

# Using Audio Objects and Spatial Audio in Sports Broadcasting

Ulli Scuda<sup>1</sup>, Hanne Stenzel<sup>1</sup> and Dennis Baxter<sup>2</sup>

<sup>1</sup>*Fraunhofer Institute for Integrated Circuits, Erlangen, Germany*

<sup>2</sup>*DennisBaxterSound, LLC, Atlanta, Georgia, USA*

Correspondence should be addressed to Ulli Scuda ([ulli.scuda@iis.fraunhofer.de](mailto:ulli.scuda@iis.fraunhofer.de))

## ABSTRACT

This paper gives an overview on how immersive sound and interactive audio objects can be used in sports TV broadcasting with regard to the characteristic sound elements of different sport families. Some of these are presented and investigated with the question in mind, how audio objects can be captured in real time and how spatial audio can be used to make sports TV more exciting. New broadcast standards target at bringing object based audio and 3D-Audio to the homes. Once the technical infrastructure has been, content is needed to make use of the innovative features. Sports TV broadcasting is a promising field for new audio standards because of its wide range of sound scenarios and its high market value.

## 1. INTRODUCTION

Today, sports is a huge business that on the one hand involves a multitude of kinds of sport, such as indoor and outdoor sports, single person activity sports and team sports or even equestrian and motorized sports. On the other hand, lots of industries and technologies work together to make sport events an exciting experience for both on location audiences and TV viewers.

New broadcast technologies like MPEG-H that feature object based audio as well as 3D-Audio can bring sports TV to a new level of audience involvement by offering interactive applications and immersive sound.

### 1.1. Overview of MPEG-H

In MPEG-H very efficient audio and video codecs are combined and standardized for the transmission of, e.g., 8k video with corresponding high fidelity audio channels and next generation audio features [1].

These features consist of interactive audio, enabled by the transmission of audio objects together with important metadata on content, position and level of the audio or presets that feature specific sounds [2]. Audio objects could be commentators in multiple languages, specific sounds and different groups of the audience. By transmitting objects, the viewers can control the sound elements with the help of an interface that displays the individual source. Favoured objects can be boosted or unwanted sound objects can be turned down.

The high fidelity audio that is offered with MPEG-H consists of a very bitrate efficient transmission of audio channels for 3D-Audio formats such as 7.1+4, which is a regular 7.1 setup with four additional height speakers [3]. These additional height channels are a production advantage as they enable immersive sound experience and add greater definition to the audio reproduction. 3D-Audio reproduction setups offer great possibilities to unmask a lot of sound, give a precise three-dimensional soundscape of the places shown in TV and thus enhance the viewer's anticipation.

Adaptive to the production format of different parts of the program, MPEG-H supports a number of different reproduction setups and transmission formats. In this way 3D-Audio can be transmitted as channel based content, as object based content or with the use of Higher Order Ambisonics (HOA) encoded signals.

Delivery of audio is improved by adaptive playback and intelligent reproduction by the end-user's device. The high fidelity audio objects are rendered at reproduction by the consumer as a final mix. In the course of this rendering process, also the reproduction setup is taken into account. In this way, formats with a higher number of channels than available speakers can be played back by taking advantage of the active format converter, which is part of MPEG-H.

## 1.2. Sports TV

Sports video and audio is generally produced with regard to the audience expectations to be entertained. Thus, sports events are produced to be adequate for the electronic media and consumer viewing habits. The expectations and experience of an on-site audience and a TV audience differ a lot. For example, the sound of the swish of the net in basketball is the expected sound on television even though it is unlikely anyone in the arena hears that sound. Similarly, the TV viewer is able to look at the action from many different perspectives and can even choose between different points of view (POV) with the help of modern second screen technologies. In parallel to cinematic sound design, sports events are made larger than the real event itself for TV.

The significance of TV and sports is owed to the fact that large sporting events such as World Cup and The Olympics are a driving force for the advancement of all technologies and operations including timing, and scoring, drug testing, security and particularly picture and sound. New media technologies often experience their premiere in front of the largest global audience in sports TV, because sports like soccer or American football often have high viewing rates. The London Games 2012, for example, were partly produced with 8k video and 22.2 sound [4].

## 1.3. The Role of Sound in Sports TV

To catch the overall feeling and excitement of the live event is most important about capturing the sound for sports TV. A natural and realistic representation of the original sound scene is desired, but it is also essential to produce a higher level of experience for electronic media.

