



OBJECT-AUDIO CAPTURE SYSTEM FOR SPORTS BROADCAST

Dr.-Ing. Renato S. Pellegrini Dr.-
Ing. Alexander Krüger Véronique
Larcher Ph. D.

Sennheiser AMBEO, Switzerland

ABSTRACT

Object-audio workflows for traditional flat broadcasts have recently appeared after the introduction of new audio formats such as MPEG-H and ATMOS. These formats allow for the creation of object-based mixes that can be dynamically rendered at the end user depending on their reproduction hardware. Until very recently, only post-produced content was being created for these formats but new broadcast standards in the US and Asia, as well as new hardware encoding engines for live production have made live sports in these formats more feasible. These formats allow for a fuller, more immersive sound design and allow for some possibilities of personalization.

The issue then arises on how to capture live action from the field that would provide these object-audio workflows with the desired isolated sounds and accompanying metadata. Current field action capture systems provide a suboptimal amount of isolation from the crowd to highlight individual action sounds and dialog from the field. And in most cases, placing traditional microphones near the action is not possible. In this paper, we are presenting new microphone techniques and systems enabling better performance for sound capture that fulfill the needs of the future object-audio broadcast formats. This includes beamforming techniques, automatic steering, and systems management of arrays of microphones.

INTRODUCTION

There are many benefits to using audio as a primary method of delivering new experiences. During a sports event, it is the audio that best engages consumers with the atmosphere of the stadium, with the passion and excitement of a commentator, or with the tension surrounding a moment of silence. The sound mix of a sports event is a fundamental component to create a sense of presence for the viewer/listener. Not only is sound suitable to create an ambient field that reflects the atmosphere at the event, it also carries important information, such as some sense of how intense a punch was in a boxing event. In soccer, the sound of the ball being kicked or the sound of the ball hitting any object may explain the decision of a referee during game play. Questions such as “Did the player touch the ball before it went off the field?” are answered more easily by the sound of

an impact than using a camera's view from a possibly sub-optimal angle. Capturing such sounds that carry additional information complementing a camera's view is non-trivial. In this paper we are presenting a novel microphone-array dedicated to such tasks, offering vastly improved performance while being easy to operate and suitable for live broadcast. We will focus on sound capture for soccer events, although similar use-cases can be thought of for many other sports.

Sound capture & rendering in sport events

There are challenges inherent to recording of sounds at sport including:

- The sound event may take place at a considerably high distance from any microphone.
- The sounds of interest may be far lower in level than the general ambient level in a stadium.
- The objects creating the sound may be moving at high speed.
- Any processing of the captured sounds must allow for live broadcast of the audio stream and therefore involve no, or conceivably low latency.
- Systems should sustain adverse weather conditions such as rain or wind.
- Systems must withstand mechanical impact such as a ball hitting a microphone.
- Microphones shall not conceal any camera view
- The frequency spectrum of the sounds of interest ranges from low frequencies below 200 Hz up to 5 kHz and above, while typical crowd noises cover the same frequencies.
- Depending on the camera view, the sound may need to be panned and played back from a different angle.
- Some sounds may only be of interest if they add information to the visuals, while others (such as the referee's whistle) need to be heard independently of what is shown on screen.

In soccer, there are several sound sources that are of high interest to the game play, including:

- The sound of the ball and impact with other objects ("close-ball sound")
- The referee's whistle
- Comments from the trainer of each team
- The referee's speech and communication, (although it is debatable to what extent such communication shall be shared with the viewers during game play)

State-of-the-art recording techniques for a close-ball sound

Today, the typical setup of microphones includes more than a dozen microphones on the field. The zones around the two goals are of higher interest and therefore up to 3 microphones are placed directly behind each goal, while the rest of the field is covered more sparsely. In order to attenuate crowd noise sufficiently and capture clean sound for the sources of interest, highly directive microphones are used pointing at the field. Depending on their positioning, the sound timbre of a source can change considerably as it moves, due to the directivity pattern of the microphone. Thanks to visual tracking systems combined with Lawo's kick software, it is possible to automate the level of all these microphones such that microphones, which are too far from the current field of play are automatically faded out while those, closest to the sound of interest are faded in. Such systems and setups suffer of the following drawbacks:

