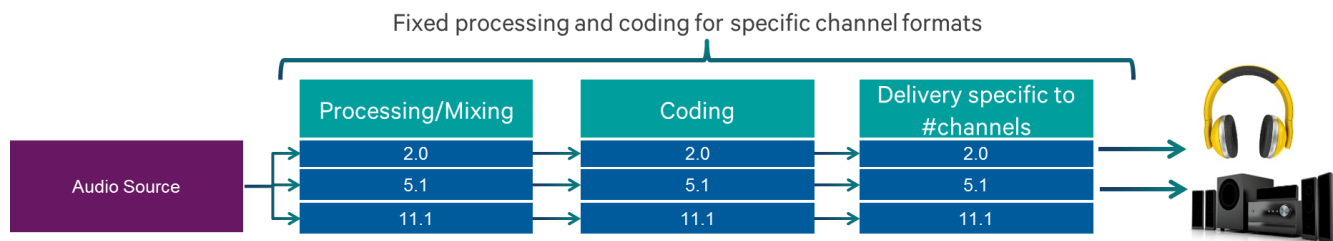
There is also a significant consumer demand for compelling content created by self-capture technology and fuelled by the popularity of social media. As a result, services such as YouTube, Daily Motion and Vimeo that enable sharing of user generated audio and video content with friends and family have experienced tremendous growth over the past decade.

## **2 The story so far – and why it can get better**

The traditional approach to coding spatial audio, which we call channel based audio coding, has largely focused on the reproduction end of the chain (“Re-create” in Figure 1). If the surround sound is to be played out in room with a 5.1 speaker system, then capture and the representation mechanisms would aim to create the signals that would go into each of the 6 speakers (aka speaker feeds or channels) so that the sound scene in the playback room matches that at the source location (in case of a live audio recording) or an artist’s intent (e.g. music or movie sound track) as closely as possible. For a different configuration of speakers in a room (7.1, 7.4.1, 11.1 etc.), a different set of signals would have to be generated to recreate the sound scene optimally. In the case of a music or movie sound track, the sound scene may be composed by a mixing artist who would place sound sources in a virtual space to create the track for a given playback speaker configuration. In a capture scenario like recording during news or sports broadcast, a mixing console generates speaker feeds from audio captured by microphones placed in the capture location.

The limitations of the traditional channel based audio representation systems are quite obvious. First, it is closely tied to the expectations that the playback location will have the exact loudspeaker count and position as the capture and representation system assumed. For

example, when the sound scene is represented by a 5.1 audio signals the playback location is expected to have the 5 speakers in predetermined position for the best audio experience. If the playback location is configured with a 7.1 or a 2.0 stereo system or if the speakers are not located in the prescribed positions, then all bets are off as far as optimal surround sound experience is concerned. So as we may observe, the constraints on the consumer side (the number of loudspeakers) adversely affects the capture/production side. The capture/production side has to make assumptions about the system over which the audio will be played back at the reproduction side and if these assumptions don't match the reality, the audio experience is not optimal.

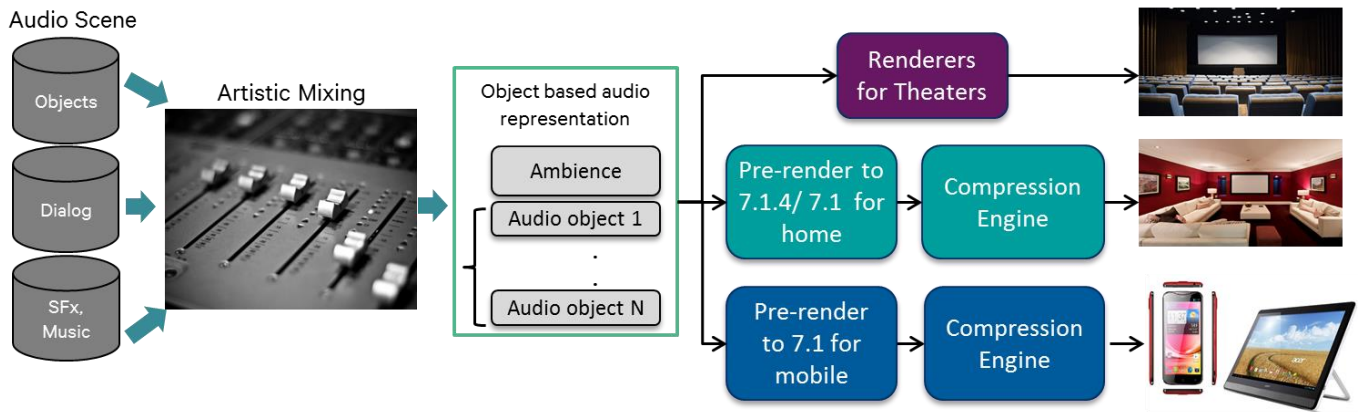


**Figure 2: Today's channel based spatial audio coding**

As there is a growing demand in the entertainment, broadcast and user generated content industries for delivering audio content that provides the best immersive experience across all playback scenarios, the constraints of the traditional channel based audio representation quickly become a disabling factor. To break the barrier, what is needed is a technology where the content creation is decoupled from the reproduction at the playback location.

