FH | JOANNEUM
University of Applied Sciences

kunst uni graz

3D Audio Sound Design

Creating Immersive Spatial Audio Experiences

Michael Bernsteiner

University of Music and Performing Arts Graz

Institute of Electronic Music and Acoustics

FH JOANNEUM University of Applied Sciences

Institute of Design & Communication

Supervisor: **Univ.Prof. Dipl.-Ing. Dr.techn. Alois Sontacchi**

This thesis is submitted for the degree of

Master of Arts

September 2019

DECLARATION

Ich erkläre ehrenwörtlich, dass ich die vorliegende Masterarbeit selbstständig angefertigt und die mit ihr verbundenen Tätigkeiten selbst erbracht habe. Ich erkläre weiters, dass ich keine anderen als die angegebenen Hilfsmittel benutzt habe. Alle aus gedruckten, ungedruckten oder dem Internet im Wortlaut oder im wesentlichen Inhalt übernommenen Formulierungen und Konzepte sind gemäß den Regeln für gutes wissenschaftliches Arbeiten zitiert und durch Fußnoten bzw. durch andere genaue Quellenangaben gekennzeichnet.
Die vorliegende Originalarbeit ist in dieser Form zur Erreichung eines akademischen Grades noch keiner anderen Hochschule vorgelegt worden. Diese Arbeit wurde in gedruckter und elektronischer Form abgegeben. Ich bestätige, dass der Inhalt der digitalen Version vollständig mit der gedruckten Version übereinstimmt. Ich bin mir bewusst, dass eine falsche Erklärung rechtliche Folgen haben kann.

Graz, am 4. September 2019

Michael Bernsteiner

CONTENTS

ACKNOWLEDGEMENTS

I would like to express my very great appreciation to my supervisor Alois Sontacchi for his valuable and constructive suggestions during the planning and development of this thesis. His willingness to give his time so generously is very much appreciated. Furthermore, I would like to thank my second examiner Josef Gründler, also for his support throughout the program.
My grateful thanks are also extended to all the immersive audio experts who gave their time and took part in my survey. Special thanks should be given to my colleagues Paul and Clemens for testing my questionnaire. Advice given by Stefan Warum and Daniel Rudrich has been a great help in producing my audio drama.
Finally, I want to express my thanks to Sarah for her endless support during the phase of work.

ABSTRACT

This master thesis and the associated spatial audio piece treat the subject of immersive sound design and 3D audio production techniques. The theoretical part of this paper provides basic information on 3D audio formats as well as their advantages and disadvantages. The practical part describes the production process of the Ambisonics audio drama “Spinnenbank”, which is based on a text written and narrated by language artist John Sauter. The documentation of the artistic contribution and the production environment used should provide an insight into the production effort of Ambisonics audio content. Additionally, a questionnaire regarding 3D audio production techniques was designed and sent out, to obtain first-hand information from immersive audio experts. The knowledge gained from the survey is compared with the experiences gained from creating the audio work and corresponding paper.

KURZFASSUNG

Diese Masterarbeit und das zugehörige Werkstück behandeln die Themengebiete immersives Sound Design und 3D Audio Produktionstechniken. Im Theorieteil der Arbeit werden die gängigsten räumlichen Audioformate sowie deren Vor- und Nachteile vorgestellt. Der praktische Teil umfasst den Produktionsprozess des Ambisonics Hörspiels "Spinnenbank", das auf einem Text des Sprachkünstlers John Sauter basiert. Die Dokumentation des künstlerischen Beitrags und der verwendeten Produktionsumgebung soll einen Einblick in den Produktionsaufwand von Ambisonics Audioinhalten ermöglichen. Ein Fragebogen zu 3D-Audio Produktionstechniken wurde entwickelt und ausgesendet, um Informationen aus erster Hand von immersiven Audioexperten zu erhalten. Die aus der Umfrage gewonnenen Erkenntnisse werden den eigenen Erfahrungen, die während der Erstellung des Werkstücks gemacht wurden, gegenübergestellt.

1 INTRODUCTION

As gradually more content for virtual reality (VR), augmented reality (AR), mixed reality (MR), cinematic reality (CR) and other immersive media formats is being produced, the need for 3D audio is increasing.

Aim of this thesis is to investigate what production techniques could be used to immerse the listener of a 3D audio sequence more deeply compared to stereo production and what should be considered in sound design in order not to endanger the immersion of sound space creation. In order to find an answer to these questions, the work is divided into two parts.
In the first part the production steps of an audio drama are documented. The story and sound concept of the story are described broadly. The methodological work of the author is discussed in more detail, including description of the resources used for recorded or created sounds. The author's production should make an artistic contribution in the field of 3D audio, however, to focus entirely on audio, 360° video or VR content has not been used. Furthermore, audio drama offers the advantage of having no boundaries set by a screen and is the most effective medium when it comes to stimulating the imagination.[1]

The second part of the thesis is an online survey that was developed in order to explore the approach of 3D audio experts and give an overview of their used production setup, workflow and the areas of work they engage in. It was sent out via e-mail to 29 selected audio experts from Austria, Germany, Switzerland, United Kingdom, United States of America, Australia, Canada, Ireland and Norway. Results from the survey furthermore provide information about how long the audio experts have been active in their field, whether it was possible for them to transfer their experiences in the field of stereo production to 3D audio and how they see the future of 3D audio. Experts were selected not only due to their experience within various immersive media formats but also because of their research in the field of 3D audio. Out of 11 accesses, 8 respondents completed the questionnaire, 2 of them hold a PhD in spatial sound or spatial music.

