

Guideline

Document name

Content production guidelines: Ambisonic recordings and postproduction



Project Acronym:	IMMERSIFY
Grant Agreement number:	762079
Project Title:	Audiovisual Technologies for Next Generation Immersive Media

Revision:	1.0
Authors:	Jan Skorupa Maciej Glowiak
Support and review	Wojciech Raszewski Maciej Strozyk Sergio Sanz
Delivery date:	M30
Dissemination level (Public / Confidential)	Public

Abstract

This document contains a description of the practical recording and production of Ambisonic sound based on the experience gained in the European H2020 Immersify project. The document presents practical methods of Ambisonic sound recording supported by examples of three realizations in different acoustic conditions and using different equipments. In addition, it discusses how to mix the sound using mostly free software. Finally, it contains a description of a multi-speaker Ambisonic monitoring installation based on the example of PSNC laboratory. All chapters are accompanied by some practical advice and hints on how best to prepare for Ambisonic recordings.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 762079.

TABLE OF CONTENTS

1	Introduction	4
1.1	Theoretical background	4
1.1.1	Ambisonics	4
1.1.2	File formats	4
1.1.3	Orders of Ambisonics	5
1.1.4	Binaural decoding	6
1.1.5	Purpose of these guidelines	6
1	Recordings	6
1.1	Introduction	6
1.2	Recording techniques	7
1.2.1	Ambisonic microphones	7
1.2.2	Ambisonic microphone matrix	8
1.2.3	Recording independent sound sources	8
1.3	Practical use cases	8
1.3.1	Outdoors: Luxtorpeda rock band	9
1.3.2	Studio jazz septet recordings	9
1.3.3	Outdoors: String quintet “Noise and Structure”	11
1.4	Best practices	14
2	Mixing and editing	14
2.1	Introduction	14
2.1.1	Channel ordering	15
2.1.2	Normalisation type	15
2.1.3	Ambix and Furse-Malham	15
2.2	Software	16
2.3	Mix	16
2.4	Rendering files	18
2.5	VR implementation	20
2.6	Best practices	20
3	Auditory space	21
3.1	Introduction	21
3.2	PSNC setup description	21
3.3	Hardware	22
3.4	Software	22

1 Introduction

1.1 Theoretical background

1.1.1 Ambisonics

In the sound engineering area, there are several methods for production a multichannel spatial audio. Most of them were adapted into the movie and broadcast industry in order to give consumers better and deeper listening experiences. Commonly used surround systems such as 5.1, 7.1, 7.4.2 or Dolby Atmos are based on speakers located around the listener, where each of them have a special identifier and a dedicated location in space (e.g. “center”, “front-left”, “front-right”, etc.). It is very practical approach, but there are several limitations in the audio reproduction. In such surround systems, each audio channel is assigned exactly to one speaker. Therefore, the audio files have to be prepared according to the positions of the speakers, to achieve the correct spatial effect. There is no possibility to play 7.4.2 audio files in 5.1 setup without losing significant information and skipping some channels. However, there is a different way of thinking about spatial sound without these limitations. That method is called **Ambisonics** or **Ambisonic sound**.

Ambisonics is a breakthrough idea of a soundfield representation, where - in the contrast to typical multichannel systems - each audio channel doesn't carry signal information corresponding to the appropriate speaker. As a substitute, each channel represents properties of the acoustic soundfield. Due to such approach, the Ambisonic data file can be decoded to any speaker layouts with keeping information about sound direction.

1.1.2 File formats

It's important to distinguish the Ambisonics formats from typical multichannel audio (5.1, 7.1 etc.). The first represents and approximates the sound field around, whereas the multichannel just stores audio tracks for each loudspeaker separately. Ambisonic file can be derived only from a special microphone or encoder and carries global information about acoustic soundfield properties: sound pressure, velocity and phase. The minimal number of single channels for creating Ambisonics files are four, and each of them is called an Ambisonic component. This file is not ready to play, as a normal multichannel audio. For correct reconstruction of encoded soundfield it is require to decode it in a special way. To do so, the decoder carries complete information about the position of all speakers in surrounding space (angle, elevation, distance). On the basis of this information it is possible to decode soundfield properties to any speaker layout with preserved sound directivity. The biggest disadvantage of Ambisonics is that every speaker layout has limited **sweet spot** - the place in an audition space where all soundfield properties are preserved well. Most often it is located in a center of the speaker layout. The precision of a soundfield reconstruction and size of the sweet spot is related to the number of speakers, type of Ambisonics format, acoustic properties of auditorium, the type of decoder and encoder, as well as an Ambisonic order (See section 1.1.3).

Nowadays there are two most popular types of file to preserve acoustic space properties: **A-format** (derived from ambisonic microphone) and **B-format** (derived from A-format, encoded mono/ stereo files, multi microphone matrix).

A-format is the raw recording of the n individual cardioid capsules in Ambisonic microphones. The A-format consists of n mono channels where each of them corresponds to one microphone from capsule. For example, Ambisonic microphones in first order give the A-format composed of four channels and the second order results of 9 channels, etc. Nowadays the most popular ambisonic microphones consist mostly of four cardioid capsule where each of them are directed into different space point (front left up, front right down, back left down, back right up).

The most popular Ambisonic format that is used to store Ambisonic soundfield is **B-format**. It is the standardized Ambisonic format derived from the A-format, created from mono or stereo files by a special Ambisonic encoder or with use of special microphone matrix.¹ Commonly there are some different types of files representing B-format. Mostly this differences are connected with varied possibilities of encoding and decoding strategies, normalisation types of single audio tracks and channel ordering of components. Further information about this can be find in sections 2.1.1 to 2.1.3.

1.1.3 Orders of Ambisonics

The simplest form of Ambisonics is known as **First Order Ambisonics (FOA)** which corresponds to three figure-of-eight polar pattern microphone along each axis $X Y Z$ ² (Figure 1). FOA B-format consists of 4 audio channels (W X Y Z), where the first channel carries information about soundfield pressure corresponding to omnidirectional microphone (zeroth order) and three other channels carry information about sound directivity and phase. Sound directivity resolution represented in FOA is quite low. It is possible to enlarge the experience of sound spatial and source localisation by increasing Ambisonic order which is connected with adding more selective components in **B-format**. The increasing number of components gives us higher resolution in the domain of sound directivity. The amount of B-format components in **High order Ambisonics (HOA)** is connected with formula for 3D Ambisonics: $(n + 1)^2$ or for 2D Ambisonics $2n + 1$ where n is given Ambisonics order.

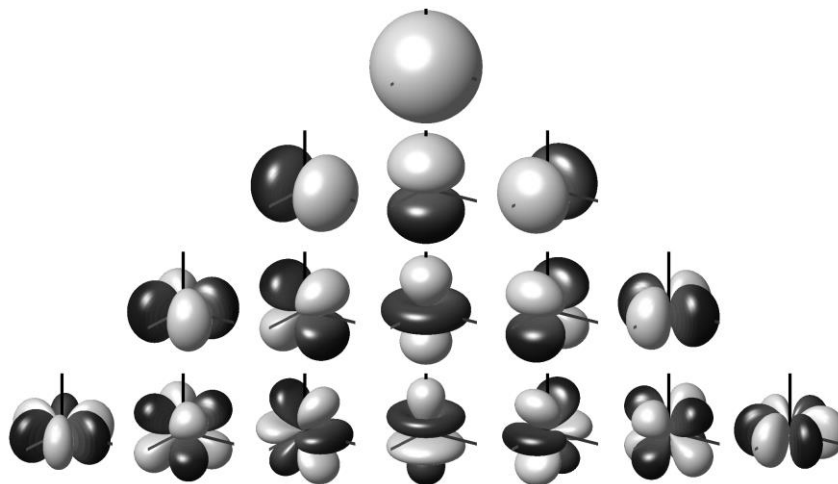


Fig.1 Visual representation of Ambisonics B-format up to 4th order³

This method is much more universal in case of audio production for immersive content like VR video. Within the scope of the Immersify project we focused on using **HOA** (over 1st order) and adapting such an approach into production of immersive and VR content.