- The achievable SNR resulting from the acoustic fixed directivity of the microphone may not be sufficient to attenuate the fan/crowd's noise sufficiently.
- As microphones are placed far away from each other, mixing more than one signal creates comb-filtering effects that adversely affect the sound timbre due to the different time of arrival of the sound waves at each microphone. Therefore, in practice only the closest microphone is typically used as a singular mono source.
- Although ball position trackers and software such as the one from Lawo are used to automate a mix at the console level, the original source position is lost in the mix, which hinders the subsequent proper panning expected for objects in the upcoming formats.

Proposed circular microphone array for sports capture

We have designed a novel circular array as shown in figures 1 & 2 of highly directive shotgun microphones. The total diameter of 1.50 m allows for proper beam steering at frequencies as low as 200 Hz and covering frequencies up to above 5 kHz with nearly flat on-axis response while ensuring an effective side and back rejection.



Figure 1
Circular mic array

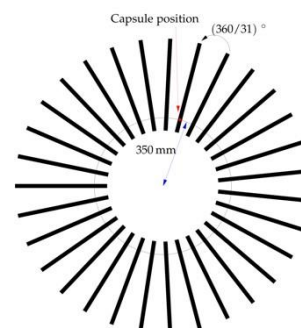


Figure 2
Array consisting of 31
MKH8070 shotguns
microphones

This circular array features a fixed narrow directivity in the vertical plane thanks to the physical directivity pattern of the shotgun microphones. This enables a very effective attenuation of any fan crowd noise above the field while focussing on sounds in the field. Figure 3 shows the original directivity pattern of a single MKH8070 microphone.

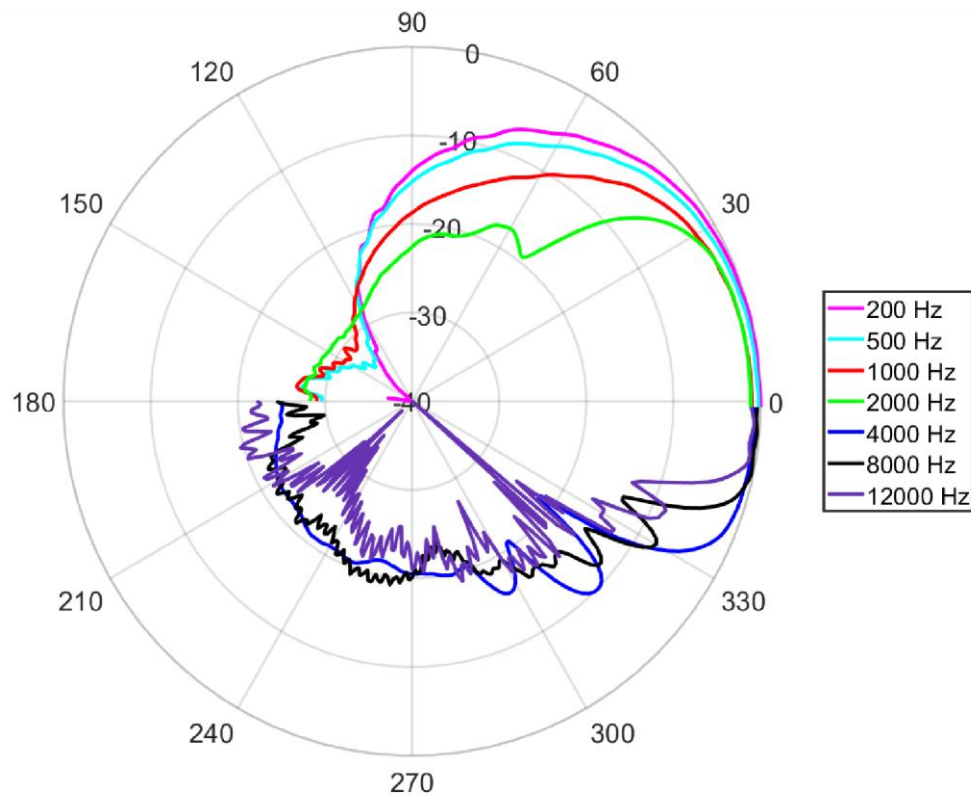


Figure 3
Directivity pattern of MKH8070

In the horizontal plane, the new circular array allows for extremely precise and narrow sound beams pointing at the source of interest. These beams outperform the directivity of all tested shotgun microphones while keeping their inherent sound timbre flat across the entire frequency spectrum of interest as shown in figure 4. Thanks to the size of the array, even low frequencies as low as 200Hz can be attenuated properly from any off-axis direction.