[1] Rodero, “Stimulating the Imagination in a Radio Story.”

1.1 IMMERSION

The term *immersion* can be interpreted in different ways and is nowadays often used as a catch phrase in marketing. The Cambridge Dictionary defines the term, referring to media, theatre and film as "seeming to surround the audience, player, etc. so that they feel completely involved in something"[2]. The Oxford Dictionary clarifies the term as "(of a computer display or system) generating a three-dimensional image which appears to surround the user."[3] In VR, immersion is the perception of being physically present in a non-physical world in which image, sound or other stimuli offer a captivating overall environment.[4]
Auditory immersion and the issues with conflicting definitions of the term's immersion and presence in relation to the content in VR are presented in a research paper by Eaton and Lee[5]. The Research includes a survey designed to gather opinions from audio professionals and consumers on the importance of perceptual and technical auditory factors for immersion. Two types of immersion have been identified, passive immersion and active immersion. Passive immersion describes the perceptual feeling of "being there", little or no user interaction is required, and it is heavily related to the sensory perception of a virtual environment. Content for passive immersion may include music or soundscapes. Active immersion, also known as cognitive absorption, describes the immersion perceived when being involved in a task or an activity and is dependent on user interaction. Content for active immersion may include video games, simulations or professional training. "The term immersive system could be described as the inherent property of how advanced the system is at providing both passive, presence based immersion and active, cognitive absorption based immersion."[6]

Bolles[7] argues that the term immersive experience should be reserved to only refer to the movement that relates to the literal definition of immersion. It should be used to describe an experience in which one is fully submerged and, even for a moment, everything that binds one to a version of reality is forgotten.

[2] Cambridge Dictionary, s.v. "immersive."
[3] Oxford Dictionary, s.v. "immersive."
[4] Wikipedia, s.v. "Immersion (virtual reality)."
[5] Eaton and Lee, "Quantifying Factors of Auditory Immersion in Virtual Reality."
[6] Ibid.
[7] Bolles, "Defining Immersive."

Guidelines[8] when defining an experience as immersive could be:

- Submerge the participants in a 360° environment
- Digital technology can help to increase the experience (but does not necessarily have to be used)
- Interaction can enhance the experience, but is not a reason for an experience to be interpreted as immersive
- Increased engagement should always be a goal, but not the only reason for defining an immersive experience

Barrett[9] points out that listener immersion is less clearly formulated, and it means that the listener is convinced of being in a real space, yet the scales are inhuman and in reality impossible.
In immersive environments, users often report a general holistic feeling of being transported to another place. This subjective feeling tends to increase as the degree of physical immersion increases, e.g. 3D tends to offer a stronger experience than 2D visualizations, as does surround sound as opposed to stereo.[10]
Various research papers[11] [12] show a positive trend towards increasing immersion and higher emotional intensity for a growing number of playback channels. Aro[13] points out that multichannel listening causes less listening fatigue as the surround sound of multichannel listening is closer to natural hearing than stereo.

In a kind of trance with music, the threshold of belief can be lowered to watch movies and immerse oneself in the story.[14] Yet the feeling of immersion should be a tool used to serve the story, as Dyar[15] argues and furthermore says, that putting immersion before storytelling is the biggest problem facing 360° video content.

[8] Bolles, "Defining Immersive."
[9] Barrett, "Ambisonics and acousmatic space."
[10] Jane et al., "Evaluating Immersive User Experience and Audience Impact."
[11] Hahn, "Musical Emotions Evoked by 3D Audio."
[12] Lawrence, "Evaluating Immersive Experiences in Audio for Composition."
[13] Aro, "Surround Sound in Radio Drama."
[14] Sonnenschein, Sound Design, 105.
[15] Dyar, "Stuck In Spheres. "

Research by Zhang, Perkis and Arndt[16] show that "emotional immersion is significantly more immersive than spatial immersion"[17] when it comes to the sense of "being there", time perception, realism, sense of engagement, emotional aspects and sensory cues. "Spatial immersion is almost as immersive as emotional immersion in terms of attention and image motion" [18]. In regard to spatial disorientation, spatial immersion is stronger than emotional. The results of this research also show that "there are individual differences in the perception of immersion"[19].