Object based audio is a first step towards representation of an audio scene that is agnostic to the configuration of the loudspeakers. Instead of focusing on the reproduction end of the chain, object based audio representation aims to represent discrete sources of sound in the auditory scene. These discrete sound sources, also called objects, are identified by their location coordinates and the sound signals that they emit. A renderer at the playback end of the chain is tasked with the job of recreating the object at the specified position. Being at the playback location, the renderer can be made aware of the number and position of the speakers and therefore can generate the appropriate signals for the speakers to create the spatial localization of the sound object. Since all sound scenes cannot be completely composed on

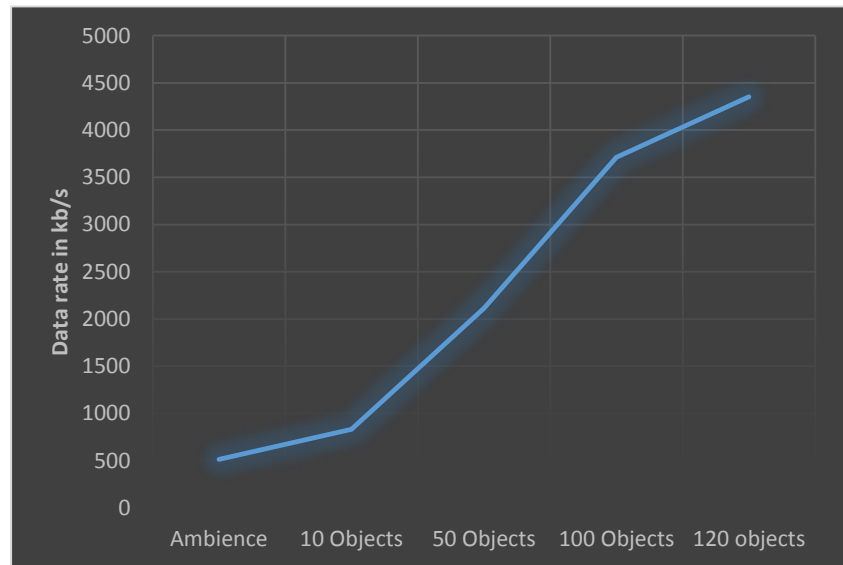
very narrowly located point discrete objects, the spatially diffused ambience is still represented in the channel based format. For example, the sound scene in a restaurant can be split into sound coming from discrete point sources (people talking, clinking cutlery etc.) which is amenable for object based representation and the background (music etc.) which is then represented using a channel based representation.



**Figure 3: Object based spatial audio coding for movie and music**

While object based spatial audio representation infuses a greater degree of realism to the reproduced sound by improving the resolutions of point audio sources, its efficacy is limited by the complexity of the sound scene. In a scene where there are hundreds of point sources (for example: claps in a stadium or concert hall), it becomes

practically impossible to represent each source as a discrete object. The amount of information required to represent all these sources and the complexity of rendering them at their specified location grow quickly with the number of objects and object based representation is no longer a viable option. It is perhaps for



**Figure 4: Cost of coding audio objects**

this reason that object based audio representation technologies revert to channel based representation for playback in homes and over mobile devices. Besides this, object based representations are not well suited for capture live audio content as isolating discrete objects during capture of a sound scene is not easy.

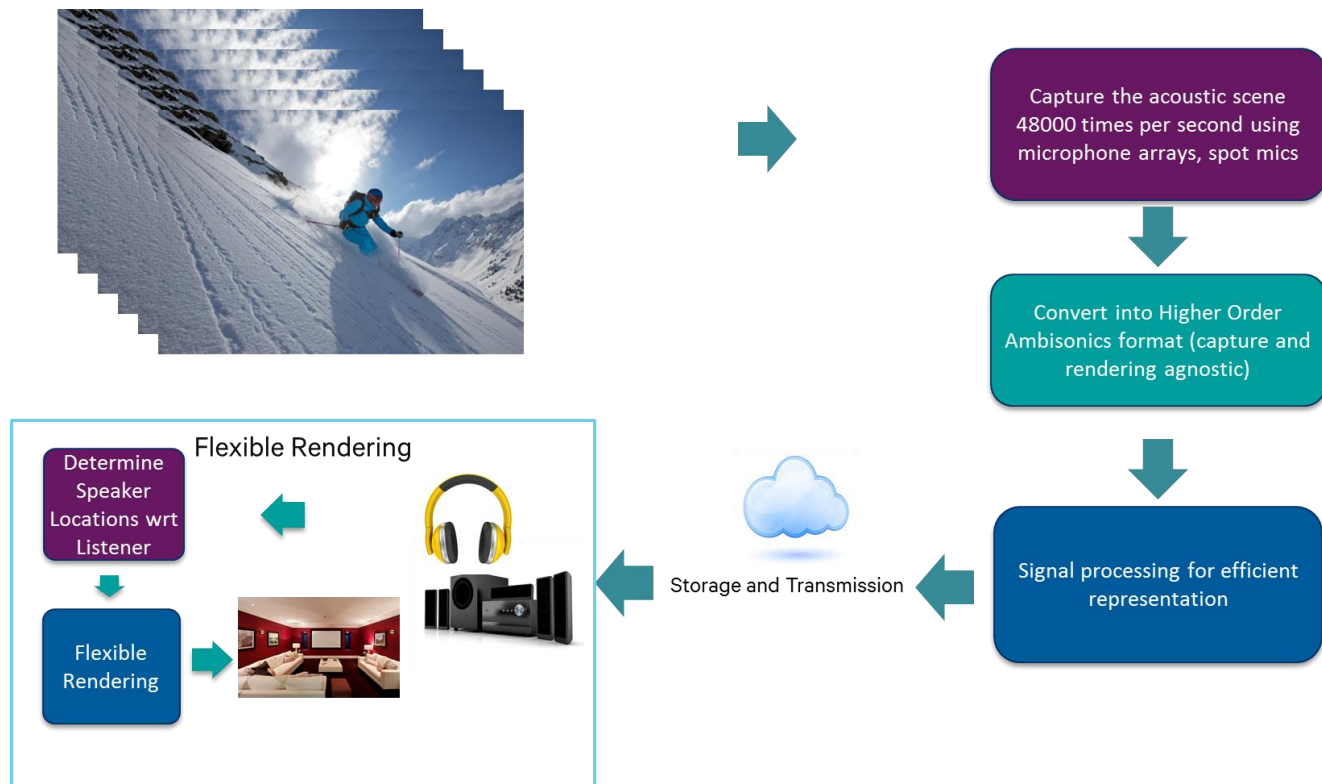
Now imagine an audio technology that truly enables decoupling the production/capture from the reproduction/ playback process without the disadvantages of the object based audio representation. Imagine a technology enables you to keep your audience immersed in enveloping sound regardless of how many loudspeakers are employed or where the loudspeakers are located as well as over headphones. Imagine the same audio technology can be used to capture live 3D audio during sporting events or news coverage and represent the sound track in a music or movie. This technology would be the panacea of spatial audio coding. This technology is scene-based audio coding.

