

ORTF-3D: Ambience recording for 3D Audio & VR

Development and application of a stereophonic multichannel recording technique for 3D Audio & VR

Helmut Wittek, Schoeps Mikrofone, 2016



Fig. 1: ORTF-3D arrangement, in a windscreen with the cover removed

Contents

One recording method for all 3D formats?	2	Recording engineers who work with 3D sound face a difficult task when choosing a suitable recording technique. The number of channels is greater than with playback systems that operate only in the horizontal plane, so the complexity increases as well.
Is first-order Ambisonics adequate for 3D?	2	
Criteria for stereophonic arrays	3	
The ORTF-3D recording method	4	When a customer demands 3D audio rather than conventional 5.1 surround it may be tempting to apply solutions that are overly simple. But when a 3D recording has been made well, using a suitable recording technique, the advantages are impressively audible.
Translating theory into practice	5	
Conversion for Dolby Atmos and Auro3D	5	
Conversion for VR	6	

ORTF-3D: Ambience recording for 3D Audio & VR

One recording method for all 3D formats?

There are various 3D audio playback systems, so the recording techniques that work best for each of them will naturally be different. For soundfield synthesis systems, multi-channel microphone arrays can be a solution, while for 3D stereo, stereophonic miking techniques are the norm. For binaural reproduction in the simplest case, a dummy head can be used.

But all these systems share one requirement when recording complex, sustained sound sources such as ambient sound: Stereophonic techniques must be used, because they alone offer both high quality sound and high channel efficiency. It is impossible or inefficient to reproduce in high quality the sound of a large chorus, for example, or the complex, ambient sound of a city street, by compiling single point sources recorded with separate microphones.

In the same way, multi-channel microphone arrays for soundfield synthesis, such as higher-order Ambisonics („HOA“) or wavefield synthesis, fall short in practice because their channel efficiency and sonic quality are too low. If on the other hand the number of channels is reduced, e.g. with first-order Ambisonics, the spatial quality becomes burdened with compromise.

For binaural playback, the dummy head technique is clearly the simplest solution—but it does not, in itself, produce results compatible with virtual reality glasses, in which the binaural signals must respond to user's head motions. This is possible only through the “binauralization” of a stereophonic array—a technique that is already well established in audio for games.

Is first-order Ambisonics adequate for 3D?

There is a common assumption that Ambisonics would be the method of choice for 3D and VR. The professional recording engineer would do well to examine the situation more closely.

Ambisonics, which has existed for a long time by now, is a technology for representing and reproducing the sound field at a given point. But just as with wavefield synthesis, it functions only at a certain spatial resolution or “order”. For this reason, we generally distinguish today between “first-order” Ambisonics and “higher-order” Ambisonics (“HOA“).

3D AUDIO

The new approaches included in „3D Audio“ reproduce sound from all spatial directions. This includes the Dolby Atmos and Auro3D stereophonic systems; binaural / virtual reality („VR“) systems; and sound-field synthesis approaches such as Ambisonics and wavefield synthesis systems.

3D Audio can give distinctly better spatial perceptions than 5.1. Not only is the elevation of sound sources reproduced, but noticeable improvements can also be achieved with regard to envelopment, naturalness, and accuracy of tone color. The listening area can also be greater; listeners can move more freely within the playback room without hearing the image collapse into the nearest loudspeaker.

First-order Ambisonics cannot achieve error-free audio reproduction, since the mathematics on which it is based are valid only for a listening space the size of a tennis ball. Thus the laws of stereophony apply here—a microphone for first-order Ambisonics is nothing other than a coincident microphone with the well-known advantages (simplicity; small number of recording channels; flexibility) and disadvantages (very wide, imprecise phantom sound sources; deficient spatial quality) of that approach in general.

Creation of a Ambisonics studio microphone with high spatial resolution is an unsolved problem so far. Existing Ambisonics studio microphones are all first-order, so their resolution is just adequate for 5.1 surround but too low for 3D audio. This becomes evident in their low inter-channel signal separation as well as the insufficient quality of their reproduced spatiality.