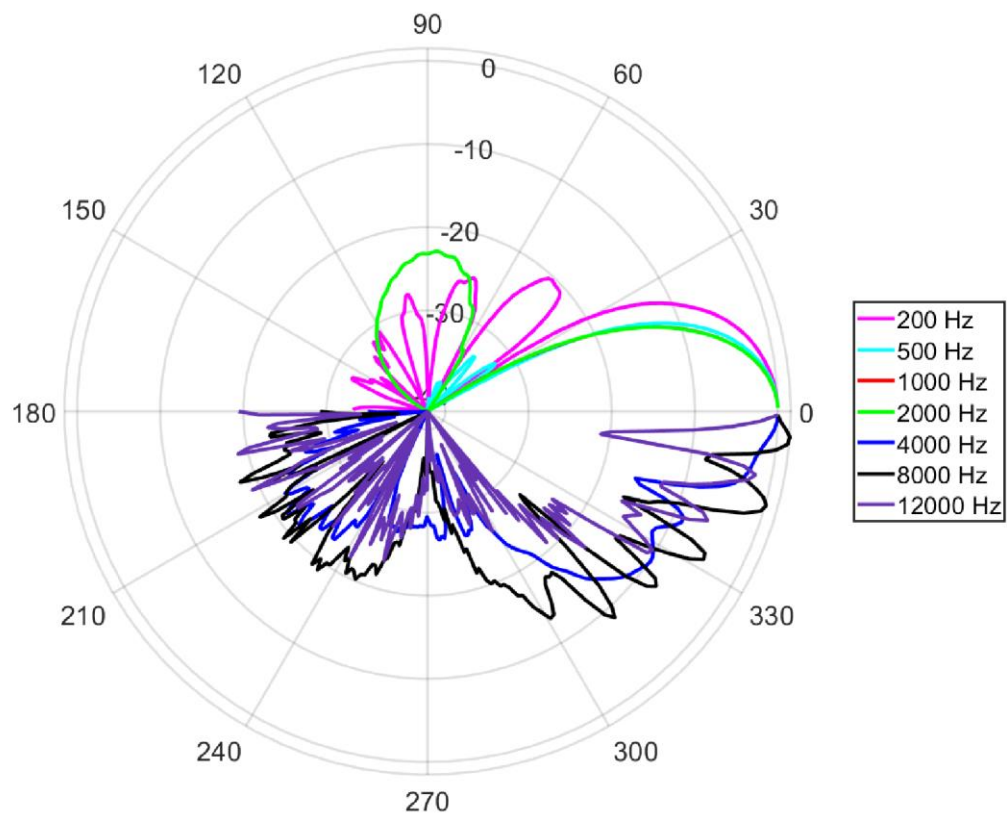


Figure 4
Horizontal beam pattern for array

Realtime beamforming using live tracking information

The proposed microphone accepts positional information of the Lawo kick software [1] to control the beam's direction with respect to the microphone-array's position on the field and its orientation towards the field. Care has been taken to ensure that latencies involved in receiving appropriate positional information are aligned with the audio captured from the field to optimize the beams target direction at any point in time. As distance information is available, the gain can be chosen so as to compensate for distance allowing for a uniform level of a close-ball sound independently from the distance between the ball and the microphone. To date, there is no standardized way to uniformly share object position and orientation information with the audio stream but future mixing and rendering hardware may provide ways to understand positional information from within the original track. Dolby has developed Dolby ED2, which enables object-audio metadata to be carried along with the audio. Dolby ED2 is one solution for delivering these new experiences through the broadcast infrastructure. Other standards are currently being prepared and established by the ITU & EBU standardization bodies to allow for "Next Generation Audio" (NGA),

including a common renderer for ADM content labelled EAR as described in [2]. Dolby is working with standards bodies and other organizations to ensure that this new metadata can be supported through a wide range of technologies and devices. The proposed microphone array is capable to share such positioning information with the subsequent mixing console and rendering hardware. Using this metadata, a full workflow for object-oriented mixing such as in Dolby Atmos for live, or Immersive Audio according to ITU-R BS.2088-0 becomes feasible.