¹ Claudio Santos, [VR Audio — Differences between A Format and B Format](#), 2017

² Daniel Artega, [Introduction to Ambisonics](#), 2015

³ Wikipedia - [Ambisonics](#)

1.1.4 Binaural decoding

One of the biggest advantage of the Ambisonics, which really distinguishes it from a typical multichannel format is that all **B-format** files can be represented in stereo domain. It allows to keep all sound spatial properties. For this purpose special binaural decoder can be used. Such decoder is based on a frequency response of our head, pinna and torso called **HRTF⁴ (Head Related transfer function)**. This function gives information about how our ear receives sound from any point of space. The averaged values of **HRTF** for humans are implemented into binaural decoder and with the use of filters (separately to left and right ear) it can create a space corresponding to the virtual loudspeakers position. After binaural decoding, the output file is typical stereo which can be played on headphones with all 3D sound aspects⁵.

1.1.5 Purpose of these guidelines

The aim of implementing Ambisonic audio for immersive video is to improve users experience and to give them high quality and realistic content for next generation media technologies, such as VR or immersive spaces. Since the approach is innovative, we needed to prepare the whole content and build an individual workflow for all steps: recording, mixing process, as well as playback at the auditory spaces. In the Immersify project, we have completed three different audio and video productions which were fully oriented on Ambisonics and VR/360. In these guidelines we provide a description of the whole audio production process and share our experience about Ambisonic audio. However, this document is not a practical tutorial on how to record, encode, and recreate Ambisonic sound step by step, but rather describes a specific approach and a workflow, giving specific examples used in the Immersify project. Where possible, we tried to provide links to further documentation.

1 Recordings

1.1 Introduction

Despite over fifty years of development, the Ambisonic technique is still not a popular approach used in audio recordings. It means that there is no regular workflow which has been created for the production of Ambisonic recordings. A significant moment in the context of popularization of this technique was the implementation of binaural Ambisonics by services, such as YouTube or Facebook which facilitated the appearance of an increasing number of tools for Ambisonic recording and processing on the market. Unfortunately, most of the products have been adapted only to the functionalities offered by the mentioned services and they are usually limited to FOA. However, the effect of such recordings with the use of Ambisonic microphones may be quite satisfactory for field recording combined with 360° video recording. In more advanced cases, such as the recording of a larger music band, such solutions turn out to be insufficient. The use of a single Ambisonic microphone does not provide an audio reproduction with satisfactory directionality and selectivity. Recordings of

⁴ Wikipedia - [HRTF](#)

⁵ Immersify - [binaural ambisonic tests](#)

this type require additional traditional microphones. For this reason, the key element of such recordings is the mixing and post-production process.

1.2 Recording techniques

In general, Ambisonic recordings can be made in three ways. The first one is to use a dedicated Ambisonic microphone, secondly - a specialized Ambisonic microphone matrix can be used and the third way is to take classic microphones, which are used to record individual sound sources, and then, with the help of dedicated software, to recreate the whole acoustic space by spatializing individual monophonic signals and reverberating them properly. Using hybrid techniques is also possible in order to achieve the best possible quality.

Each of these methods allows for a slightly different approach to the recording and brings different final sound effects. These differences can be seen primarily in two areas: concerning the reproduction of the acoustic space and the selectivity of individual sound sources. The selection of particular solution depends largely on the effect we want to achieve at the end. Based on our experience, we try to describe the benefits of each approach in the following sections.

1.2.1 Ambisonic microphones

There are some dedicated ambisonic microphones available on the market. Most of them have a relatively similar design: they usually consist of four independent condenser microphones with cardioid polar pattern, each facing a different direction. They are placed on one sheltered frame and form a closed dome. Each cardioid is responsible for recording the acoustic wave coming from different points in space. The manufacturers of these devices also provide dedicated software (e.g. a plugin) allowing to encode A-format signals to B-format. These devices are very simple to use, but usually allow only for FOA Ambisonic recordings. Examples of products in this category are, among others:

- Sennheiser ambeo VR mic (figure 2)
- Zoom H3 VR
- Rode NT-SF1

Some manufacturers also offer commercially available microphones that enable HOA recordings - the second order, like the Core Sound Octo Mic, and even the third order, like Zylia ZM-1 microphone.



Fig.2 Sennheiser Ambeo VR mic⁶

⁶ <https://sennheiser.pl/o/ambeo-vr-mic/pl/o/ambeo-vr-mic>

1.2.2 Ambisonic microphone matrix

The second way for a FOA recording is to construct a matrix consisting of three bidirectional microphones and one omnidirectional microphone. Such systems allow recording directly to B-format without the need to encode the acquired signal through dedicated software. The signal from each bidirectional microphone corresponds to the next ambisonics B-format components: X Y Z, and omnidirectional microphone to W component according to the diagram below (Figure 3). Due to the complexity of such a microphone matrix, we will not discuss it in detail here. Additional information can be found in "Ambisonics A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement and Virtual Reality."⁷

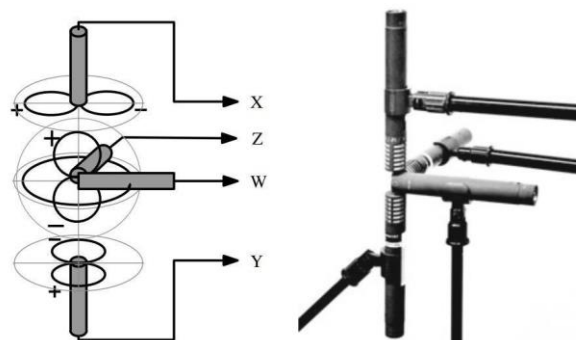


Fig.3 microphone matrix providing B-format recordings⁸

1.2.3 Recording independent sound sources

The third way of making recordings for an Ambisonic production, completely different from the ones described above, is to record individual sound sources separately with ordinary microphones and to sequentially make the mono tracks spatial during the mix with use of Ambisonic encoders. Although this method is, as far as the idea is concerned, contrary to Ambisonics, it offers more possibilities of creating Ambisonic recordings and allows to achieve greater selectivity of sound. There are tools available on the commercial market that allow to create full acoustic space in the domain of high order Ambisonics. Thanks to this, it is possible to skip the whole process of complicated microphone setup or the use of expensive capsules. The recording can be made in the classic way, assuming that the whole process of sound spatialization is transferred into the domain of post-production. Depending on the purpose of the recording, we use classic condenser or dynamic microphones for instruments.

1.3 Practical use cases

In the Immersify project, we conducted three different recording sessions with various types of music in different acoustic conditions. All scenarios required a different recording approach. The first one was a recording of an outdoor rock band concert taking place at the festival. The second one was a

⁷ Franz Zotter and Matthias Frank: "Ambisonics A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement and Virtual Reality."