### **3 Scene-based audio coding – a panacea**

Scene-based audio coding marks a big leap forward in capturing, representing and rendering spatial audio. As with many disruptive innovations, it is based on a rather straightforward and intuitive paradigm for capturing, representing and rendering spatial audio signals. In a given 3D space, sound travels as pressure waves. This means that the perfect representation of the sound in that space can be achieved if one knows the pressure at every point in that space at all times. Scene-based audio representation attempts to do precisely that – the sound scene at a given instance of time is represented as a field of pressure values at all points in that space at that time and this is done for every instance of time when the sound field is sampled.

A natural question then is how one could grapple with the massive volume of data if one has to represent the pressure values at every point in space, and do this say, 48000 times every second. This is where spherical harmonics comes to the rescue. It provides a neat mechanism for representing the pressure values at all points in a 3D space using a small number of coefficients with a rather amazing degree of accuracy. Employing spherical harmonic based transforms, the daunting task of representing the pressure values at every point in space is drastically simplified to the task of figuring out these coefficients. Since the pressure at a given point in space is related of the pressure in the neighboring points, one can place a few spatially

separated microphones in the sound scene and employ simple mathematical operations to derive the above mentioned coefficients. For content generated by music artists and movie track mixing engineers, all that they need to do is to create the sound scene just as they do for existing channels or objects based technologies. From that point, it is again the same simple mathematical operation as mentioned above to generate these coefficients that would represent the sound scene in 3D. It is really that simple!



**Figure 5: Scene-based audio coding - a path breaking approach**

### 3.1 For the mathematically inclined

Here is a deeper dive into the mathematics and physics of the representation. Feel free to skip this section if you are so inclined after noting that the coefficients mentioned in the previous section are also called the Higher Order Ambisonics (HOA) coefficients.

Consider a 3D space with the origin in the center of the space and any arbitrary point identified by the distance from the origin,  $r$ , the elevation of the point with respect the horizontal point  $\theta$ , and the azimuth,  $\phi$ . At a given time instance  $t$ , the pressure at this point is denoted as  $p(r, \theta, \phi, t)$ . As mentioned above the pressure at every point in the space constitutes the



pressure field. For the region reasonably close to the origin, the pressure field is related to the coefficients mentioned above by the following relationship:


$$p(r, \theta, \varphi, t) \approx \left[ \sum_{n=0}^N j_n(kr) \sum_{m=-n}^n a_n^m(t) Y_n^m(\theta, \varphi) \right] e^{j\omega t}$$

In the above equation, the spherical harmonic coefficients are  $a_n^m(t)$ . Since these coefficients are a function of time, we call them the higher order ambisonics (HOA) signals. As mentioned in the previous section, these signals enable us to represent the pressure field quite accurately using a small number of signals. As a side note, as  $N \rightarrow \infty$  the relationship in the equation above becomes an equality.

The value  $N$ , also called the order of the HOA determines the accuracy of representation of the pressure field. One would choose a larger value of  $N$  to obtain a more accurate representation of the pressure field at distances far away from the origin.  $j_n(kr)$  is a Bessel function of the first kind and  $Y_n^m(\theta, \varphi)$  represent the spherical harmonic basis functions. Without getting deeper into the details, it is enough to note the following:

1. The order  $N$  determines the number of coefficients one needs in the above equation. For a given  $N$ , the number of HOA coefficients is  $(N + 1)^2$ . So for a 4<sup>th</sup> order scene-based representation, one needs 25 HOA coefficients.
2. If we know the pressure at certain locations in the sound field (i.e. if we know  $p(r_1, \theta_1, \phi_1, t), p(r_2, \theta_2, \phi_2, t), \dots, p(r_k, \theta_k, \phi_k, t)$ ), then the HOA coefficients  $a_n^m(t)$  can be easily estimated. We can then plug in these coefficients back into the above equation to get a fairly accurate representation of the sound field in a region not too far from the origin.
3. In practice, a 3<sup>rd</sup> or 4<sup>th</sup> order coefficients provide sufficient perception of localization of sound sources in a scene.

The theoretical framework for first order ambisonics ( $N = 1$ ) was heuristically derived in the 1970s. However, this technology has matured since the 1970s. Today the underlying theory has evolved and is grounded in physics and mathematics and goes beyond just 1st order



ambisonics. With Qualcomm Technologies' involvement in this area, commercialization of various technologies based on HOA is imminent.