In order to create immersive sound experiences, it might be of use to have insight in the fields of psychoacoustics[20] [21], auditory perception[22] [23] , auditory scene analysis[24], acoustic ecology[25], spatial hearing[26], the perception of spatial sound[27] [28] [29] and listener envelopment (LEV)[30] [31]. In addition, it is advisable to know about the human memory and how memory affects the perception of information.[32] Sound designers can furthermore benefit from spatial music composition techniques[33] [34] and technical ear training[35] but always have to keep in mind that the perception of everyday listeners may fundamentally vary from those of audio experts precisely because of the training they have and their involvement in the production of sound.[36]

[16] Zhang, Perkis and Arndt, "Spatial Immersion versus Emotional Immersion."
[17] Ibid.
[18] Ibid.
[19] Ibid.
[20] Parncutt, "Psychoacoustics and cognition for musicians."
[21] Lennox, Myatt, and Vaughan, "From Surround to True 3-D."
[22] Forrester, "Auditory perception and sound as event."
[23] Moore, An Introduction to the Psychology of Hearing.
[24] Bregman, "Auditory Scene Analysis."
[25] Droumeva, "Understanding immersive audio."
[26] Blauert, Spatial Hearing.
[27] Roginska and Geluso, Immersive Sound.
[28] Spors et al., "Spatial Sound With Loudspeakers and Its Perception."
[29] Kendall, "Spatial perception and cognition in multichannel audio for electroacoustic music."
[30] Blauert and Lindemann, "Auditory spaciousness."
[31] Morimoto, "Auditory spaciousness and envelopment."
[32] Ferster, "The Magical Number Seven, Plus Or Minus Two."
[33] Baalman, "Spatial Composition Techniques and Sound Spatialisation Technologies."
[34] Barrett, "Spatio-musical composition strategies."
[35] Corey, "Technical ear training."
[36] McGregor and Cunningham, "Comparative Evaluation of Radio and Audio Logo Sound Designs" 883.

1.2 OUTLINE

Chapter 2 gives a brief introduction to various spatial audio formats.
In *chapter 3* the artistic application is presented, including the story of the audio drama, the sound concept and the methodology.
Chapter 4 deals with the production environment used regarding hardware, DAW, plug-ins, software instruments and sounds.
Selected Scenes of the audio drama are presented in *chapter 5*.
Chapter 6 contains the results of a survey among immersive audio experts regarding their production setup, their experience in spatial audio and immersive media and audio formats they produce for. Furthermore, their spatial audio production workflow, knowledge in stereo production and their opinion on the future of 3D audio is included.

2 SPATIAL AUDIO FORMATS

When creating immersive audio, it is important to consider the format the content will be produced for as well as the end user's playback configuration. The production of audio content is traditionally based on the manipulation of sound objects, which can be defined as audio waveforms (audio elements) and associated parameters (metadata).[37]

To enable high-quality, bit-rate efficient distribution and flexible playback of 3D sound, the Moving Picture Experts Group (MPEG)[38] has developed the MPEG-H Audio Codec for the universal carriage of encoded 3D audio from channel-based, object-based and higher order Ambisonics (HOA)-based input.[39]

Spatial Audio Coding (SAC), used in MPEG spatial audio standards, is not a pure compression method, it rather can be considered as a technique to represent multichannel audio signals of a lower number of channels. While maintaining the spatial characteristics of the audio signals, a down-mix to mono or stereo is created. In order to reconstruct the original multi-channel audio signals at a later time, spatial parameters are stored as additional data of the down-mix audio signal.[40]

The down-mix signal needs to be encoded using an existing compression technique, such as MPEG-1 layer 3 (mp3). The spatial parameters can be regarded as side information, which allows the reconstruction of multi-channel audio signals at a later moment by enhancing the down-mix signal with spatial parameter guidance.[41]

Characteristics as well as advantages and disadvantages of various 3D audio formats will be described below.

[37] Roginska and Geluso, Immersive Sound, 244.
[38] https://mpeg.chiariglione.org/
[39] Herre et al., "MPEG-H Audio - The New Standard for Universal Spatial / 3D Audio Coding."
[40] Elfitri and Luthfi, "Reviews on Technology and Standard of Spatial Audio Coding."
[41] Ibid.

2.1 CHANNEL-BASED AUDIO

A summary of production, distribution and playback features of channel-based audio is illustrated in *figure 1*. When producing channel-based audio it is advisable to know the playback system in advance, as each output controls a separate channel. In regard to production, sound objects are first mixed, including metadata such as panning, level, equalization and reverberation, and then rendered as individual channel tracks. These can be distributed as streams or via electrical data storage media (e.g. DVD, Blue-ray Disc) and played back in the format selected during production, e.g. 5.1.[42]

The MPEG Surround standard represents channel-based audio signals as mono, stereo or 5.1 down-mix signals and is compatible with any older audio codec for encoding down-mixed signals. Its main use is in audio broadcasting or teleconference.[43]

The 5.1 channel sound system has been specified in the International Telecommunication Union (ITU) Recommendation ITU-R BS.775 to enhance the directional stability of the frontal sound image and the feeling of spatial reality for e.g. ambience. The system is widely used as part of digital broadcasting and the recommendation defines the loudspeaker arrangement with regard to horizontal angles from centre, height and inclination.[44]

However, despite affordability, according to the ITU recommendation, small multi-channel home audio systems such as the 5.1 channel sound system, can rarely be placed in the end user's listening room due to various reasons such as room characteristics.

[42] SCHOEPS Mikrofone, "Florian Camerer (ORF)." YouTube video, 1:00:46, April 18, 2018, https://www.youtube.com/watch?v=osZ842Zaj5Q.