The original first-order Ambisonics microphone was the Soundfield microphone, built the same way as for example the Tetramic or the new Sennheiser VR microphone. The Schoeps “Double M/S System” works in similar fashion, but without the height channel.

Ambisonics is very well suited as a storage format for all kinds of spatial signals, but again, only if the order is high enough. A storage format with only four channels (first-order Ambisonics calls them W, X, Y, Z) makes a soup out of any 3D recording, since the mixdown to four channels destroys the signal separation of the 3D setup.

Ambisonics offers a simple, flexible storage and recording format for interactive 360° videos, e.g. on YouTube. In order to rotate the perspective, only the values of the Ambisonics variables need be adjusted. Together with the previously mentioned small first-order Ambisonics microphones, 360° videos are very easily made using small, portable cameras.

ORTF-3D: Ambience recording for 3D Audio & VR

As a result, most of the advantages of first-order Ambisonics do not come into play in VR. On the contrary, its disadvantages (poor spatial quality, crosstalk among virtual loudspeaker signals) only become more prominent.

If practical conditions allow for a slightly larger microphone arrangement, an ORTF-3D setup would be optimal instead as an ambience microphone for VR.

Criteria for stereophonic arrays

Stereophonic arrays are thus the approach of choice for all 3D formats. The requirements for 3D are the same as in two- and five-channel stereophony:

- Signal separation among all channels in order to avoid comb filtering: No one signal should be present at significant levels in more than two channels.
- Level and/or arrival time differences between adjacent channels to achieve the desired imaging characteristics.
- Decorrelation of diffuse-field sound for optimal envelopment and sound quality.

2-channel stereophony

These demands are still easy to fulfill in two-channel stereophony; a suitable arrangement of two microphones and two independent channels can provide the desired imaging curve. Tools such as the Imaging Assistant application (available as an iOS app or on the Web at www.ima.schoeps.de) have been developed for this purpose. They take into account not only the creation of phantom image sources, but also the ever-important channel decorrelation.



Fig. 2: Two-channel ORTF system in a suspension designed for use within a windscreen; two cardioids, 17 cm, 110°.

A classic, positive example is the ORTF technique, which has a 100° recording angle and delivers a stereo signal with good channel decorrelation.

5-channel stereophony

The above requirements are distinctly more difficult to meet with five channels, and there are numerous geometries that fail to meet them, e.g. a microphone that looks like an egg the size of a rugby ball, with five omni capsules that can deliver only a mono signal at low frequencies.

Five independent channels simply cannot be obtained with any coincident arrangement of first-order microphones. A coincident arrangement such as first-order Ambisonics is thus a compromise for 5.1, though highly workable because of its advantages in compactness and post-production flexibility.

One optimal solution for ambient recordings in multi-channel stereophony is the "ORTF surround" system, in which four supercardioids are arranged in a rectangle with 10 x 20 cm side lengths. Here the distances between microphones help with decorrelation, and thereby lend the sonic impression its spatial openness. The microphone signals are routed discretely to the L, R, LS and RS channels. The signal separation in terms of level is ca. 10 dB; thus the sonic image during playback is stable even in off-axis listening positions



Fig. 3: Four-channel "ORTF Surround" system; four supercardioids, 10 / 20 cm spacing, 80° / 100° angles.

ORTF-3D: Ambience recording for 3D Audio & VR

8 or more channels

With eight or nine channels, the arrangement of the microphones becomes very difficult if the above-mentioned requirements are to be met. The simplest method for maintaining signal separation is to set up eight or nine microphones far apart from one another. Thus a large nine-channel “Decca Tree” arrangement is very well suited for certain applications, although it has severe disadvantages that limit its practical usability. For one, the sheer size of the arrangement is greater than 2 meters in width and height. And the signal separation in terms of level difference is nearly zero; every signal is more or less available in all loudspeakers. Thus this array can represent a beautiful, diffuse spaciousness, but stable directional reproduction isn’t achieved beyond the “sweet spot.” This can be helped by adding spot microphones.