⁸ [Franz Zotter, Matthias Frank: Ambisonics A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement and Virtual Reality 2019](#)

studio recording of a jazz septet and the last one was an outdoor recording of a contemporary piece written for string quintet and real-time processed electronics presented in an octogonal system.

1.3.1 Outdoors: Luxtorpeda rock band

Integrated Ambisonic microphone and compact 360 camera

The first Ambisonic project was a live recording of Luxtorpeda rock band during the Luxfest festival in Poznan. Unfortunately, in this case, due to the lack of possibility to build our own microphone setup, we had to base only on the Ambisonic microphone and sums from the mixer of the sound engineer at the concert. In this case, we wanted to recreate the full stage situation where, combined with a 360° video, the recipient would have an impression of being between the band and the audience during the concert. We placed the ambient microphone together with the 360° camera at the edge of the stage. This setting made it possible to avoid placing the microphone within the array of the main speaker system. We managed to get spatial recordings by locating the ambisonic microphone close to the reference monitors of the musicians. In addition, the presence of a large audience resulted in sounds typical for a concert such as audience singing or applause. The sound was recorded with a portable zoom recorder. The fact that we had to pay special attention during the mixing of the material was the incompatibility of the phase of the signals from the microphone and the mixer console.

1.3.2 Studio jazz septet recordings

Independent sound sources and compact 360° camera

Our primary goal was to record a music band with a fairly large instrumental composition and extensive musical material, so that in the final recording we could easily fulfill 360° degrees of auditory space with instruments. We invited the jazz septet called "Anomalia", which consisted of: drums, double bass, alto saxophone, tenor saxophone, trombone, trumpet, and electric guitar. The recordings were made in a specially adapted studio of Poznan Supercomputing and Networking Center. Since the studio is well adapted acoustically, we could create good recording conditions without early reflections or large reverberations. This acoustic environment allowed us to achieve the best possible directionality in the final mix. Above all, to experiment with HOA, we decided to approach the recordings in a classic way (as described in Section 1.2.3), focusing on the precise recording of each instrument individually, and used an additional ambisonic microphone only to record the overall sound plan. We wanted the audio recording to be combined with the accompanying 360° video as well, so that it could be played later in a VR installation or Head Mounted Display.

The musicians were placed in a circle, with the Insta360 Pro camera in the center of it. This device was recording the omnidirectional video in 8K (7680 × 4320) resolution (Figure 4). In order to preserve the naturalness of the sound recording, we deliberately did not focus on eliminating crosstalk between the individual microphones and assumed that this approach would allow us to achieve a more real acoustic space imitating a more natural listening experience.



Fig.4 Jazz band recordings setup source: PSNC documentation

In order to record wind instruments, we used DPA 4099 microphones. An additional condenser SE2200a microphone was used for the tenor saxophone in order to obtain more accurate recording of lower frequencies. For the recording of the double bass, we used line output from the amplifier. This solution was not fully satisfactory because in this configuration it was impossible to record the interactions between strings. So, to achieve the desired effect, we additionally installed a DPA instrumental microphone. In the case of the electric guitar, in order to get the deepest possible sound, two different types of microphones were used: dynamic Shure SM57 and condenser Audio Technica AT 2050. Additionally, to obtain ambient sound and a general acoustic image, Ambeo VR Mic from Sennheiser was hanging from the ceiling above the Insta360 Pro camera, which made FOA recordings possible. The sound captured by this microphone served in the further stage (mix) as the basis for multi-speaker projection. The live recording was made using Midas Pro's microphone preamps and converted to Dante through KlarkTeknik DN9650. The digital signal was then sent to a recording workstation with Reaper DAW software. The schematic view on the recording setup has been depicted on Figure 5.

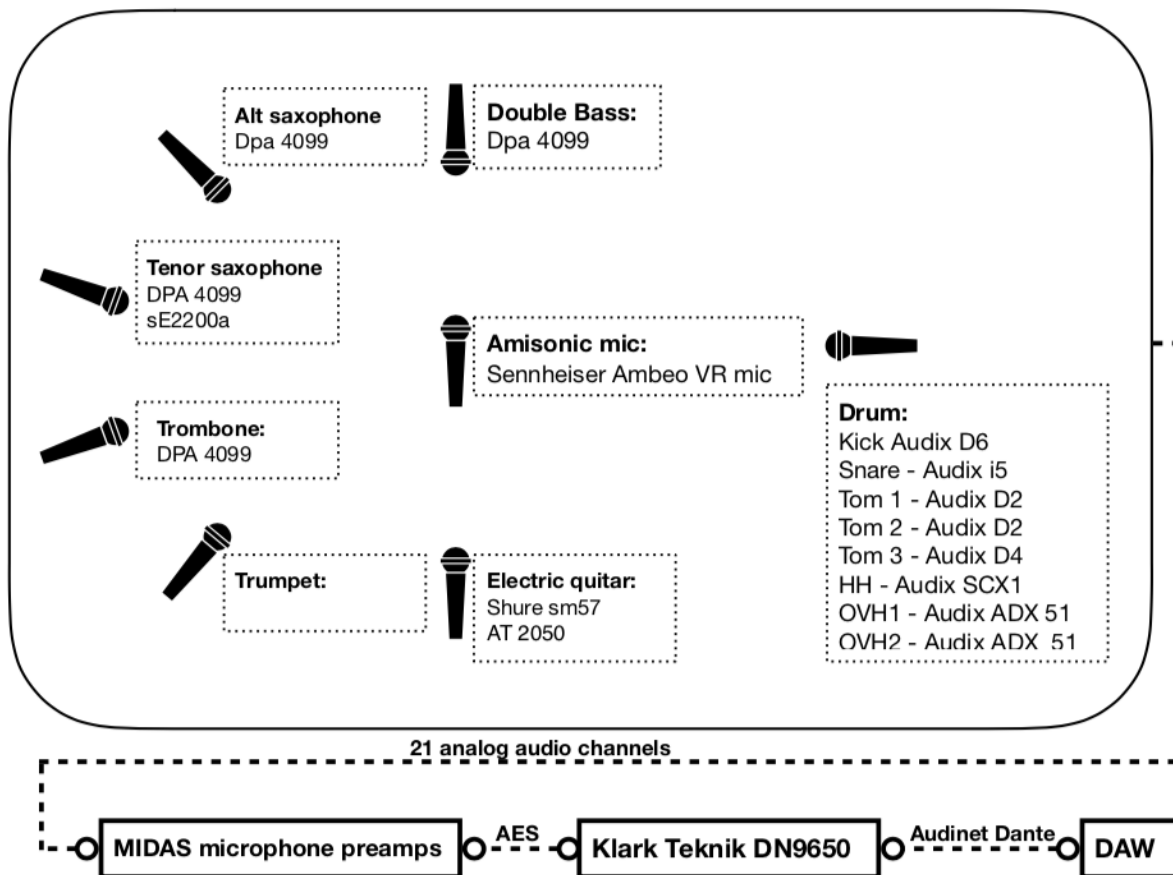


Fig.5 Jazz band recording setup

1.3.3 Outdoors: String quintet "Noise and Structure"

Independent sound sources and custom high-resolution 360° camera rig

The last Ambisonic recording of the Immersify project was much bigger undertaking. The idea was twofold: 1) to record Ambisonic sound in outdoor conditions; and 2) to obtain both interesting background sound and 360° image in higher resolution than 8K (7680 × 4320). High resolution recordings was provided by a custom rig constructed of 8 BlackMagic Micro Studio 4K cameras with Olympus 8mm lenses (Figure 7). The total resolution provided by this custom 360° camera rig is 16K (15360 x 8640). Due to the 360° video recording, we wanted the choose place to be visually attractive. For this purpose, we decided to use the roof of the building of Poznan Supercomputing and Networking Center, which provide us an amazing scenery of the city. From the musical point of view, we recorded Jan Skorupa's original contemporary composition entitled: „Noise and Structure". This composition is written for string quintet (double bass, cello, viola, two violins), five crystal bowls and an electronic layer performed live in an eight-channel octagonal system. All the musicians, as before, were placed on a circle, in the middle of which an Ambisonic microphone and the 360° camera rig was set up.