Rendering and processing of objects

In today's mixing practices, target sounds such as a close-ball capture or the referee's whistle are not spatialized but added in mono to the mix neglecting their spatial positioning with respect to the image seen. Future VR applications will benefit from properly positioning and spatializing such sounds as objects fitting with the image seen, because audio-visual congruence is essential to the sense of presence and immersion. However, this step required knowing the camera's orientation and involves a complete spatial mapping of the site. Broadcasters such as Sky as well as BT Sport have started producing and broadcasting their British Premier League and Champions League productions in Dolby Atmos since January 2017. The 2018 Olympic Winter Games were among the first ones to be produced with Dolby Atmos by Comcast and DirectTV, although both of them were only offering the content in a non-live fashion. DirectTV was using their 4k channel that has a time delay and Comcast Xfinity offered its productions on demand.

So far, the mixes have not used the full potential of object-oriented sources but focused on another specific feature of Dolby Atmos allowing for increased immersion by adding height loudspeakers. As the technical infrastructure for live productions improves, the production studios will start to exploit the full potential of the new object-oriented mixing formats. The microphone array introduced in this paper is an important step towards a seamless workflow using clean signals for audio objects and their positional real-time information.

Using the microphone array in production

The microphone array is interfaced through the processing unit with the mixing console and receives tracking information through the console. This allows the mixing engineer to access the trackers positional data through Lawo's kick software and make adjustments to the target array beams. In general, various target beams may be processed simultaneously, each of them with a different target beam steering direction. An overview is given in figure 5 below.