The ORTF-3D recording method

An optimal ambience arrangement for eight channels is offered by the new “ORTF-3D” system (developed by Helmut Wittek and Günter Theile). It is more or less a doubling of the “ORTF Surround” system onto two planes, *i.e.* there are four supercardioids on each level (upper and lower), forming rectangles with 10 and 20 cm side lengths. The two “ORTF Surround” arrangements are placed directly on top of one another, see Fig. 6.

The microphones are furthermore tilted upward or downward in order to create signal separation in the vertical plane, see Fig. 5. Thus an 8-channel arrangement is formed, with imaging in the horizontal plane that somewhat corresponds to the “ORTF Surround” system. The microphone signals are discretely routed to four channels for the lower level (L, R, LS, RS), and four for the upper level (Lh, Rh, LSh and RSh). In VR applications, virtual loudspeaker positions forming an equal-sided cube are binauralized.



Fig. 4: A prototype of the ORTF-3D system at the ICSA conference in 2015. Eight supercardioids, horizontal distance 20 cm, vertical distance 0, angle 90°.

Imaging in the vertical dimension is produced by angling the microphones into 90-degree *XY* pairs of supercardioids. Such a two-channel coincident arrangement is possible due to the high directivity of the supercardioids, and the imaging quality and diffuse-field decorrelation are both good.

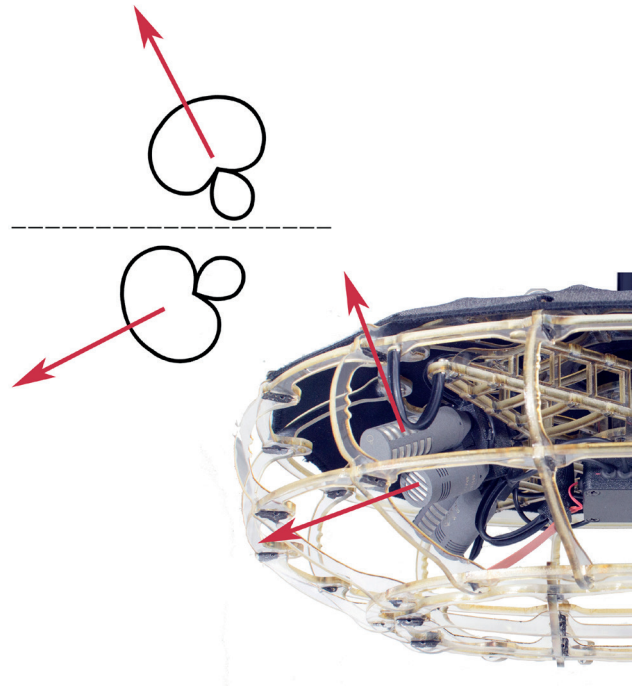


Fig. 5: Orientation of the capsules: one vertical *XY* microphone pair for each vertical pair of loudspeakers

This results in an eight-channel array with high signal separation, optimal diffuse-field correlation, and high stability within the listening area. All requirements are optimally fulfilled, yet the array is no larger than the compact ORTF Surround system—a decisive practical advantage. Numerous test recordings have shown that the ORTF-3D approach produces very beautiful, spatially open and stable 3D recordings.

ORTF-3D: Ambience recording for 3D Audio & VR

Translating theory into practice

For the SCHOEPS ORTF-3D Outdoor Set, eight compact supercardioid CCM studio microphones are used. All microphones, as well as the windscreen itself, are elastically suspended in order to decouple vibrations. Each vertical X/Y pair is composed of one front-addressed CCM 41 and one radially-addressed CCM 41 V. This enables a space-saving parallel arrangement of the microphone housings.

The windscreen and suspension have been developed by Schoeps together with the company Cinela. As with the "ORTF Surround" windscreen, elastic suspensions are also available for the ORTF-3D windscreen; fur, optional rain protection, multicore cables with breakout cables and integrated heating are standard. The windscreen is designed to be mounted by hanging. Thus long-lasting outdoor installations, e.g. from the roof of a stadium, are possible.