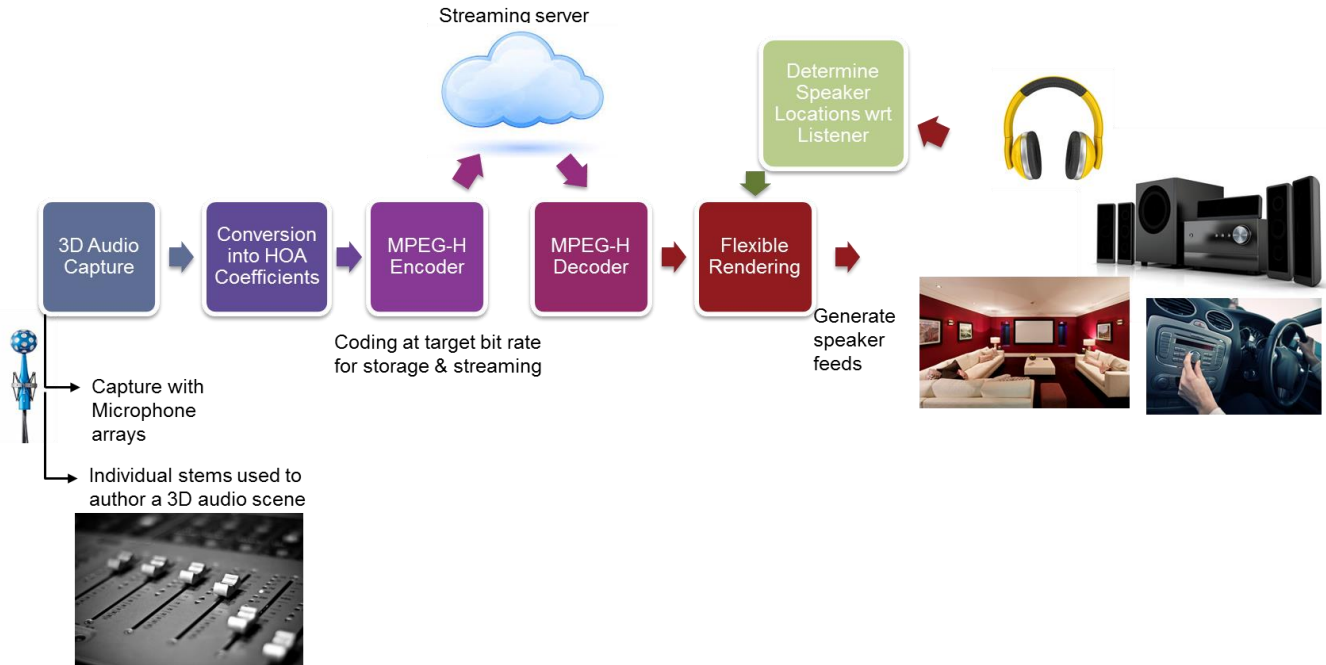
### **3.2 Break free from shackles**

Mathematical details aside, it is worth reiterating that once we derive these coefficients, we now have a complete representation of the sound scene and effectively, this means that we have decoupled the representation from the capture. In other words, these coefficients represent the pressure field and one no longer needs to bother about where the microphones were placed, how many and what type they were and their characteristics at the capture location or how the sound scene was created by a mixing artist.

Likewise, the representation is also decoupled from the system that is used to playback the audio. The flexible renderer at the playback location receives the representation of the pressure field that was captured and or created originally at the original sound scene. Being at the playback location, the renderer can be made aware of the number and location of the speakers. The job of the renderer now boils down to generating signal feeds for these speakers in such a way that the sound field created in the playback location matches that of the capture location (or the artistic intent) as closely as possible.

## **4 The big picture**

Before we proceed to reviewing the benefits of the scene-based spatial audio coding, lets briefly look at the entire end to end chain of a scene-based 3D audio capture, representation and rendering system.



**Figure 6: Scene-based audio coding workflow**

There are several ways one can create spatial audio content which will get represented in the scene-based audio representation format. Live events such as news, sport shows or end-user generated content (e.g. Home videos) can be captured using an array of microphones. Again, the goal is to have enough microphones in appropriate locations so that the HOA signals can be easily estimated. There are ideal configurations for such microphones where 32 microphones are optimally located on the surface of a sphere. This microphone ball with 32 microphones can help capture 4<sup>th</sup> order HOA coefficients. An example of such a microphone is the Eigenmike (1). The HOA coefficient can also be derived from use fewer microphones and at suboptimal locations with marginal inaccuracies. One can also place individual microphones at spatially separated arbitrary locations and use the capture from all of them to derive the HOA coefficients. The capture system also offers the flexibility of using any combination of the ball microphone with one or more spot mics. The signals captured from these microphones can be used to estimate the HOA coefficients that represent the sound field at the capture location.

Movie and music mixing artists often compose their sound scenes by spatially placing individual components of the scene (called stems) such as dialog, music and special effects employing a digital audio work station (DAW). Since the scene-based audio format can easily


represent depth and distance perceptions current DAW can be augmented with tools that can help these artists place objects in a truly 3D space. Finally the output from the DAW can be converted into the HOA coefficients similar to the live capture.

Mixing live captured content with pre-recorded content is also straightforward as the HOA coefficients representing each of these can be summed up. Also, since the HOA coefficients use spherical harmonics as the basis, it is mathematically easy to rotate, stretch or compress the sound scene.

As mentioned earlier, for an  $N^{th}$  order presentation of the sound scene, one needs  $(N + 1)^2$  HOA coefficients. So for a 4<sup>th</sup> order presentation, we would end up having 25 HOA coefficients. That would be the representation of the sound scene at a given instance of time. For good sound quality, one would then need the sound scene to be captured 48000 times every second. Other sampling rates (higher or lower than 48000 samples/second) can also be considered as per quality and system requirements. In the end, one would have a 25 digital PCM signals (the HOA signals) sampled at 48 KHz, each representing on HOA coefficient. The data rate at this stage is approximately 28 Mb/s, making it challenging to transmit over band limited channels. Qualcomm Technologies' spatial encoding module comes to the rescue here. It takes these 25 HOA signals and by employing signal processing tools for dimensionality reduction such as singular value decomposition, reduces it to as few as a set of 6 PCM signals and a small side channel of metadata without any significant quality loss. This reduced PCM signal representation is called the "mezzanine format". This makes it conducive for distributing HOA signals over TV plants. The MPEG-H compression engines may be used to compress the HOA signals to an effective bit rate in the range of 96 kb/s to 1.2 Mb/s.