Fig.6 Noise and structure rooftop recordings source: PSNC Documentation



Fig.7 Eight cameras rig for 360 video recordings source: PSNC documentation

In this case, it was crucial to direct the recordings in such a way that the acoustic background was also recorded very well. For this purpose, we set microphones for each instrument in two ways: first, we used the DPA 4099 instrumental microphones, which provide exceptional dumping to any background noise, what finally gives quite accurate recording of the direct signals from each instrument; second, we placed the Audio Technica AT897 super-directional microphones further from the instruments on the tripods, allowing for the recording of each instrument with a well-balanced background level. Those microphones were simultaneously directed to crystal bowls. To avoid wind noise, each microphone was inside a wind capsule. In order to reproduce the electronics, we set octagonally eight Genelec 8010A studio monitors. To avoid possible crosstalk of the electronic part, the speakers were positioned in front of the musicians. The recordings were made with a rack build of microphone preamps Behringer ADA8200, which was connected via ADAT protocol to a Focusrite Rednet 3 interface. This interface was connected to a computer with Dante protocol. We also used this setup because the electronic part was mostly based on real-time sound processing of acoustic instruments. By the Dante protocol it was easy to route the signal from the microphones to the electronics performer computer and directly route the electronics to an omnidirectional system. Figure 7 presents a detailed diagram of connections between audio equipment. This setup allowed us to record the whole composition while maintaining a perfect reproduction of the sound scenery. In the further stage of work on the recording, individual plans of each instrument and a general ambience from the Ambisonic microphone guarantee to build the space of the final material.

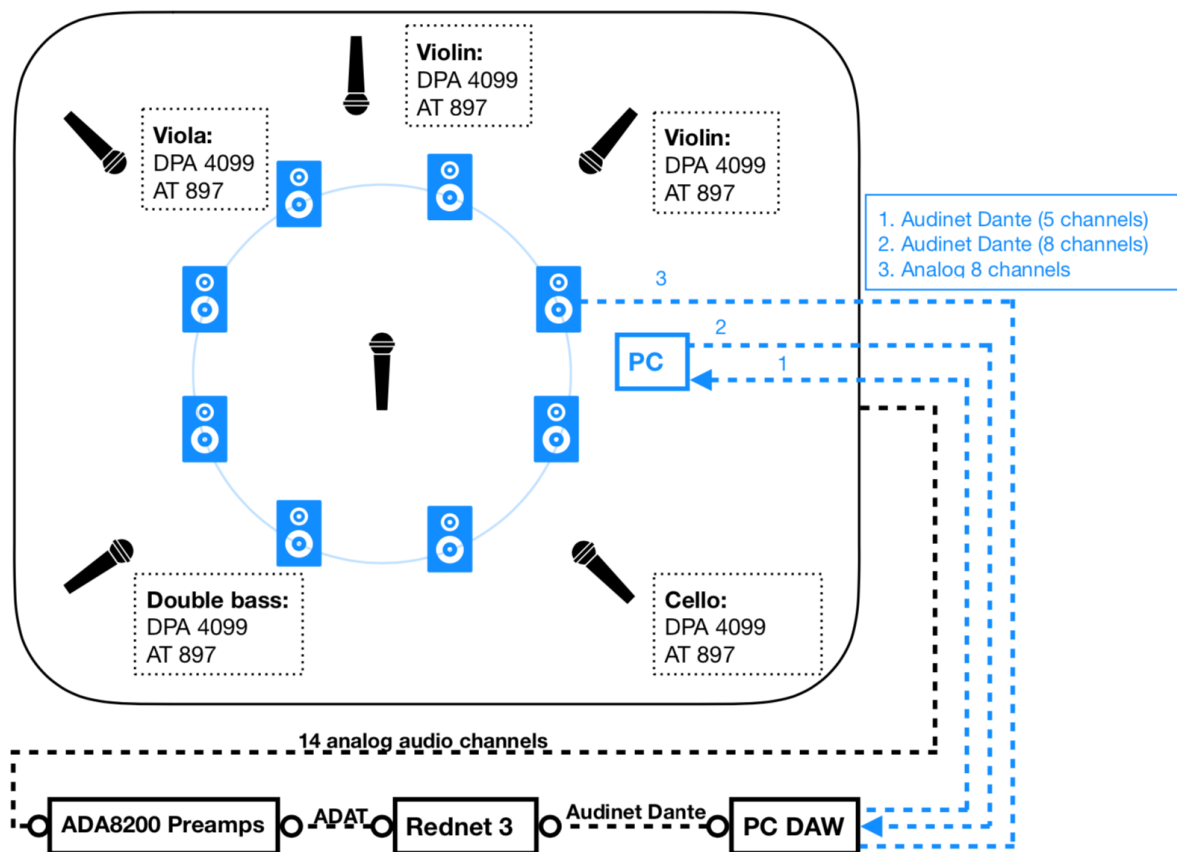


Fig.8 "Noise and structure" recording setup

1.4 Best practices

Ambisonic recordings hints:

- The recording method and selection of microphones is closely related to the acoustic environment. In the case of studio recordings with larger music bands, the use of an Ambisonic microphone does not make much sense.
- In the case of music recordings, even if we use an Ambisonic microphone, it is worthwhile to catch each instrument separately in order to be able to complete the recording from the Ambisonic microphone at the post-production stage to achieve better resolution in the domain of source localisation.
- The Ambisonic microphone is ideal for ambient recording in a spatial manner.
- In case of creating acoustic space in post-production, it is worthwhile to record each instrument from two different distances.

2 Mixing and editing

2.1 Introduction

The mixing process in the Ambisonic domain is similar to the typical audio production. For this purpose we also use tools such as: compressor, noise gate, correction, etc. However, Ambisonics mixing is associated with a few technical limitations. First of all, not every DAW⁹ (Digital Audio Workstation) software gives the possibility to work with such material. This is due to the fact that the B-format which we have to encode each audio signal is a multi-channel format (Chapter 1). Secondly, after the encoding process we are no longer able to use any popular Virtual Studio Technologies¹⁰ (VST) plugin because commercial tools are rarely able to support B-format files.