[43] Elfitri and Luthfi, "Reviews on Technology and Standard of Spatial Audio Coding."

[44] International Telecommunication Union (ITU). "Multichannel sound technology in home and broadcasting applications."

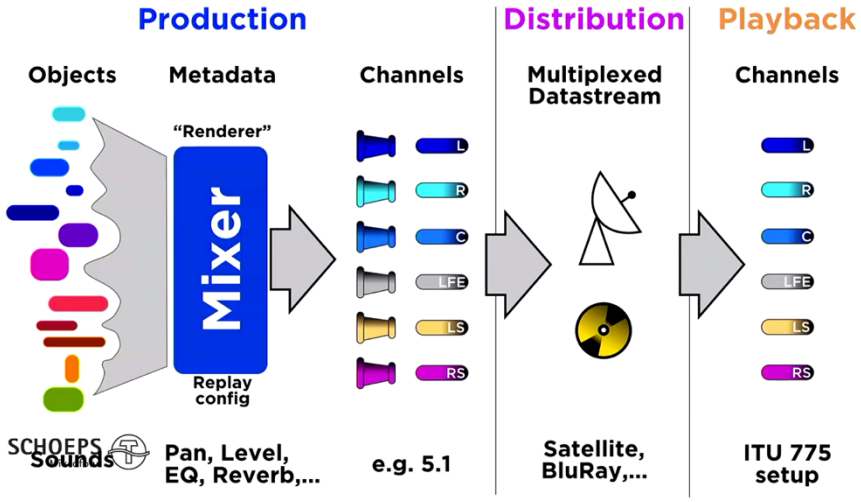


Fig. 1. Channel-based audio - production, distribution & playback

Screenshot of SCHOEPS Mikrofone, "Florian Camerer (ORF)." YouTube video, 9:13, modified

In regard to production, sound objects are first mixed, including metadata such as panning, level, equalization and reverberation, and then rendered as individual channel tracks. These can be distributed as streams or via electrical data storage media (e.g. DVD, Blue-ray Disc) and played back in the format selected during production, e.g. 5.1.[45]

2.2 OBJECT-BASED AUDIO

When producing object-based audio, sound objects and metadata are stored and distributed separately e.g. as stream or on electrical data storage media (e.g. DVD, Blue-ray Disc). The playback renderer creates a mix optimized for the respective format (e.g. 5.1, 9.1, 4.0, binaural).[46]

Using the MPEG Spatial Audio Object Coding (SAOC) standard, the encoder side can capture every audio object that might be considered for adjustment in the decoder side. Object parameters and metadata, such as the position of each audio source, are thus generated and transmitted to the decoder.[47]

[45] SCHOEPS Mikrofone, "Florian Camerer (ORF)." YouTube video, 1:00:46, April 18, 2018, https://www.youtube.com/watch?v=osZ842Zaj5Q.
[46] Ibid.
[47] Elfitri and Luthfi, "Reviews on Technology and Standard of Spatial Audio Coding."

An advantage of the system is that the end user is able to adjust the mix, e.g. changing the language or the volume of the narrator. Disadvantages are the demand of cost-intensive renderers and that large amounts of data need to be transferred.[48] Production, distribution and playback of object-based audio are outlined in *figure 2*.

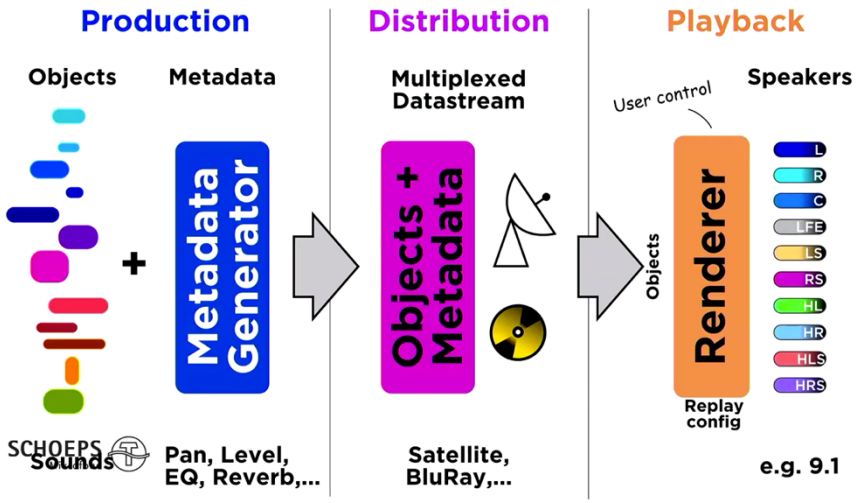


Fig. 2. Object-based audio - production, distribution & playback

Screenshot of SCHOEPS Mikrofone, "Florian Camerer (ORF)." YouTube video, 36:57, modified

When producing object-based audio, sound objects and metadata are stored and distributed separately e.g. as stream or on electrical data storage media (e.g. DVD, Blue-ray Disc). The playback renderer creates a mix optimized for the respective format (e.g. 5.1, 9.1, 4.0, binaural).[49]

A hybrid version of channel-based and object-based audio is used for Dolby Atmos[50] technology. Based on the replay configuration, some of the sound objects used are processed in the same manner as channel-based audio. These so-called beds contain e.g. ambience tracks and music. Individual sounds are transferred separately as objects with the metadata and rendered for the individual playback system. More than a hundred single objects can be processed in Dolby Atmos.[51]

Production, distribution and playback of the hybrid version of channel-based and object-based audio is shown in *figure 3*.