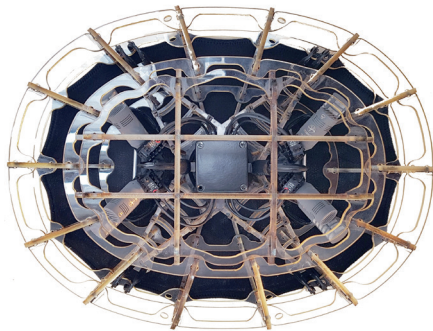


Fig. 6: ORTF-3D windscreen with the cover removed; view from below

This microphone arrangement, which was initially introduced as a prototype at the end of 2015, has already been sold or rented in considerable numbers to customers in the sports and VR sectors. Tests have been made with great success during the past two years, including several well-known sporting events

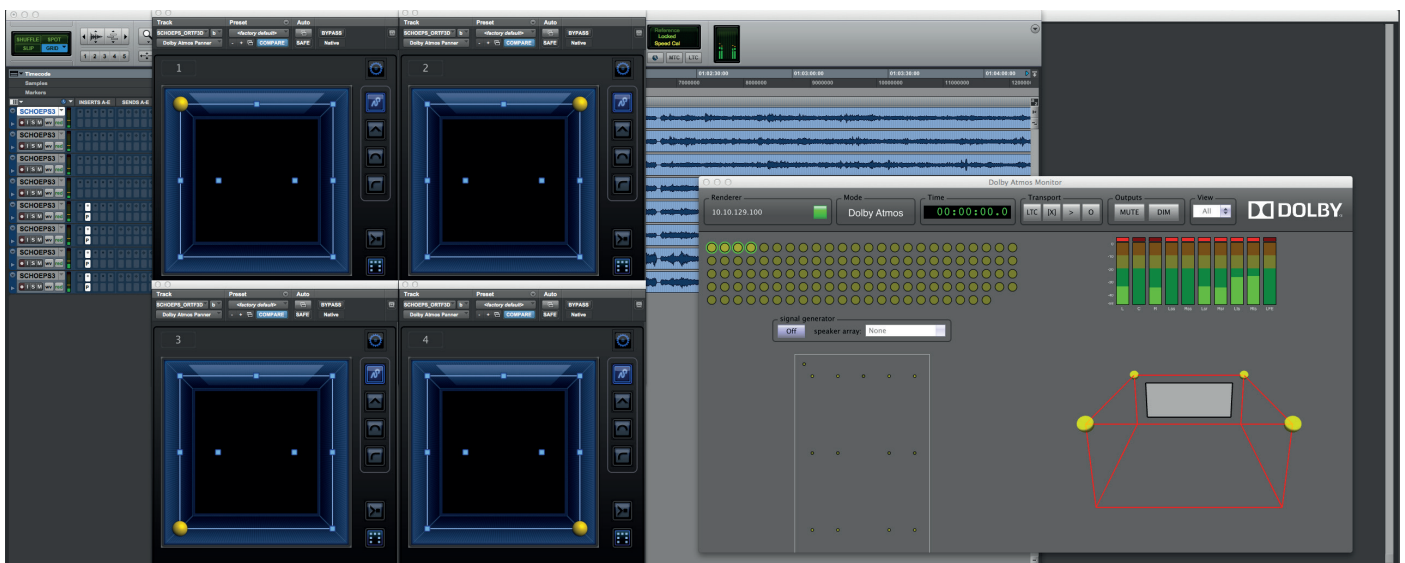


Fig. 7: Routing of the eight channels from the ORTF-3D in Dolby Atmos (ProTools plugin)

Conversion for Dolby Atmos and Auro3D

The eight channels of the ORTF-3D are L, R, LS, RS for the lower level, and Lh, Rh, LSh and RSh for the upper level. They are routed to eight discrete playback channels without matrixing.