Another difficulty is that we are using both classic microphones (recording any mono or stereo source) and Ambisonic microphone recordings. These two types of output recording materials require a slightly different approach, so during the mix we have to separate them from each other. In case of using mono tracks with appropriate B-format encoders, we have the possibility of using any spatial arrangement of sound sources. However, if an ambience microphone is used, the recorded material is a representation of the acoustic field recorded in A-format, which requires an appropriate converter to be transposed to B-format. It means, that the output format coming from an Ambisonic microphone capsule is Ambisonics A-format (4 mono channels corresponding to each cardioid: front left up, front right down, back left down, back right up) and needs to be converted into B-format (W, X, Y, Z corresponding to the three bidirectional microphones and one omnidirectional). In this case, we do not have the possibility of separating and arranging single sound sources in space, but only rotating the whole acoustic scenery. With respect to an Ambisonic mix, it is important to choose the order of Ambisonics (described in Chapter 1) in which we encode our files. The Ambisonic order defines the resolution in the domain of sound directivity. Theoretically, the higher Ambisonic order, the better

⁹ DAW stands for Digital Audio Workstation and is a device or a software application used for audio recordings, edition, post production and music production. For more details see [DAW at Wikipedia](#).

¹⁰ VST is an audio plugin standard that is commonly used for audio recording and production. For more details see [VST at Wikipedia](#).

sound localization can be achieved. Practically in our case (layout with 24 speakers) the biggest difference in the domain of source localisation was between first to third order. For higher orders than the third one, the difference was noticeable but not huge. Regarding this fact in our productions we mainly oscillated between 3rd and 5th order, but testing also higher orders.

The final and the biggest difference is that for such productions we need a special multi-speaker auditory space. You can use decoders implementing Head Related Transfer Function (HRTF) and listen to the mix in the domain of binaural Ambisonics, but this approach does not give us as many possibilities as the multi-speaker space.

2.1.1 Channel ordering

Another aspect of the Ambisonic mix is the ordering of successive B-format components. This basically does not affect the sound quality of the final material, but is very important when decoding the file to a particular speaker constellation. The type of decoder ordering components must always match the one used in the encoder. Currently there are several different ways of channel ordering (Fig. 9)

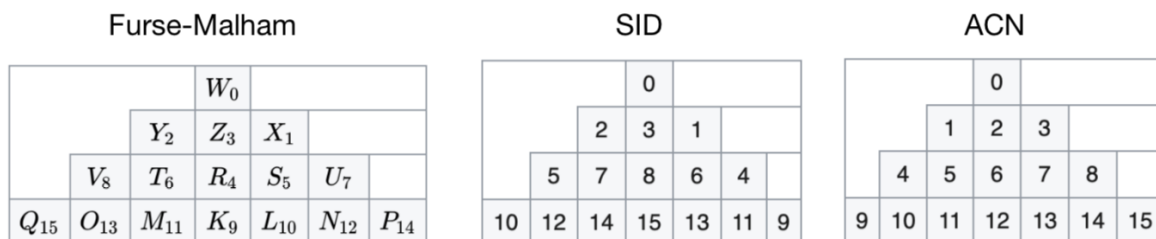


Fig. 9 Ambisonic components ordering systems¹¹

2.1.2 Normalisation type

When working with a B-format, the type of normalisation of its individual components is an important element. Currently there are several different approaches to this issue:

1. **maxN** - normalizes each single component to never exceed a gain of 1.0 for a panned monophonic source.
2. **SN3D** - With SN3D, no component will ever exceed the peak value of the 0th order component for single point sources.
3. **N3D** - Ensures equal power of the encoded components in the case of a perfectly diffuse 3D field.¹²

2.1.3 Ambix and Furse-Malham

Currently, the most popular Ambisonic format are **B-format Furse-Malham (FuMa)** or **Ambix** - and both of them are really similar¹³. Main difference is connected with aspects mentioned before, channel ordering and normalisation types. **Furse-Malham** are using: Furse-Malham Channel ordering

¹¹ [Wikipedia - Ambisonics data exchange formats](#)

¹² [Wikipedia - Ambisonics data exchange formats](#)

¹³ [Ambisonics data exchange formats, Wikipedia](#)

(W, X, Y, Z) and **maxN** normalisation. **Ambix** are based on **ACN** channel ordering (W, Y, Z, X) and **SN3D** normalisation. In the majority of our ambisonic works provided during Immersify project we used Ambix format with ACN channel ordering and SN3D normalization.

2.2 Software

The choice of the right DAW, combining Ambisonic encoders and decoders, has a significant impact on the Ambisonic production process. As the B-format is a multi-channel file, it requires the software to support multi-channel audio files within one bus. The best DAW we tested was REAPER COCKS¹⁴, which allowed for up to 64 audio channels within one track. Although, the application is commercial, its price of \$60 allows for use even in non professional productions.

There are also a number of Ambisonic tools available on the Internet in the form of VST plugins allowing for encoding and decoding and basic processing of sound files in B-format such as: compression, correction or even reverberation. These toolkits or plugins are:

- **Ambisonic Toolkit (ATK)**¹⁵. Basic encoders and decoders to work in first order Ambisonics. Recently they have been implemented into the basic DAW Reaper COCKS tools.
- **IEM Plugins**¹⁶. The comprehensive free of charge package tools for Ambisonics up to the seventh order created at the Gratz University. In this set are available Ambisonic reverb and delay working in B-format.
- **O3A core VST**¹⁷. Plugins created by Blue Ripple Sound. The basic "Core" set includes free tools that allow you to work in the 3rd order B-format but with limited decoding capabilities. More advanced tools from this manufacturer require purchase.
- **SurroundZone2**¹⁸. A tool that allows you to convert the A-format signal from an Ambisonic microphone to a B-format
- **AmbiFreeVerb**¹⁹. Free package containing encoders and decoders to work in third order Ambisonics. In addition to the basic tools, an Ambisonic reverb is also included.

For our productions we have chosen the complete IEM Plugins toolkit. In addition to basic ambisonics encoders and decoders, this package includes tools for music production. This software supports Ambisonic files up to seventh order. To convert A-format into B-format we used a dedicated tool from Sennheiser Ambeo A-B converter.

2.3 Mix

Due to the lack of a large number of tools for mixing sound in the Ambisonic domain, it is advisable to divide the work into two separate stages:

¹⁴ <https://www.reaper.fm>

¹⁵ [The Ambisonic Toolkit, Tools for soundfield-kernel composition](#)

¹⁶ [The IEM Plug-in Suite](#)

¹⁷ [Blue Ripple Sound, O3A Core plugin](#)

¹⁸ [Soundfield surroundzone2](#)

¹⁹ [The Blog of Bruce, Ambisonic Music Creation for the Masses... and other bits of bobs - Wigware](#)

- The first one is **pre-processing** of the recorded sound material by denoising, pre-equalization and dynamic range compression,
- and the second stage is the **encoding** of the prepared material to B-format. At this stage, we also created the entire acoustic space by using reverberation or additional correction or dynamic compression.

At the very beginning of the Ambisonic mix we have to choose an Ambisonic order in which we want to prepare recordings. This determines the way we have to prepare the project template. For example, if we assume 5th order Ambisonics, which consists of 36 Ambisonic B-format components, the project will have to be configured with a total of 36 tracks. In the project we place on timeline all the mono files from the recordings and, by using the encoder set to the appropriate order, we transform them into B-format (Figure 10). We use IEM encoders, which also allows for spatialization in the space of given sources. It can be done in a very simple way: by using the graphical interface of the encoder the azimuth, elevation and width of each instrument is set. In our case, due to additional 360° video recordings, we define the location of individual instruments that is identical to the location of sound sources in the camera image.