[48] Elfitri and Luthfi, "Reviews on Technology and Standard of Spatial Audio Coding."
[49] SCHOEPS Mikrofone, "Florian Camerer (ORF)." YouTube video, 1:00:46, April 18, 2018, https://www.youtube.com/watch?v=osZ842Zaj5Q.
[50] https://www.dolby.com/us/en/brands/dolby-atmos.html
[51] SCHOEPS Mikrofone, "Florian Camerer (ORF)." YouTube video, 1:00:46, April 18, 2018, https://www.youtube.com/watch?v=osZ842Zaj5Q.

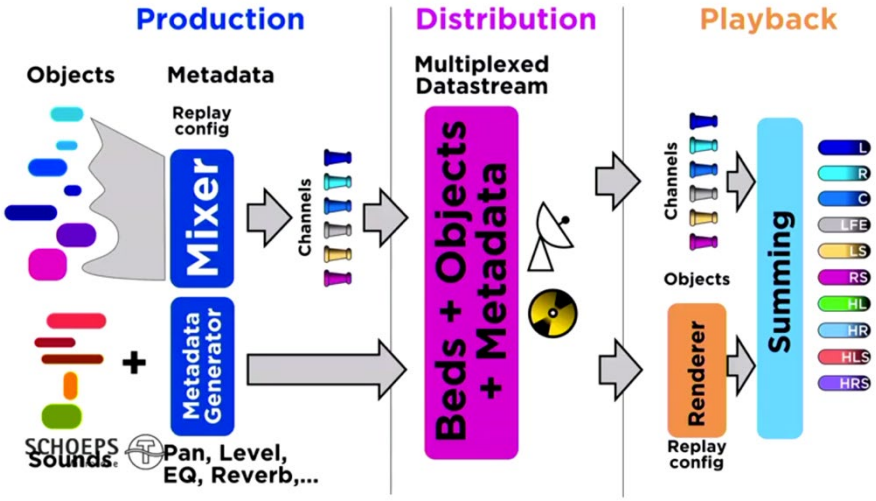


Fig. 3. Hybrid version of channel-based and object-based audio - production, distribution & playback

Screenshot of SCHOEPS Mikrofone, "Florian Camerer (ORF)." YouTube video, 40:02, modified

A hybrid version of channel-based and object-based audio is used for Dolby Atmos[52] technology. Based on the replay configuration, some of the sound objects used are processed in the same manner as channel-based audio. These so-called beds contain e.g. ambience tracks and music. Individual sounds are transferred separately as objects with the metadata and rendered for the individual playback system. More than a hundred single objects can be processed in Dolby Atmos.[53]

2.3 SCENE-BASED AUDIO

The attempt to physically reconstruct a sound field is called a holophonic approach. The following chapter introduces the two main representatives of holophonic systems, Ambisonics and Wave Field Synthesis (WFS).[54]

2.3.1 AMBISONICS

Ambisonics can be described as a mathematically motivated procedure.[55] "It is based on the representations of the sound field excitation" [56]. The sound scene is described spatially, and so-called spherical harmonics represent the richness of detail on the surface of a sphere in

[52] https://www.dolby.com/us/en/brands/dolby-atmos.html
[53] SCHOEPS Mikrofone, "Florian Camerer (ORF)." YouTube video, 1:00:46, April 18, 2018, https://www.youtube.com/watch?v=osZ842Zaj5Q.
[54] Sontacchi, "Dreidimensionale Schallfeldreproduktion für Lautsprecher- und Kopfhöreranwendungen."
[55] Ibid.
[56] Frank, Zotter, and Sontacchi, "Producing 3D Audio in Ambisonics."

space. In the production process sound objects, including metadata such as level and equalization, are first mixed and then encoded into the Ambisonic domain. The stored information is distributed in the Ambisonics format as an audio file or via streaming.[57] Production, and distribution of Ambisonics is illustrated in *figure 4*.

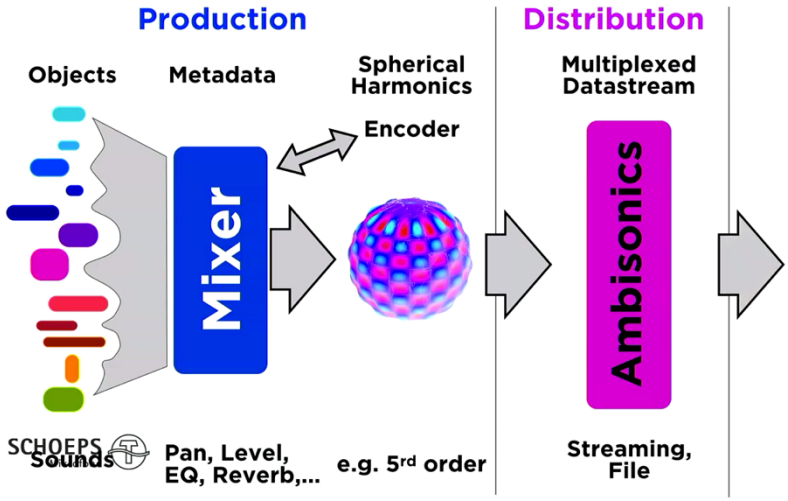


Fig. 4. Ambisonics - production & distribution

Screenshot of SCHOEPS Mikrofone, "Florian Camerer (ORF)." YouTube video, 13:43, modified