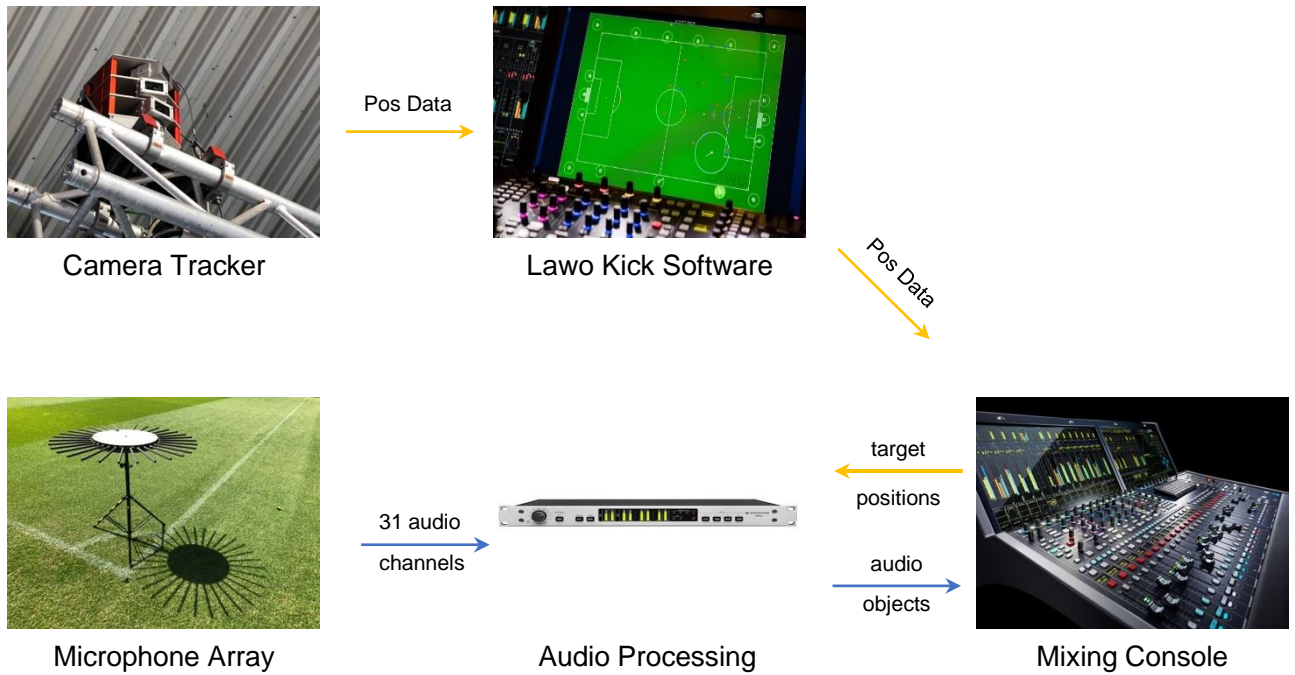


Figure 5 – Component block diagram

For the use case of a soccer game live broadcast we suggest 4 microphone arrays to be placed on the long side-lines of the field as shown in Figure 6.

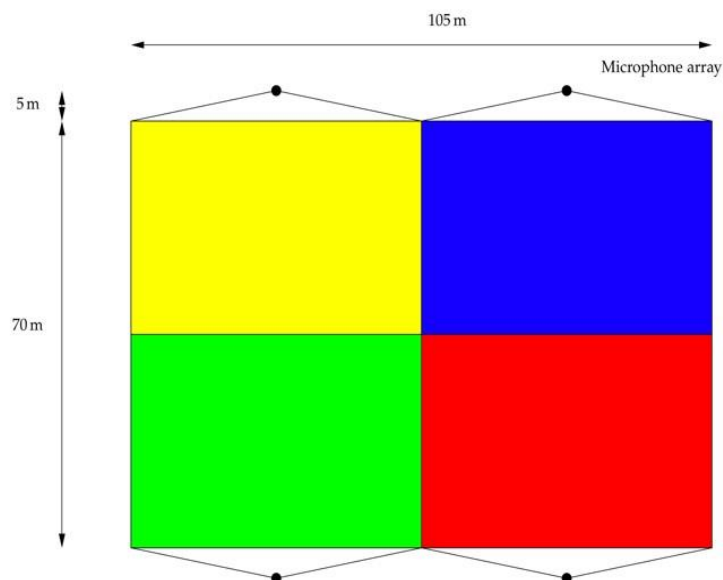


Figure 6

Proposed array placement around soccer field

In this configuration, the maximum distance that the array needs to cover is approximately 50m in this configuration. Alternatively, one could mount two arrays behind each goal and two more on the middle of the long sides which may interfere with moving cameras in some stadiums. Although the distance to the sources of interest remains challenging, the setup as described in Figure 6 is more practical as there is usually more space available on the side lines to place the microphones.

A good positioning seems to be within the video-add walls as there usually is an open space between the front and the back wall as illustrated in Figure 7.



Figure 7

Proposed array placement

Array processing and performance

The internal processing of the array is using a modified modal beamformer such as described in [3]. Such beamforming will result in nearly flat frequency responses on axis independent of the target direction. Thanks to the circular shape of the array, there are no side effects coming from the array ends, as are known to deteriorate performance of linear arrays. Even following a moving sound source will result in very limited disturbing audible artefacts. Further, all filtering is done statically and changing the target beam's direction only involves simple gaining and summing of each filters output. The resulting processing is therefore very efficient and can deliver several target beams simultaneously covering not only one target sound such as a close-ball, but simultaneously can record other objects such as the referee's whistling or some conversation, while keeping background noise out of the recorded signal with high efficiency. At the time of writing we only had simulated data to proof our concepts but during IBC 2018 we are happy to present our findings in direct comparison to today's best practices.

The general processing for one target is shown in Figure 8.