Sound for sports TV tries to keep up with the extraordinary performances of the popstar-like athletes by using dramatic sound design as used in the cinema or video games. The aim is to give the viewer an illusion of reality instead of using a neutral documentary approach. For this reason, electronic media need to deliver a bigger than life experience to transfer the feeling into the living room.

Viewing habits of the audiences have been influenced by movies and games, but the production situation as well as the reproduction circumstances are very different for sports TV. At the event, everything is produced in real time and there are only few possibilities to do any kind of post production. On the reproduction side, the common TV set is not as good in terms of audio quality as in

cinema, where people only watch and listen and where there is no distraction as there is in the homes.

The sound of sports television can be a marketing differentiator between competing broadcasters as well as a merchandising tool for the fan experience. People are willing to pay more money for premium tickets in the stadium in order to be able to be closer to the action. With pay TV, customers can see more if they pay more. Why should a dedicated fan not pay extra money to get premium sound objects such as the personal microphone signal of their favourite athlete or compelling 3D sound?

## 1.4. The Role of the Sound Designer

With MPEG-H, the creative possibilities of sound for sports extend with the use of audio objects and immersive audio. The sound engineer has always been both, a technician who makes sure all signals are recorded and broadcasted and a creative sound director, whose task it is to deliver a realistic interpretation of the recorded signals to the listener. With each event and sport the immersive experience varies with the imagination and subjective interpretation of the sound designer, as is also the case with the video director.

Creating the immersive experience begins with identifying the characteristic sound elements that identify the event and the given sport in general – the acoustic fingerprint of the location, such as crowd and sport sounds. The next thing to do is to find the most important sound elements to articulate the action and the dramatic story of the athlete, for example the kicking sound of the soccer ball or the clicking sound on a snooker table.

The broadcasting of today offers HD picture and surround sound. The sound capabilities will soon be extended to MPEG-H. To use the new features for more audiences entertainment and excitement, a close cooperation between the video director and sound designer is essential.

## 1.5. Sound and Picture

Next generation High Definition (HD) picture will most likely be 4K (Super High Definition), but like its predecessor – HD1080 and standard definition (SD), as well as the iPad and every other current visual platform, these devices are only two dimensional and it is the sound that gives them the illusion of dimension.

The sound has to complement the picture and glue together the individual camera perspectives so that the story line continues. There is a tremendous amount of creative latitude for this task. Sound design can vary from the essential noises needed to convey the experi-

ence to an elaborate soundtrack that delivers an experience beyond the visuals. The audio not only underscores the picture but steers the attention and imagination of the viewer through the landscape.

When 3D-Audio comes into play, the additional height dimension can be used to accentuate a close video perspective by bringing the sound of audiences, e.g. clapping and cheering, from the upper layer. When switching back to a wide shot, the audience would move further down and to the front. This stylistic device makes closely captured sports like table tennis – the number one sport in Asia – more dramatic and transfers the athletes' tension to the living room.

The fairly new audio elements, the change between different acoustical perspectives, and the acoustical description of movement in the horizontal and vertical plane need to be supported by the video cuts to reach their full potential. Video editors could, for example, allow for longer shots, showing the complete movement of an athlete. Directors and video editors could possibly cooperate closer with audio experts to allow for video cuts that support these new possibilities in audio.

## 2. CLASSIFICATION OF SPORTS

To get an understanding of capturing objects or 3D sound, most relevant sports can be divided into several classes. Within these classes, capture techniques are similar. The following classification is limited to some of the most important sports. There are lots of other relevant sports that can not be covered in this paper.

Most relevant for sports TV are field sports, which can be split into outdoor field sports, such as European football, American football or baseball, even though some of the used stadiums may feature a roof. Indoor field sports can be basketball, ice hockey or net sports such as volleyball. For indoor field sports the acoustic focus is on the sound of the ball or the squeaking of the shoes. The crowd sound may be different, to outdoor sports, because the audience usually is closer to the field.

Another sports class is motorsport which includes car races, motorcycle races, air races or boat races. Here, the most dominant sound object is the engine of the vehicle. The sound of the audience and the location is often less important and hard to capture due to the high sound pressure levels of the engines. Sound objects that make these sports even more exciting can be multiple commentators, on-board sound or team radio.

Wintersports cover a wide range of different sports like skating or curling. The athlete with their sports equip-

ment in the snow or on the ice gives a unique sound. Outdoor winter sports often are practised on large areas, where the cross-country ski track is prepared or the ski jumping ramp is installed. The starting and finishing area is crowded with many fans, whereas there are zones in-between without any or with only a few spectators. For outdoor winter sports, the ambience and audience sound differ a lot. Therefore, an additional ambience track needs to be captured. Because it may be difficult to place microphones for the athletes action at the desired spot, sometimes sound samples are used, to give a natural and realistic feeling.