## 5 The value proposition

Flexible rendering is one of the key benefits of the scene-based audio representation. With the representation no longer tied to a specific speaker configuration, the flexible renderer can generate speaker feeds for any number and location of speakers so that the best possible reproduction of the sound scene is created. It is also possible to generate feeds for headphones that provide the same immersive user experience as in the room.



Scene-based audio representation is a disruptive innovation that overcomes some of the biggest challenges in spatial audio coding based on the traditional channels and object based representations. Before moving on to some of the key applications of scene-based audio representation, let's summarize the significant benefits this technology brings to the world:

- True 3D sound: The technology enables content creators to easily capture or create truly 3D sound scenes including proximity and depth components that would not be possible with legacy formats,
- Live recording of HOA scenes in both a professional and consumer settings with little or no human intervention is a tremendous advantage. Both of these are ideally suited to the mobile device, which is essentially a conduit to enable both the acquisition and playback over a large and diverse set of transducers.
- Efficient representation: Using Qualcomm Technologies' spatial processing and the MPEG-H compression engines, the captured 3D sound scenes can be compressed to any bitrate among a wide range of bitrates.
- Universal Format: Loudspeaker agnostic format that can be adapted to the local loudspeaker geometry and acoustic landscape to offer optimal immersive sound playback in any location. Single format for all rendering scenarios. There is no need to encode content separately for 2.0, 5.1, 7.1 etc.
- One format for theaters, home or for mobile.
- Uniformity of Experience: Flexible enables you to stay immersed in sound everywhere: in theaters, at home or on the go

## 6 Future of spatial audio - in a product dear to you

Qualcomm Technologies has invested considerable resources in the development of the scene-based 3D audio coding

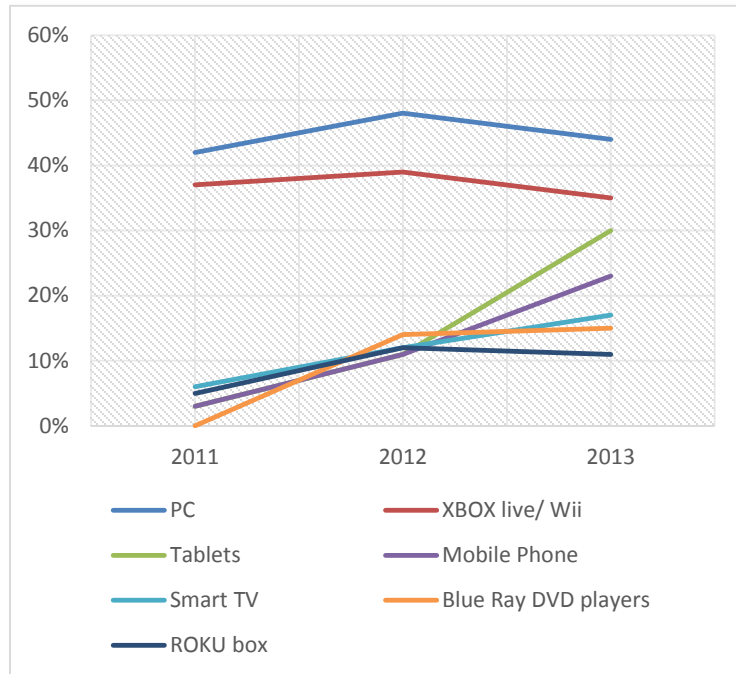
technology with an aim of advancing the state of the art in providing an immersive audio user experience to the billions of consumers of audio and visual content across the globe.

Mobile devices are playing an increasingly important role as consumer's preferred devices for consumption of AV content. For example, Nielsen research (2) over the past few years has shown a steep growth in subscribers to content streaming services like Netflix who enjoy content on their mobile devices.

Today, smartphones and tablets

interface with home theater systems or speaker bars to deliver surround sound experience in homes. As technologies that enable mobile devices to wirelessly connect to speakers mature, one can envisage multichannel audio being played out through a set of speakers wirelessly connected to a smartphone or a tablet in the near future.

The advantages of scene-based 3D representation provide content creators, including broadcasters, movie and music artists, media production houses and user generated content creators the ability to deliver their audio content immersively across several end user devices without compromising on the artistic intent. In this section, we present 3 case studies that demonstrate how our scene-based 3D audio can be integrated into workflows for creation and consumption of spatial audio content.




**Figure 7: Netflix viewership on various CE Devices (2)**

## 6.1 Commercial Broadcasting – benefits of scene-based coding

Scene-based audio coding is a natural fit for live broadcasting applications such as a news or sport events. By employing scene-based audio representation, next generation television and radio broadcast systems can provide a compelling and immersive audio user experience over various consumer electronics devices include mobile devices connected to headphones, home theater systems with stereo or multiple speakers and speaker bars. Here is why -

- **Easy capture of live audio scene:** Signals captured from microphone arrays and/or spot microphones can be converted easily into HOA coefficients in real time.
- **Universal format:** With scene-based audio for MPEG-H, there is no longer a need to create multiple mixes for stereo, 5.1, 7.1.4 etc. The underlying sound field representation ensures a consistent and accurate playback of what was actually present (or artistically intended).
- **Interactive dialog and commentary options:** Scene-based audio representation can be combined with audio objects representing commentary or dialog to provide interactivity and personalization features to the end consumer. For example, people watching a sporting event will now be able to choose commentary in their preferred language or a preferred commentator or even mute commentary altogether.
- **Audio scene manipulation during playback:** Scene-based audio representation includes all information about the source audio scene. As a result, it provides the end user the flexibility to alter the point of view, focus on a specific direction in the sound field and rotate the sound field. This capability, unique to scene-based audio representation, opens up a plethora of opportunities for manufacturers of consumer electronics devices to provide value added and differentiated features to their viewers.
- **Flexible rendering:** And again, flexible rendering allows the reproduction of the immersive auditory scene regardless of speaker configuration at playback location and on headphones.