In the production process sound objects, including metadata such as level and equalization, are first mixed and then encoded into the Ambisonic domain. The stored information is distributed in the Ambisonics format as an audio file or via streaming.[58]

The spatial resolution and the minimum density of the loudspeakers required for playback are defined by the maximum order N, the number of channels per (bus) track is determined by $(N+1)^2$.[59] *Table 1* shows the channel count of 1st to 7th order Ambisonics.

As the track count of 3rd and 7th order Ambisonics is exactly divisible by 8, these orders are more efficient with regard to playback equipment. Most audio interfaces have inputs and outputs in groups of 8. For 5th order Ambisonics at least 4 channels of a new input or output group are occupied, whereby at 4th or 6th order Ambisonics only one channel is occupied.

[57] SCHOEPS Mikrofone, "Florian Camerer (ORF)." YouTube video, 1:00:46, April 18, 2018, https://www.youtube.com/watch?v=osZ842Zaj5Q.
[58] SCHOEPS Mikrofone, "Florian Camerer (ORF)." YouTube video, 1:00:46, April 18, 2018, https://www.youtube.com/watch?v=osZ842Zaj5Q.
[59] Frank, Zotter and Sontacchi, "Producing 3D Audio in Ambisonics."

Ambisonics order	Track count
1st order = first-order Ambisonics (FOA)	4
2nd order = higher-order Ambisonics (HOA)	9
3rd order = higher-order Ambisonics (HOA)	16
4th order = higher-order Ambisonics (HOA)	25
5th order = higher-order Ambisonics (HOA)	36
6th order = higher-order Ambisonics (HOA)	49
7th order = higher-order Ambisonics (HOA)	64

Table 1. Channel count of 1st to 7th order Ambisonics

Ambisonics order and its corresponding track count.

For playback, a decoder is used to reproduce the sound scene correctly on the available sound system (5.1, Auro 9.1, binaural).[60] One of the playback methods of Ambisonics is illustrated in *figure 5*. The reproduction and decoding of audio signals in random loudspeaker arrangements was a challenge during the introduction of Ambisonics. With the introduction of the All-Round Ambisonic Decoding (AllRAD)[61] method, which decodes to an optimal virtual t-design[62] loudspeaker arrangement by sampling, there is now a flexible procedure that can handle irregular layouts very well. Signals of the virtual loudspeakers are mapped to the real loudspeakers using vector-base amplitude panning (VBAP).[63] A comparison of amplitude panning approaches on ITU BS.2051 loudspeaker layouts with height can be found in Romanov's master thesis[64], in which he focuses on amplitude panning methods for direction rendering of sound objects and their realization on standard ITU loudspeaker layouts. A research paper by Frank[65] discusses the effect of "quality elements (e.g. reproduction room, number and equalization of loudspeakers, order weighting, and decoder design) on perceived quality features (e.g. localization, source width, and coloration)"[66].

[60] SCHOEPS Mikrofone, "Florian Camerer (ORF)." YouTube video, 1:00:46, April 18, 2018, https://www.youtube.com/watch?v=osZ842Zaj5Q.
[61] Zotter and Frank, "All-Round Ambisonic Panning and Decoding."
[62] Zotter, Frank and Sontacchi, "The Virtual T-Design Ambisonics-Rig Using VBAP."
[63] Frank, Zotter and Sontacchi, "Producing 3D Audio in Ambisonics."
[64] Romanov, "Comparison of Amplitude Panning Approaches on ITU BS.2051 Loudspeaker Layouts with Height."
[65] Frank, "How to make Ambisonics sound good."
[66] Ibid.

Ambisonics offers the advantage that one does not have to know where the production is reproduced and therefore it is independent of the renderer during production. It is especially favourable for VR content as the sound field can be easily manipulated and e.g. rotated or augmented. Contrary, Ambisonics, if not scaled to 2D, is not really suitable for home use as most likely the required number of loudspeakers is not available. However, a binaural render can be played back on headphones.[67]

Fundamentals of binaural technology can be found in this article[68], a 3D Ambisonic based binaural sound reproduction system is presented in this research paper[69] and binaural sound reproduction from both a technical and a perceptual perspective is discussed in Pike's PhD thesis[70].