The center channel remains unoccupied. A center channel is seldom desired in ambience recording; it would distort the energy balance between front and rear, and require significantly greater distances among microphones in order to maintain the necessary signal separation. If a center signal should be necessary for a specific reason, e.g. to cover the shut-off of a reporter's microphone, a simple downmix of the L and R signals at low level is sufficient.

In Auro3D the loudspeaker channels L, R, LS, RS, HL, HR, HLS and HRS are fed.

With Dolby, the integration in the Atmos production environment is equally simple; the channels L, R, LS, RS are simply laid down in the corresponding channels of the surround level, the so-called "Atmos bed," whereas the four upper channels are placed as static objects in the four upper corners of the Cartesian space in the Atmos panning tool. These are then rendered in playback through the corresponding front or rear loudspeakers.

The below screen capture from ProTools, with the four Atmos panners as well as the monitoring application, illustrates this.

ORTF-3D: Ambience recording for 3D Audio & VR

Conversion for VR

In a virtual reality ("VR") environment, 3D video and binaural sound are reproduced via VR glasses with headphones. Head position and rotation are processed in real time. 360° videos can also contain binaural sound, but only head rotation is processed, not the head position.



Fig. 8: VR glasses (Samsung)

If binaural sound is to respond to head tracking, a dummy head cannot be used as the recording method since it allows only for one head angle. Instead, the following sound components are gathered separately and assembled:

- "Audio object" with dry sound
- Binaural filters: "HRTF" (with room: "BRIR")

Usually the audio object e.g. a character in a VR video game, is a single source with a certain distance and 3D direction. It consists of dry sound, which is then processed via binaural and room filters (= "binauralized") depending on its 3D direction. This direction is determined by the position of the audio object and the position and head rotation of the listener within the VR scene.

The acoustical background signal of a scene, or "ambience/atmo", is a very special kind of audio source. It cannot be recorded dry, nor can it be mapped to a single point source. In principle it could be produced by the superposition of numerous audio sources in space, but often this would either be inefficient (e.g. trees in a forest) or impossible (live ambience from a venue).

Thus a group of several audio objects forming an array of virtual loudspeakers is used to reproduce a stereophonic recording of the ambience. These group of loudspeakers can be chosen from a 3D preset, for example the Dolby setup 5.1.4, or the Auro3D setup 9.1, in each case without a

center loudspeaker. If no preset is available, one can define an equal-sided cube around the listener.

These audio objects are "diegetic", i.e. they do - exactly as their visual counterparts - not move in response to head rotation. This does imply that their incidence angle in relation to the head changes with head rotations and thus the HRTFs change. The eight signals of the ORTF-3D microphone are utilized in this way to build up an optimal 3D live ambience in the VR environment.

The use of a first-order Ambisonic microphone for this purpose cannot be recommended as described above. Being a small, coincident setup, its output lacks sufficient separation among channels, thus reducing the quality of its spatiality and 3D stereophonic imaging.

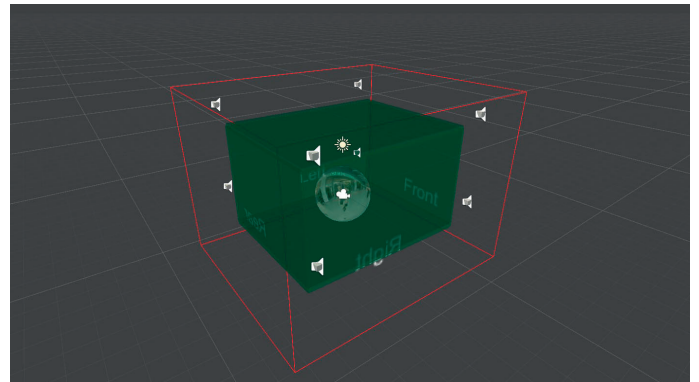


Fig. 9: Virtual 8.0 loudspeaker setup to reproduce live recorded ambience within a binaural environment

For more information:

www.hauptmikrofon.de/stereo-3d/3d-audio/ortf-3d