The existing infrastructure for audio broadcast that is currently employed for transmitting channel based spatial audio (e.g. 5.1 etc.) can be utilized without making any significant changes to enable transmission of HOA representation of the sound scene.

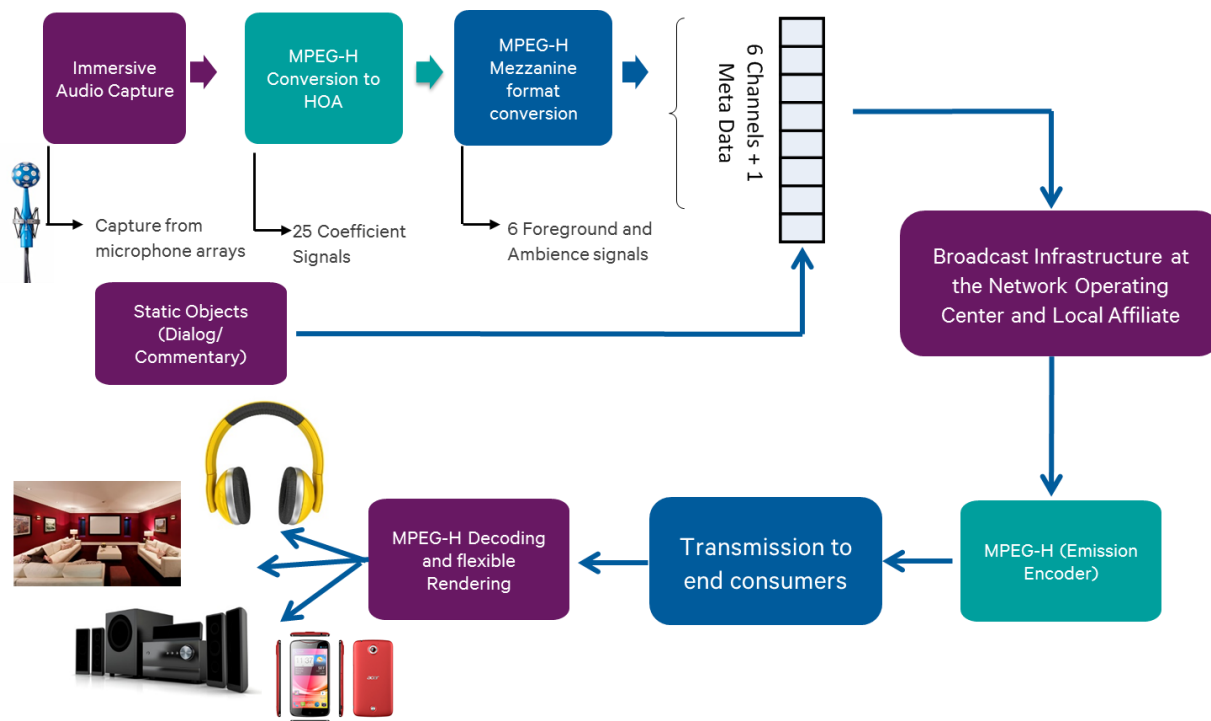


Live events, TV shows and sports can be captured real time by suitably placing microphone arrays (like the Eigenmike) and/or spot mics in the field. The sound capture from the microphone arrays is first fed into a mixing console where additional commentary objects can be added to the sound field or sent as separate elements, if desired. The mixing console can help the mixing engineer place these objects in the 3D space that includes the microphone capture. The process of estimating the HOA coefficients can be integrated into the mixing console so that the output from the mixing console is a set of HOA coefficients that represent the captured audio scene. The audio signal from the broadcast truck is then transmitted to the network operating center (NOC).

Many interconnects in TV plants use SDI infrastructure. Legacy SDI (SD-SDI) devices can accommodate at most 8 channels of audio besides video signals. To fit HOA signals into legacy SDI interconnects, a spatial encoding is performed on the HOA coefficient signals. As described in Section 3, for a 4<sup>th</sup> order HOA representation, the spatial encoding process reduces the 25 HOA coefficient signals into as few as 6 transport signals (called the mezzanine signal) with an additional low bit rate meta-data channel. These 6 transport channels and the meta-data channel can be accommodated in 7 of the 8 channels in the SDI interface. The additional channel left over in the SDI interface can then be used for transporting objects such as advertisement, commentary or effects. It is worth noting that adding such objects would not be feasible or would come at significant cost with the traditional channel based audio coding technologies. Given that the one would have to transmit both a 5.1 mix and a stereo mix to accommodate for different rendering possibilities, the channel based audio transmission uses up all available 8 channels for surround audio and therefore leaves no room commentary or advertisement options. In other words, scene-based audio representation allows interactivity (ability of the end user/local affiliate manipulate the object signal that is transmitted over the 8<sup>th</sup> slot of SDI). Furthermore, the 6+1 mezzanine format can provide a complete 3D immersive audio listening experience over any number of loudspeakers (2.0, 5.1, or even 22.2). In comparison, a traditional 5.1 audio format provides only limited 2D surround over 5+1 speakers.



The local affiliates then receive these 6 transport channels, the Meta-data channel and any object channels that come through the SDI interconnect. At this stage, the local affiliate has the flexibility to insert local advertisements. The local affiliate can encode the 6 mezzanine signals using the MPEG-H encoder to compress the data for transmission to the end user.