Indoor winter sports like ice hockey or figure skating take place in stadium-like arenas with thousands of fans in the stands. Here, the ambience consists mainly of crowd sounds and sounds of the athletes are hard to catch.

For sports like tennis, table tennis or snooker the athletes are moving constantly in a relatively small confined field-of-play. The atmosphere can be very dynamic from passages of loud audience reactions and quiet moments during the plays that are rigorously enforced by an umpire. Because of its controlled acoustics and fairly small dimensions, microphones can be placed precisely to catch the personal reactions of the players, such as breathing, grunting and body movements. The performance of the athlete can be put forward and dynamic variations between wide shots with the audience in the background and close shots of the athlete can be supported by the sound to a great extent.

## 3. SPECIFIC SPORTS: CHARACTERISTICS, CAPTURING AND MIXING

Every sport and venue has common sound elements: atmosphere, ambience, sports specific sounds, commentary, music and special effects (s. Tab.: 1). This list of sound items directly resolves into possible sound objects for each sport. Key questions for each single sport event are: What objects are offered by the sport that could benefit the TV user? How can these objects be captured and processed? How can spatial audio be captured at the location?

Every sports event features a unique sound, even each location for the same kind of sport may sound different. The general atmosphere must be transferred to the living room using main microphones or a mix of different microphone positions. To capture that acoustic fingerprint of the location, microphone techniques strongly depend on the type of sport. The nature of each sport and activity will dictate the important sound elements that contribute

Sound Object	Example	Source
atmosphere	general location sounds, wind, helicopter sounds	ambience mics
audience	cheering, applause	spaced mics, mic arrays
sports sounds	sounds of athletes action (kick of the ball, shouting, car engine)	spot mics, radio mics
commentary	TV comment (several languages), expert, interview, stadium announcer	personal mics
music	background music on location	local PA system
electronic effects	audio-visual effects of event, sponsor or producer	playback

**Table 1:** List of most common sound objects for sports TV and examples for each.

to the immersive and the interactive experience. An analysis of former productions of a specific sport can show what the usual editing style for the image is and how that can be extended to objects and 3D sound.

### 3.1. 3D-Audio Microphone Capture

The physical characteristics of a venue, its shape, size and roof status – indoor or outdoor, will have a profound effect on the acoustics and microphone capture in the venue. The acoustics and background noise will certainly impact the audio clarity of the capture.

Similar to surround recordings, there are different methods of placing and spacing microphones for the capture of audience sound for 3D-Audio (s. Tab.: 2). Depending on the size of the crowd, single or multiple spaced pairs of microphones at different acoustic zones can be used to capture the atmosphere and to make a surrounding ambience that features texture and details [5]. In order to capture signals for the height speakers, the regular middle layer microphones are complemented by microphones, that are mounted approx. 1 m or more above. The characteristic and direction of the upper layer microphones will have a big impact on the produced sonic space.

More compact setups were used for ambience. These arrays were mounted on one stand with adjustable distances between the microphones. As these arrays are very easy to move, they let the sound engineer find the best sound spot or the opportunity to move away from interference, such as PA speakers or other unwanted noise. As in every recording, spot microphones are important to give the details of the sport for the final mix.

### 3.2. Football – North American and European

Many sports like Football are presented from the ‘fan in the stand’ experience with all the sports sound up front and the viewer is surrounded in the venue ambience and atmosphere. This is definitely a natural and realistic rep-

resentation of the scene that gives good orientation to the viewer because it is a perspective, the director always comes back to, even though it is not necessarily inspiring. Having that as a basis, sports specific perspectives are used to reproduce the action that happens on the field. The POV of the athlete and close shots are used when the gameplay gets more exciting, for example for a free kick. The most important sound element is the atmosphere of the fan crowd that is in close relation to the action on the field. The sound of kicking the ball as well as the shouting of the players and coaches make up the sounds that directly derive from the sports action. Sound that is created by the athletes and coaches on the field can be captured for detailed transmission with microphones close to the action or even wireless microphones at the players – as it is done for the NFL – or it can be considered less important, as is the case with FIFA events. A stadium announcer is audible to the fans in the stadium as well as to the TV consumers.