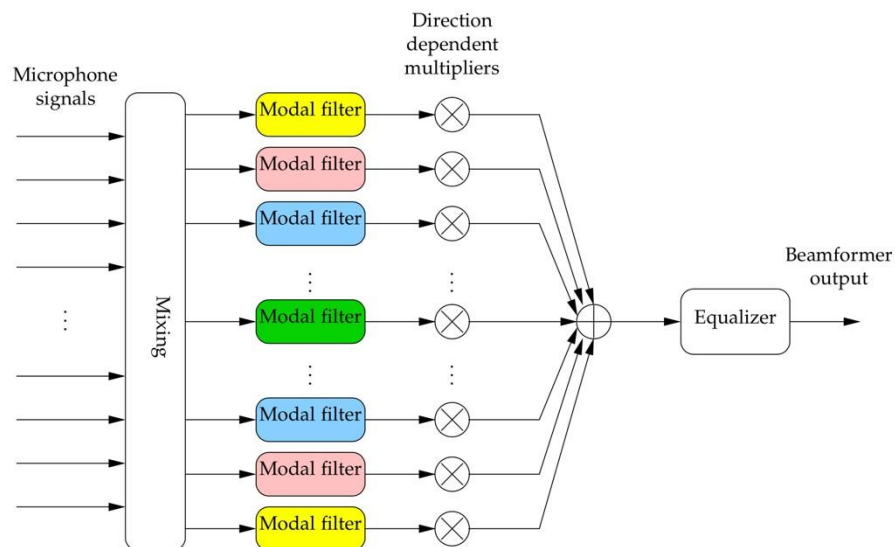


Figure 8
Beamforming processing

Choice of positions and number of microphones

In general, the number of microphones determines the resolution of the achievable target beam-pattern. In particular, it controls the maximum directivity index, which is a ratio between the beam-former output power with respect to the target direction and the output power integrated over all other directions.

In context with modal beam-forming, it does especially make sense to choose the number of microphones Q in dependence of the required maximum degree of the target beampattern M as $Q = 2M + 1$. In order to favourably condition the discrete Circular Harmonics transform, it is recommended to use a uniform distribution of microphones on the circle. This ensures a uniform beam-former performance over all directions, as is intended with modal beam-forming.

However, for certain applications it might be also reasonable to arrange the microphones differently, e.g. on a circle segment with a certain opening angle. The disadvantage of using circle segments compared to full circles is that in case of steering to a direction close to the segment edge, the attenuation of the incident sound for directions where microphones are missing usually cannot be accomplished well enough. This problem might be partly compensated for by extending the circle segment sufficiently over the expected operating angle range. For instance, if the microphone array is placed at a corner of a soccer stadium, a potential operating angle range would be 90° , such that a semi-circular array might be appropriate.

Further, for such circle segment array configuration other algorithms than modal beam-forming can usually achieve a better performance because they do not rely on the full

circular arrangement. However, the disadvantage of using such alternative algorithms is that they typically require variable filters instead of fixed filters and variable pure gains, and these filters vary with the target direction leading to additional computational load. Further, since the design of these filters is in most cases computationally demanding, it preferably done offline instead of in real-time. As a result, an efficient approach is to precompute and store the filters for a discrete set of directions beforehand, and to select suitable filters from the stored database depending on the target direction at runtime.

Conclusion & Outlook

A new circular microphone array has been proposed that outperforms existing recording approaches and simplifies the workflow for upcoming object-oriented mixing approaches. Given the outcome of early tests and simulations we are expecting array technologies to play an important role in sports broadcast as the object-oriented formats are being deployed in the broadcast industry. Position of the microphones and of the target sound objects is an information that object capture systems need to acquire, and this will be an area where we expect continued improvements. The resulting greater flexibility and ease to operate of such object audio capture systems make their underlying principles applicable to a variety of sports situations, each of them potentially dictating adjustments in the array configuration.

References

- [1] LAWO Kick product presentation. Available at <https://www.lawo.com/en/products/audio-production-tools/kick.html>.
- [2] EBU Operating Eurovision and Euroradio, Tech 3388, *ADM renderer for use in next generation audio broadcasting, 2018*
- [3] H. Teutsch. *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition*. Lecture Notes in Control and Information Sciences. Springer Berlin Heidelberg, 2007.