**Figure 8: Using HOA in an existing broadcast workflow**

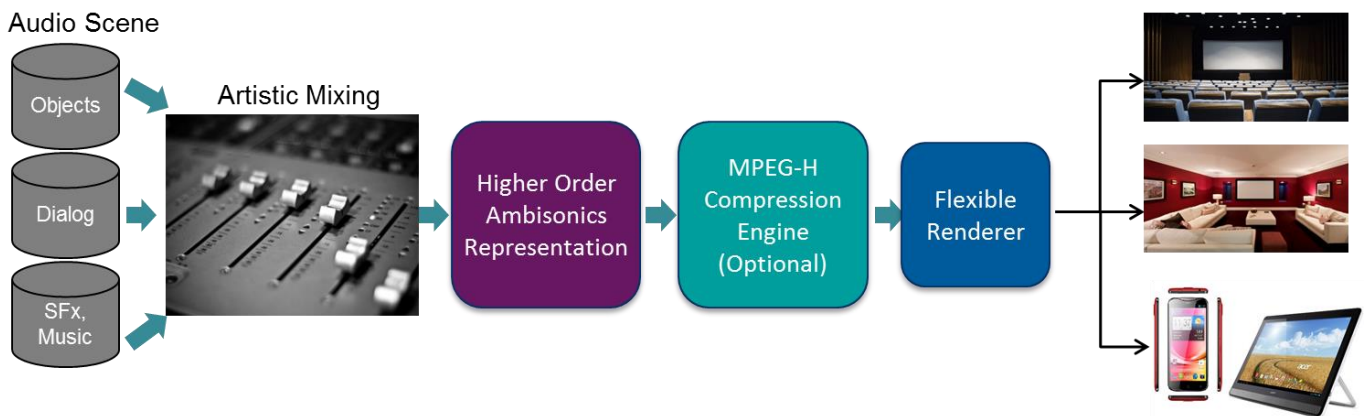
Monitoring the HOA signal quality at different points in the broadcast system is also straightforward. At any point where the mezzanine signals can be tapped, the audio quality of the signal can be assessed by a) first converting back the mezzanine signals back to HOA signals and b) rendering the HOA signals through a flexible renderer configured to generate the ideal speaker feeds for the monitoring loudspeaker configuration.

## 6.2 Movie post production – New opportunities with scene-based audio coding

The diversity of consumer electronic devices over which people watch movies and listen to music has grown dramatically over the last few years as shown by the Nielson report Figure 7.

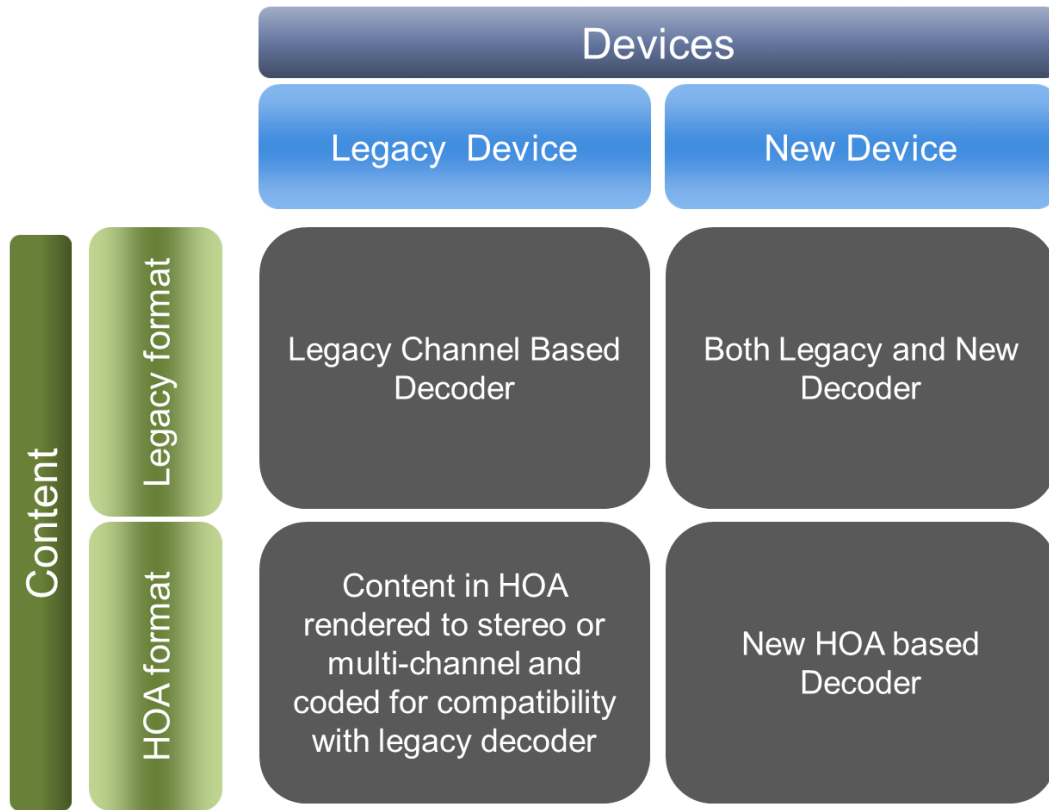
While this provides movie producers and artists several new venues for displaying their creations, ensuring that the audio quality preserves the artistic intent and quality is a significant challenge. Scene-based audio coding provides a way for achieving this.

Scene-based audio representation can be efficiently used for generating movie or music sound track. During movie post production, creative artists and mixing engineers create the sound scene by mixing dialog stems, special effects and music sound tracks from a database of these effects using a mixing tool such as ProTools by AVID. The scene created conveys the artist's intent. With current channel based audio technologies, the post production process has to come up with one mix for delivery in movie theaters, one other for home theater playback (typically 5.1), a 7.1 mix for Blu-ray playback, and maybe one for stereo playback. With the scene-based audio representation, all that the mixing artist and engineers need to create is a single mix of HOA coefficient signals. The flexible renderers can then generate the appropriate signal feeds based on the number and location of the loudspeakers for each playback location (theater, home or on mobile devices).