3D-Audio would benefit these sports by separating the sports effects from the atmosphere into the upper sonic space. That de-masks and gives definition and detail to each element of the mix. Ambience and atmosphere completely surround the listener like it does the athlete on the field. With little visual support, the sound clearly delivers auditory clues about the spatial volume of the venue, the size of the audience and their mood. This design would provide an extended horizontal experience, covering the complete viewing area in the front and thus gives more sonic room to produce the detail of the sports sounds.

The use of the surround and height channels is subjective to the venue and mixer but clearly contributes to the immersive experience. The sound designer influences the balance between the sound of the stadium and the sound of the athlete.

Sound Object	Example	Source
ambience	wind, helicopters	spaced mics, pointed upwards
audience	crowd noise	spaced mics, pointed upwards
vertical motion	jumping or flying sounds	spot mics, fixed panning
effetscs	special panning effects	dynamic or automated panning

**Table 2:** List of 3D-Audio objects and examples for each.

The use of objects for the sound of sports TV would open many interesting applications for the user. For the outdoor field sports, characteristic sound elements are the crowd, the field sound and the commentary. Microphones at the sides of the field capture the kicking sound and the players shouting. The coach and the substitute bench can be recorded with boundary layer microphones or microphones on small stands. Microphones on the athlete have drastically changed the TV experience for US NFL fans because it is possible to provide much more insight into the teams' strategy. Commentary can be recorded and treated as an object fairly easily because each commentator speaks in an acoustically moderate environment and has his or her own microphone. The crowd sound usually is created with a mix of several spaced microphones or coincident systems. A challenging task can be to position the crowd microphones without catching too much of the PA sound. When these sound elements are transmitted as objects, the user can boost the crowd sound of their favourite team, have the signal of the coaches microphone at one side all the time or raise the level of the commentary when it is hard to understand (s. Tab.: 3).

### 3.3. Bobsled and Motor Sports

Although bobsleds are not driven by an engine they can be considered as a kind of motor sport, because the nature of a fast moving vehicle with the athletes inside is similar. Bobsled or motor sports have a completely different character than field sports. Usually, the cars on the track are in the video. Audiences are only seen in total shots, fast camera panning or when an athlete is in the finishing area.

Commonly, the sled is shown at the starting area with the audience in the background. A full and surrounding ambience can give an immersive feeling of the event, using 3D sound. During the race, almost no elevated sound sources can be recorded, because there is not such a big audience along the track. In steep turns, the sled

Sound Object	Signal Source
Stadium Crowd	spaced mics or mic array
Fans Team A/B	spaced mics or mic array
Commentary, USA	com. mics
Commentary, GER	com mic
Commentary ESP	com mic
Expert or interview partner	com mic
Kick Sound	spot mics
Stadium Announcer	com mic
Bench Team A /B	spot mic, spaced mics
Coach A/B	personal mic
Player A/B	personal mic

**Table 3:** List of sound objects that are usable for football broadcasting.

sound can be panned across the upper layer to indicate to height of the curve. For special effects like fly-bys, where the camera is built into the floor of the track for the bobsled to drive over, the height channels can be used to stress that effect. Using only few 3D sound effects during the ride gives an impressive contrast, when the sled arrives at the finish and the sonic space is filled with the audience again.

A key component of the sound design for this visual is to deliver a sense of danger and speed with close shots of the athlete and vehicle in motion. The sound should be equally compact and completely surround the viewer utilizing all available speakers to steer the attention laterally and vertically according to the motion. Sonic motion uses sound to steer the attention of the viewer/listener in the desired direction and can be used effectively to enhance the illusion of speed by having a rocking effect on the viewer subconsciously. Addition-

ally the Doppler shift enhances direction, motion and speed.

If a spaced pair of microphones is placed along one side of the track, with the first microphone panned hard to the left and the other one to the right, a fast left-to-right movement can be created when a bob cart drives by. No manual editing to the audio material is needed for that effect. This technique is known as hard panning, because panning happens by placing the correlating signals hard left and right. In contrast to that, for dynamic panning, a sound object or microphone signal is panned manually through the sonic space with a panning tool.

Sound designers for movies have long used low frequencies to vibrate the seating to deliver a sensation of motion. Low frequency accentuates certain feelings and sensations in the human body and can be effectively used in the immersive sound design. In addition, high levels of the engine or the skids with its aggressive sound nature can effectively transfer the action.

Prior field tests have shown that meaningful sound objects can be commentators, PA (stadium announcer), in car sounds, pit lane sound and crowd sound [6]. Most of the signals that can be used as objects are available in the OB van. Instead of being mixed together for the final mix, they can be provided with metadata and handled as objects.