Fig.10 IEM Ambisonic encoder

To obtain a better timbre of spatialized sources, encode tracks can be additionally equalize and dynamic range compressed. All multi-channel Ambisonic tracks should be sent to the decoder. For this purpose, we create another track, in which we place the Decoder (in our case also the IEM tool) in the effect chain, where we could define the whole listening space in a virtual way (more in Section 4.4). The Decoder allows to prepare the whole mix for listening in the speaker space. Our installation is equipped with an additional sub-bass speaker, but unfortunately the IEM decoder does not allow to add an LFE channel. In order to enrich our recordings with the lowest frequencies in the project, we usually create an additional track, which after going through a 150 Hz low-pass filter is routed directly to the subwoofer. Using this track we can create a separate mix containing only frequencies below 150Hz.

Recordings from an Ambisonic microphone require a different approach. First, we need to create a group of 4 tracks, on which we place each separate mono channel, which is a record of the acoustic field in A-format. These tracks must then be converted to B-format. For that purpose the dedicated "Ambeo A-B Converter" tool (Fig. 11) can be used.

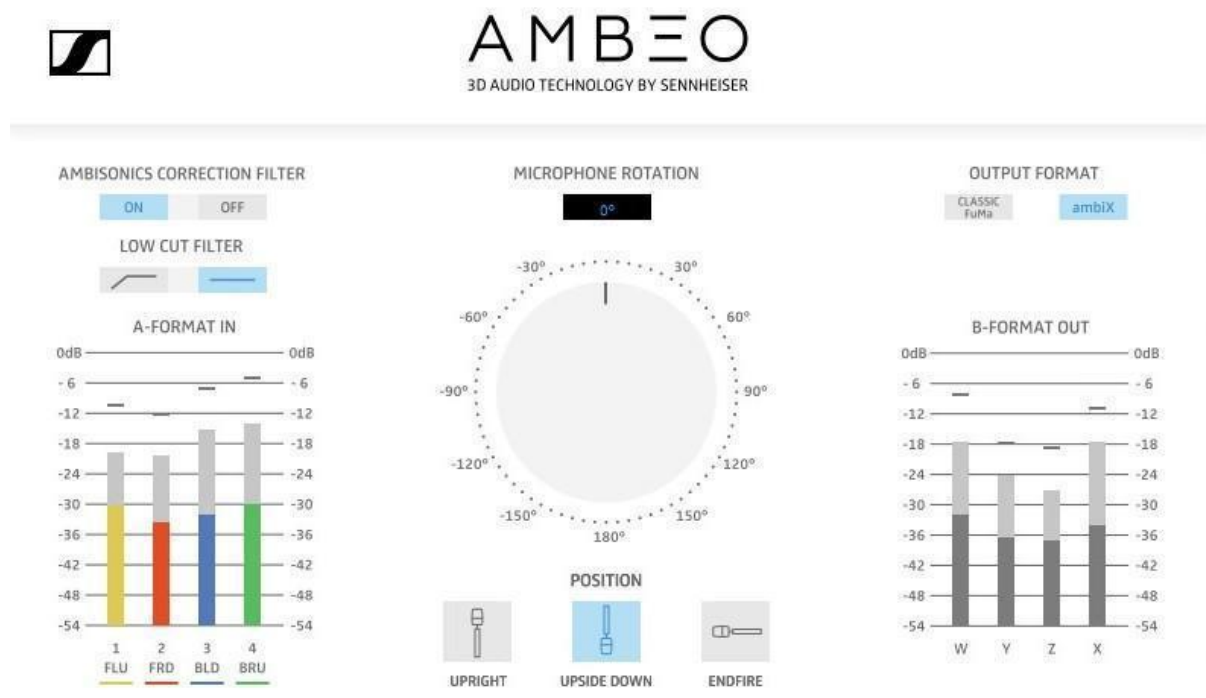


Fig.11 Dedicated tool for Sennheiser Ambeo VR MIC A-B converter

With this plugin we obtain B-format ambiX or Classic FuMa tracks. Due to the fact that we use SN3D normalization and ACN ordering, it is necessary to select as an output the ambiX format. When using Classic FuMa you should use an Ambisonic format converter, because in this format, the way in which the tracks are normalized and ordered is different. As an output the FOA B-format is obtained. Due to the fact that each successive order also contains the previous ones, it is not necessary to decode microphone recordings separately. If you keep the ordering and normalization, these tracks can be sent to a decoder set to a higher order.

In order to get a consistent space in the mix, we do not use reverb on each of the sources, but we only create two separate sent buses: one with a long reverberation time and the other with a short one, where we send out single tracks controlling the reverberation level of the dispatch. For the reverberation we also use an IEM-package tool. Note, that the reverberation track should also be multi-channel according to the Ambisonic order which we used. Then the reverberation bus should also be sent to the decoder.

2.4 Rendering files

The ready Ambisonic material can be rendered in two ways:

- The first one is used when the mix is **dedicated to listening on a particular speaker installation** where all the work relate to the mix was performed. The recordings can then be render after decoding the B-format to a specific speaker layout.
- The second one is used when the material is **intended for listening on other installations or binaural projection**. In this case, the file should be save in a specific B-format appropriate for the Ambisonic order.

In order to render a ready mix, it is good to prepare a separate track on which all other encoded tracks should be sent (except the LFE channel). You should remember that it should also be a multi-channel track. In the render options we chose as source: "Stems (selected tracks)" and as output WAV file format with option: "Force Wave64". (Figure 12).