**Figure 9: Scene-based audio coding in movie workflow**

One of the key concerns for the movie industry pertains to compatibility of legacy consumer electronic devices (TV, STB etc.) to the new format, if they create content in the scene-based coding format. The table below addresses this concern. For legacy devices, content in the scene-based format can be pre-rendered to the format supported by the legacy device and played back.



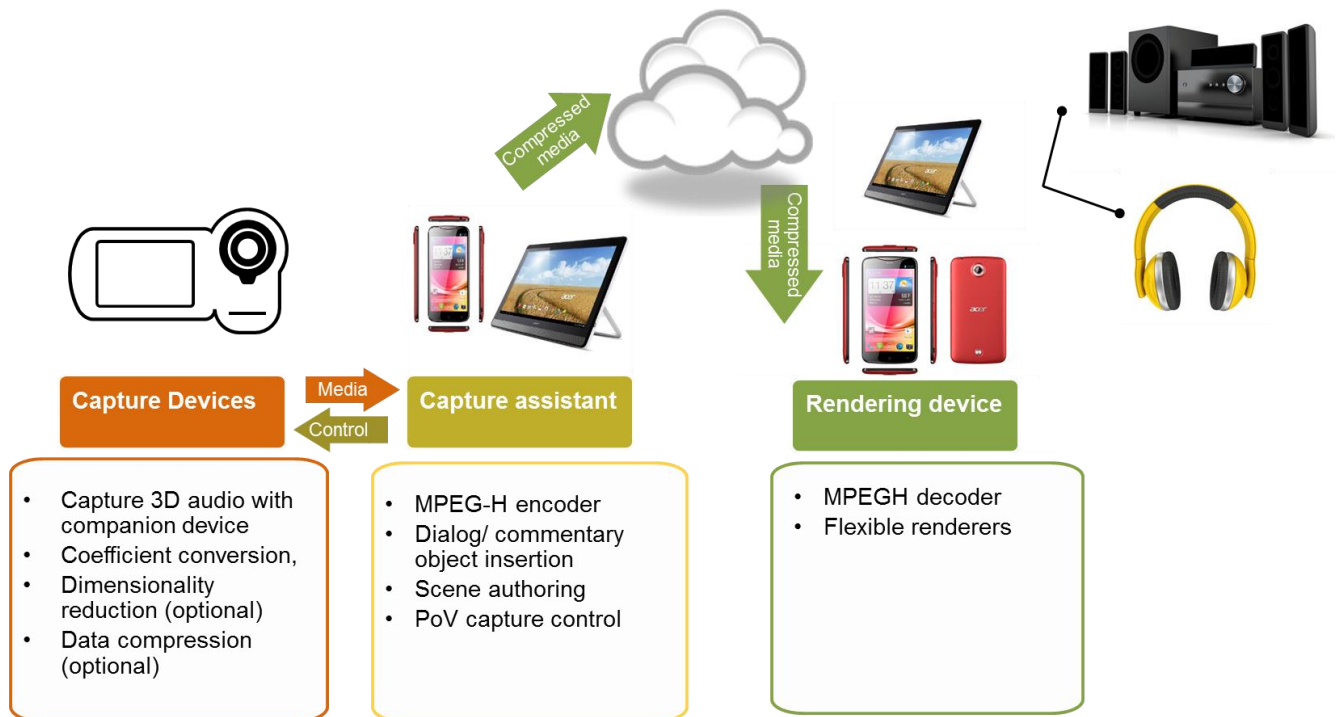
**Figure 10: Dealing with the legacy**

### 6.3 User generated content

Recently, there has been growing consumer demand for compelling content created by self-capture technology and shared via social media. Cameras including smartphone and tablet cameras, standalone professional camcorders, sports cameras that are used for personal action video photography have helped the common man generate amateur personal videos which are shared on social media websites. The growth of this market segment has been precipitated by the significant technological advances in image capture, sensor, video coding, and display technologies.

While most of these devices visually capture a user’s point of view or a panorama surrounding the user, the audio capture is largely limited to stereo. The pairing of scene-based audio capture, representation and playback technology with these camcorders and sports cameras can enable end users to generate even more compelling and immersive audio content. The

figure below shows how a scene-based audio capture technology can be paired with a camera product.




**Figure 11: Pairing immersive acoustic capture with cameras**

With scene-based audio representation, manipulating the sound field is relatively straightforward. One can stretch and compress the sound field, rotate it or focus the capture in a certain direction. One can also create simple post authoring tools that allow viewers to manipulate the sound field during playback.

## 7 So, in a nutshell

Scene-based audio representation is novel and revolutionary paradigm that offers some fundamentally new value propositions to the professional audio, broadcast, user-generated content and streaming industries. By effectively decoupling the audio representation from the capture and the rendering mechanisms, scene-based audio coding overcomes some of the key limitations of traditional systems to provide an immersive user experience across a wide range of listening scenarios including theaters, home, automotive and on mobile devices. Scene-based audio also lends itself well to personalization of the listening experience by



allowing listeners to focus in on a specific zone or rotate the sound field, or can be used in combination with existing techniques, such as audio objects. Deployment of new audio systems based on scene-based audio representation provides true 3D immersive audio while significantly leveraging the existing infrastructure for audio broadcast and streaming, thereby minimizing the capital expenditure for upgrade. Further benefits in terms of greater audio immersiveness and interactivity can be derived from the scene-based audio format with upgrade in the infrastructure.

## 8 References

1. **mh acoustics**. Eigenmike microphone. [Online] <http://www.mhacoustics.com/products>.
2. **Neilson**. How are Netflix and Hulu users Streaming. [Online] 2011-2013.