To capture 3D sound, few additional microphones are necessary to catch crowd sounds for the upper speaker channels that correlate to the ambience sounds for the middle layer.

### 3.4. Extreme Sports – Big Air and Half Pipe

In the group of modern extreme sports, Skateboard Big Air and Snowboarding Half Pipe competitions have large broadcast and streaming audiences.

Both sports feature steep downward movement with jumps and tricks, where athletes are judged on creativity, height and distance.

The visual presentation of these sports is a nice blend of wide shots mixed with shots looking up and down the jumping ramp. These picture and angles give the viewer a sense of height and size of the course. The sound design must give a feeling for dimension and must support the feeling of extreme height and fast movement.

With immersive reproduction systems as supported by MPEG-H, small details such as the athlete's sound waiting at the top of a jump or other elevated sounds such as flag flapping can establish the feeling for height by being

panned to the top speakers.

Starting from such elevated audio elements, the sound of the athlete moving down the ramp and across the screen can be emphasized by a good sound design. Detail in the effects and sonic movement following the actual position of the athlete give a sense of motion and speed to the viewer. This type of sound design demands dynamic movement from the audio object to support the vision. The athlete's sound and his speed could be emphasized even more with wireless microphones on the athletes.

Similar to most other sports, the ambience and audience is a constant source for a 3D soundscape at extreme sports events. It informs the viewer about the nature of the place and about the size and reactions of the audience.

Out of these scene elements, the athlete's sound, the audience and added effects can be defined as objects. Furthermore, at least two commentator signals, the network commentator and the PA commentator, are usually present at these events. As shown in field tests, these signals can be gained very easily [Stenzel2014].

## 4. SUMMARY

The new TV standard MPEG-H offers innovative features, such as interactive audio objects, HOA and 3D-Audio which add significant marketing value to sports TV by enhance the experience for listeners and viewers. Sports TV is rewarding material for these features, because there are multiple commentators the viewer can chose from. Additional sound objects such as emphasized sports sounds – car engines, ball kicks, athletes' microphones signals – offer exciting possibilities to personalize the TV sound.

Immersive audio opens new creative ways for sound design. Thus, the listener can be engulfed by the ambience to be closer to the action.

In this paper the authors took a closer look at sports that have different, but interesting possibilities for MPEG-H broadcasting. Television sound helps to give the viewer the illusion of reality. *The Immersive Experience* is a very powerful production tool for the innovative sound designer. All sports have potential for listener immersion. Clearly, certain sports contribute to the illusion of height and depth. Other sports that do not feature vertical movements or vertical audience distribution benefit from 3D-Audio by offering more sonic space for a dramatic sound design. Demo material that has been captured and produced demonstrates the effectiveness of these techniques.

Usually, interaction and object control will have little impact on engineering and production design, because most of the required signals are available without further effort.

At the moment, there are no boundaries or rules for the creative use of 3D-Audio or interactive sound objects.

## 5. REFERENCES

- [1] Herre, Jürgen and Hilpert, Johannes and Kuntz, Achim and Plogsties, Jan, “MPEG-H Audio - The New Standard for Universal Spatial/3D Audio Coding,” presented at the AES 137th convention, Los Angeles, USA, 2014
- [2] Fueg, Simone and Hoelzer, Andreas and Borss, Christian and Ertel, Christian and Kratschmer, Michael and Plogsties, Jan, “Design, Coding and Processing of Metadata for Object-Based Interactive Audio,” presented at the AES 137th convention, Los Angeles, USA, 2014
- [3] Scuda, Ulli “Comparison of Multichannel Surround Speaker Setups in 2D and 3D,” presented at the 2nd International Conference on Spatial Audio (ICSA), Erlangen, Germany, 2014
- [4] Sugawara, M. and Sawada, S. and Fujinuma, H. and Shishikui, Y. and Zubrzycki, J. and Weerakkody, R. and Quested, A., “Super Hi-Vision at the London 2012 Olympics”, SMPTE Motion Imaging Journal Vol. 122, 2013
- [5] Baxter, Dennis “A Practical Guide to Television Sound Engineering,” Focal Press, Burlington, 2007.
- [6] Stenzel, Hanne and Scuda, Ulli, “Producing Interactive Immersive Sound for MPEG-H: A Field Test for Sports Broadcasting,” presented at the AES 137th convention, Los Angeles, USA, 2014