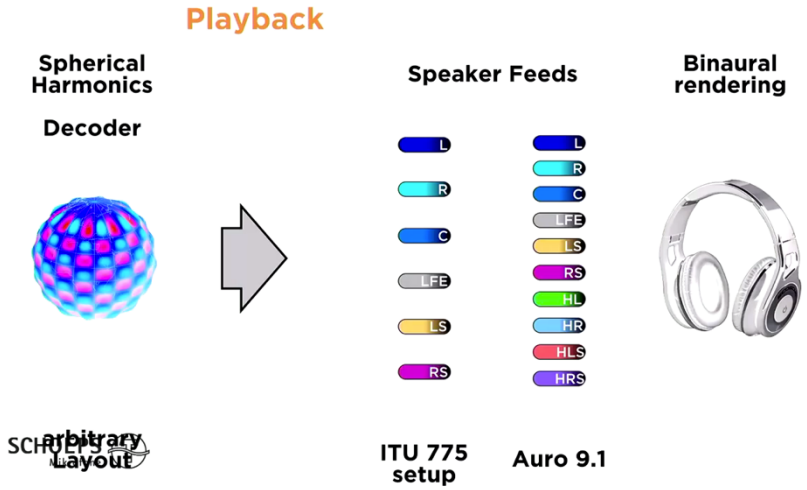


Fig. 5. Ambisonics – playback

Screenshot of SCHOEPS Mikrofone, "Florian Camerer (ORF)." YouTube video, 14:37, modified

For playback, a decoder is used to reproduce the sound scene correctly on the available sound system (5.1, Auro 9.1, binaural).[71]

Detailed information about 3D audio theory for recording, studio production, sound reinforcement and VR can be found in Franz Zotter's and Matthias Frank's book *Ambisonics*.[72]

[67] SCHOEPS Mikrofone, "Florian Camerer (ORF)." YouTube video, 1:00:46, April 18, 2018, https://www.youtube.com/watch?v=osZ842Zaj5Q.
[68] Møller, "Fundamentals of Binaural Technology."
[69] Noisternig, "A 3D Ambisonic based Binaural Sound Reproduction System."
[70] Pike, "Evaluating the Perceived Quality of Binaural Technology."
[71] SCHOEPS Mikrofone, "Florian Camerer (ORF)." YouTube video, 1:00:46, April 18, 2018, https://www.youtube.com/watch?v=osZ842Zaj5Q.
[72] Zotter and Frank, Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality.

2.3.2 WAVE FIELD SYNTHESIS

The second representative of scene-based audio is the Wave Field Synthesis (WFS) which is a physical modelling of the sound field with attempts of a global wave front reconstruction.[73] Loudspeaker arrays are used to generate a complete sound field in the listening zone, which is identical to a corresponding real sound event.[74]

The concept of WFS is based on Huygen's Principle. Which states that a propagating wave front of a primary source can be synthesized by an infinite number of secondary sources that are fed by the signal emitted from the primary source. The superposition of all secondary source signals leads to an exact copy of the primary source wave front.[75] The MPEG-H 3D Audio standard supports audio playback systems such as 10.2 and 22.2, as well as a larger number of speakers for (WFS) technology.[76]

This system is inapplicable for non-professional end users as most likely they will not have the contingency to set up such an extensive audio system at home. Yet it might be of use for events that require a large audio set-up such as audio conferences as well as for research institutions and universities.[77] However, high transmission capacity and high computing power are required. [78]

More information about 3D audio, multi-channel and binaural sound theory and applications can be found in the Audio Engineering Society's (AES) book *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio*[79].

[73] Sontacchi, "Dreidimensionale Schallfeldreproduktion für Lautsprecher- und Kopfhöreranwendungen."
[74] Theile, Wittek and Reisinger, "Wellenfeldsynthese-Verfahren."
[75] Roginska and Geluso, Immersive Sound, 312.
[76] Ibid.
[77] SCHOEPS Mikrofone, "Florian Camerer (ORF)." YouTube video, 1:00:46, April 18, 2018, https://www.youtube.com/watch?v=osZ842Zaj5Q.
[78] Theile, Wittek and Reisinger, "Wellenfeldsynthese-Verfahren."
[79] Roginska and Geluso, Immersive Sound.

3 ARTISTIC APPLICATION

The audio drama “Spinnenbank” is based on a text written and narrated by language artist John Sauter and has a total length of 39 minutes. The intention was to create both realistic and surreal sound scenes while always keeping the focus on the story line. This chapter depicts the conceptual and methodological approaches.

3.1 STORY

“Spinnenbank” is a story about two people experiencing a strange journey through a city. It begins with a ride in a tramway, which is interrupted by a short flashback of the main character and his girlfriend. Afterwards, the listener is back in the tram again until the two characters leave at the last stop. The tram journey is followed by a very intense and bizarre trip in a park where the characters are gradually immersing into a surreal world. Throughout the audio drama, the listener is confronted with the question of whether the experience is still real or if it only takes place in the mind of the main character.

3.2 SOUND CONCEPT

The first step was to read through the text and split it into sound scenes. Then, keywords were tracked down, as Sonnenschein[80] recommends, and categories were defined to group them in.

Those are:

- *Ambience sounds*: defining the place where the scene takes place
- *Sound effects*: describing objects and actions
- *Emotions*: keywords that give clues to emotions

[80] Sonnenschein, Sound Design, 2.