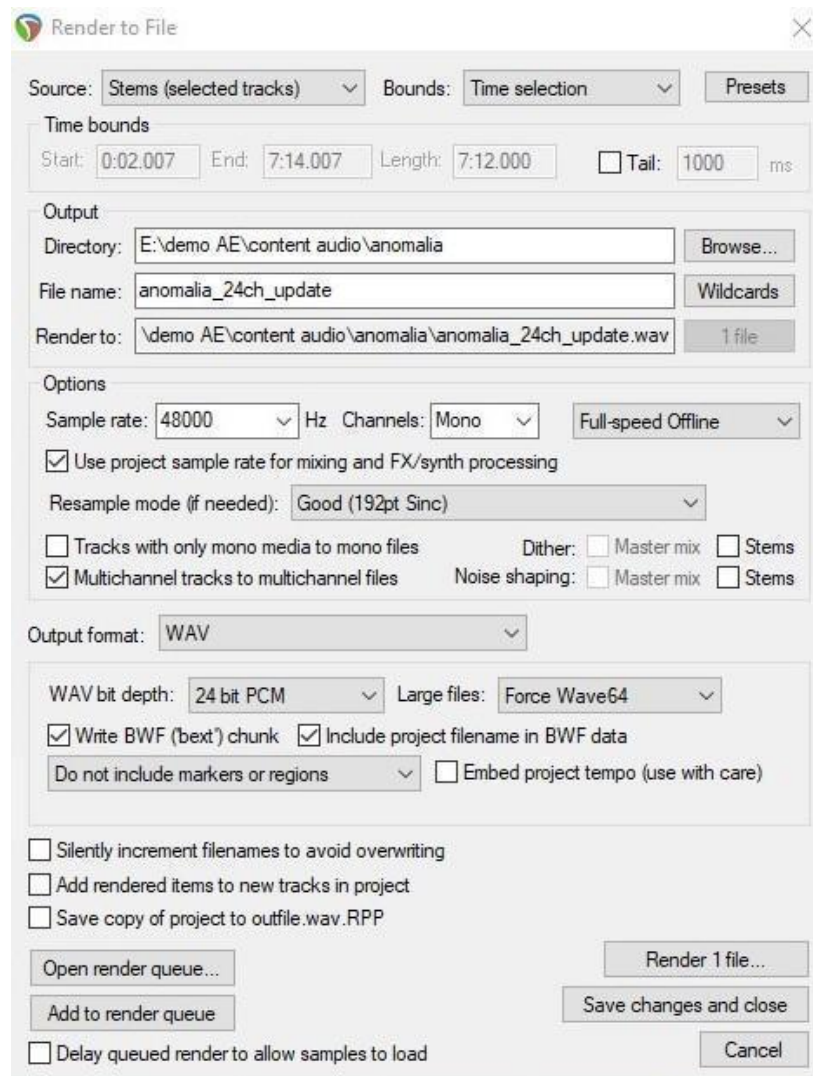


Fig.12 Render window

After rendering, one multi-channel WAV file with the number of channels corresponding to the Ambisonic order is generate - in the given example we can see 5th order B-format with 36 components (Figure 13). Separately as mono WAV we have to render the channel with the subwoofer mix. In case of rendering files for a particular audio installation, the steps remain the same, but the output number of channels shall be consistent with the number of speakers on which we will listen to the recordings.

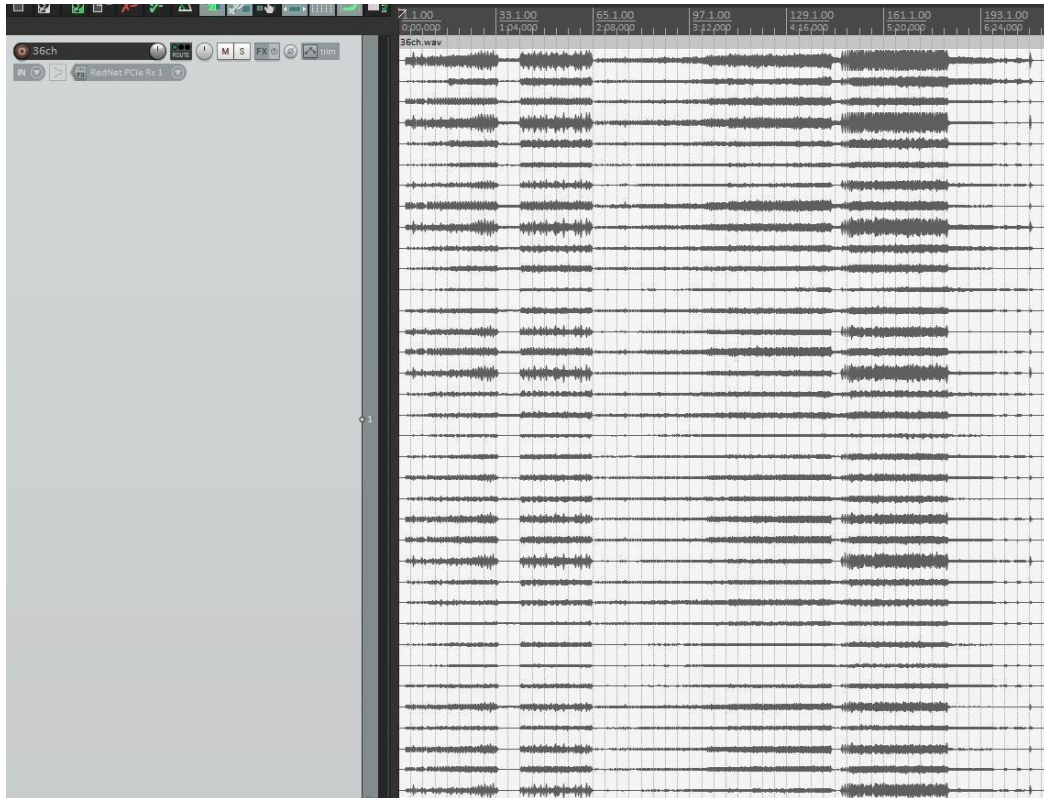


Fig.13 5th order Ambisonic B-format wav file

2.5 VR implementation

The last step of the Ambisonic production was a combination of the 360° video presented in VR glasses with simultaneous sound played on our installation. The problem we challenge at this stage was the issue of synchronization of both materials. For this purpose we used a generated linear timecode LTC, which was placed under the audio track in the presented video file. The signal with the timecode was sent to DAW using the bluetooth interface, where the current video position was read and synchronized with the audio.

2.6 Best practices

Ambisonic mix hints:

- Recorded files intended for the ambience mix should already be denoised, pre-compressed and equalized.
- The Ambisonic order defines the resolution of the final mix's sound directionality.
- The template of the project is strictly related to the defined order of Ambisonics.
- When using many different Ambisonic tools, it is important to pay attention to what types of channel ordering and normalization are implemented in them. In case of incompatibility, use encoders.
- Ambisonic microphone signals are represented in the A-format, which requires conversion to B-format.
- Ambisonic encoders are used to convert mono or stereo signals into B-Format.
- Channel ordering and the type of B-format normalisation components in encoder must be compatible with that in the decoder.

3 Auditory space

3.1 Introduction

There isn't any standardized speakers setup for the Ambisonic production, so we decided to build our own 24.1 (with possibilities to enlarge into 25.2) spherical auditory space. Our aim was to create a setup which would be very flexible and easy to install. We wanted to have permanent control of height, diameter and full flexibility of placing all the speakers. Such an approach gave us scenery which could be easily transformed to workspace dedicated also to typical multichannel production.

3.2 PSNC setup description

The auditory space is based on Genelec 8010A active studio monitors and subwoofer Genelec 7030B. Conventionally we use 24 independent speakers which are located around the listener. Speakers are set up on three different heights in octagonal order (Figure 14). All the speakers are creating a full sphere with sound sources above and below the listener's head. Such construction provides the possibility of reproducing sound horizontally and also vertically with full azimuth and elevation.

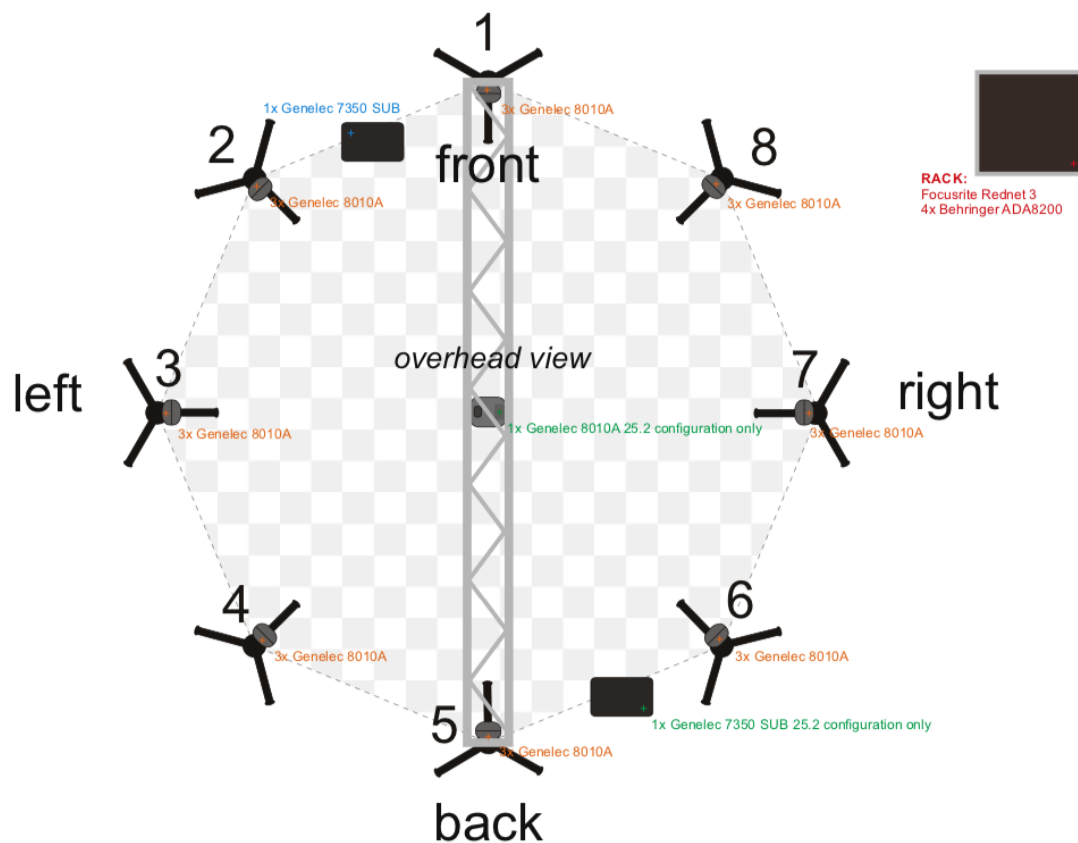


Fig.14 PSNC 24.2 speaker setup

3.3 Hardware

The auditory space is based on Focusrite RedNet 3 audio interface which is connected through ADAT protocols to four Behringer ADA8200 preamps. The interface is also equipped with Dante protocol, which is connected to the Focusrite RedNet PCIe card through a network cable. This connection provides up to 32 physical audio outputs in sample rate up to 48 kHz (Figure 15).



Fig.15 hardware connection speaker setup

3.4 Software

The whole auditory space is operated by one PC machine with the use of Dante control matrix. It provides simple routing of the audio channels directly into all speakers. This connection is also very universal and allows easy and fast changing the speakers layout from simple 5.1 setup up to full 24.1 layout.

At the beginning, we had to calibrate and calculate a dedicated decoder for our installation. As mentioned in the previous sections, the toolkit IEM contains²⁰ an appropriate decoder for that purpose. We had to virtually set all speakers' locations by adding azimuth, elevation angle and radius from the listening position (Figure 16).

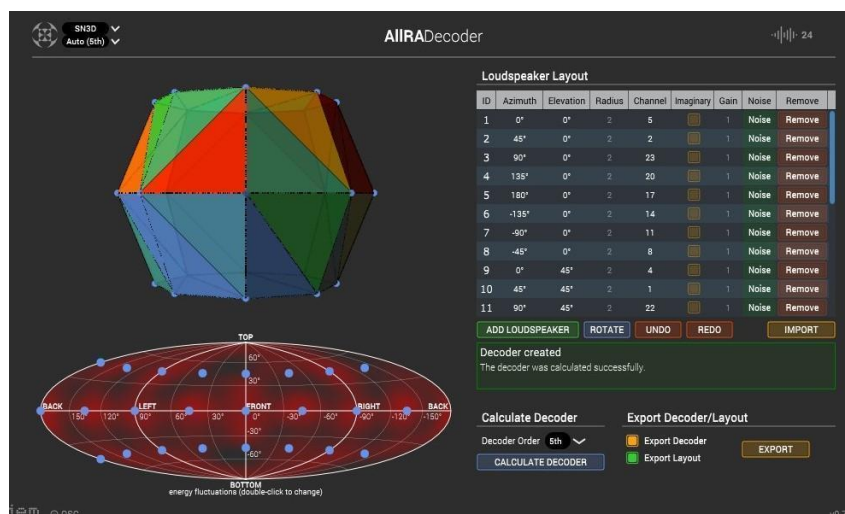


Fig.16 IEM AIIRA decoder with PSNC speaker layout

This tool enables the decoding of any B-format Ambisonic files into a virtually created speakers layout. The IEM plugins also allow for advanced settings of playback scenery. With the use of the Distance Compensator we can make some additional gain and delay correction which can improve the quality of Ambisonic space (Fig. 17).

²⁰ <https://plugins.iem.at>

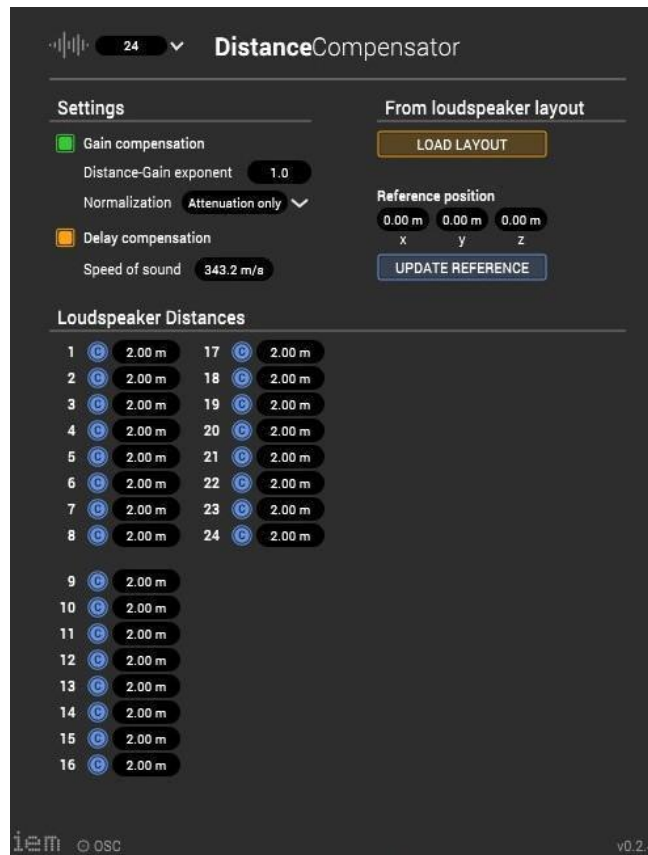


Fig.17 IEM distance compensator with implemented PSNC speaker layout

4 Conclusions

In these guidelines we presented our methodology for recording, producing, mixing and playing the Ambisonic sound, supported by practical examples. Many of our productions are experimental in nature, and the knowledge about software and hardware integration has been acquired largely experimentally. Therefore, this description is certainly not fully complete, but we hope that it will allow readers to get the necessary information to test the Ambisonic sound on their own.

We also encourage you to watch the documentary footage from ambisonic sound productions, which can be found at www.immersify.eu, especially:

- [Experimenting with ambisonics and binaural audio](#), April 5, 2019 by Jan Skorupa, Wojciech Raszewski, Bartłomiej Idzikowski and Maciej Glowiak
- [Recording and production of ambisonic sound for Immersify](#), March 10, 2019 by Jan Skorupa, Wojciech Raszewski, Eryk Skotarczak and Maciej Głowiak

among other Immersify project news and blog articles about recording and demonstrating Ambisonics sound.