The keywords for the three categories were each marked in a different colour and a scene list was created, generating an overview of the entire audio drama and a framework of what sounds could be used. Based on the scene list sample libraries were searched for matching sounds.

3.3 METHODOLOGY

3.3.1 NARRATIVE VOICE

At a recording session the author John Sauter read through the audio drama three times and in several editing sessions a single track was created. The duration of the vocal recording determined the length of the overall project. A study by Emma Rodero[81] shows that a dramatised story would "achieve a greater level of stimulation of the imagination than a radio story with a narrated presentation structure" [82].

The narrator's voice was placed in the front of the mix, as is usually done in movies, to ensure good comprehensibility. Ambience sounds, sound effects and musical elements were placed in the background of the mix, according to the taste of the author.

3.3.2 AMBIENCE SOUNDS

Usually, the first step of an audio project is to define the given space with basic ambience layers. This process also provides an idea of what sounds could be missing in each scene. Although there is probably plenty of room for additional sounds in a 3D audio environment than in a limited stereo image, even in 3D audio, less is sometimes more.

Most of the sound scenes would be designed as realistic as possible, capturing sound images form the real world to use within the immersive environment and creating an illusory world.[83]

[81] Rodero, "Stimulating the Imagination in a Radio Story."
[82] Ibid.
[83] Lennox, Myatt, and Vaughan, "From Surround to True 3-D."

The listener should be able to believe what he or she is listening to. The term realism is not interchangeable with the term real as a composed sound scene will probably never replace reality.[84] Still from an artistic point of view, one could try to get as close as possible.
Multichannel ambience tracks (Ambisonics, 5.1) were used either individually or used in combination to create a wider sound image in order to place one multichannel layer on the upper half of the hemisphere and the other on the lower half of the hemisphere. In other words, ambience tracks with a larger proportion of e.g., bird and wind sounds should be perceived mainly from above and ambience tracks with a larger proportion of e.g., human and city traffic sounds should be perceived from all around.

Table 2 shows the multi-channel ambience sounds used throughout the timeline. Furthermore, it provides information on where the spatial field recordings were performed, and which microphones were used. For more information on the microphones used, refer to chapter 4.1.3. In chapter 4.2 and 4.4 detailed information on the samples used can be found.

Time code	*Ambisonics recording (microphone)*	*5.1 ambience sounds (source)*
0:00	Inside a tramway (SoundField SPS200[85])	-
0:37	In the woods (Oktava MK-4012[86])	City park sample (Logic Pro X sample library[87])
0:49	Inside a tramway (SoundField SPS200 and Sennheiser Ambeo VR[88])	-
1:07	In the woods (Oktava MK-4012)	City park sample (Logic Pro X)
2:12	-	City night and town after rain samples (Logic Pro X)
4:42	-	Room tone, room lights, two different rain and thunder samples (Logic Pro X)
6:10	-	City night and town after rain samples (Logic Pro X)
6:46	Inside a tramway (SoundField SPS200)	-
7:27	In the woods (Oktava MK-4012)	City park sample (Logic Pro X)
7:51	Near a tram stop (Sennheiser Ambeo)	City park sample (Logic Pro)

[84] Lennox, Myatt, and Vaughan, "From Surround to True 3-D."
[85] https://www.soundfield.com/#/products/sps200
[86] http://www.oktava-shop.com/Small-and-medium-diaphragm-condenser-mics/Oktava-MK-012-01-12.html
[87] https://www.apple.com/logic-pro/plugins-and-sounds/
[88] https://en-us.sennheiser.com/microphone-3d-audio-ambeo-vr-mic

Time code	*Ambisonics recording (microphone)*	*5.1 ambience sounds (source)*
8:03	Inside a tramway (SoundField SPS200) and in the woods (Sennheiser Ambeo)	City park sample (Logic Pro X)
8:29	In the woods (Sennheiser Ambeo)	City park sample (Logic Pro X)
11:28	In the woods (SoundField SPS200)	City park sample (Logic Pro X)
18:57	-	Various dark drone samples (Logic Pro X)
26:28	Nature sample by Free to Use Sounds[89] (Zoom H3-VR[90])	Various dark drone and town after rain samples (Logic Pro X)
27:55	Nature sample by Free to Use Sounds (Zoom H3-VR)	Town after rain sample (Logic Pro X)
37:10	-	Dark Drone Sample (Logic Pro X)

Table 2. Multi-channel ambience sounds used throughout the audio drama

The exact timecode is given and whether the ambient tracks used are spatial field recordings (including information about microphones used) and/or 5.1 ambience sounds.

3.3.3 MUSICAL ELEMENTS

As the story progresses, gradually more musical elements are introduced. Like in movies they were used to conduct the relaxation-tension cycle and take the listener on a ride[91]. Emotions can be communicated with the help of basic sound characteristics according to the valence-arousal model.[92] Very low sound effects course subtle associations and interrelations of emotion and often play a decisive dramaturgical part.[93] The use of deep drone sounds creates tension and occasionally discomfort. As it is difficult for the human auditory system to localise low-frequency sources, a sense of immersiveness can be achieved.[94]

The density of the musical elements increases with time and culminates in a surreal sound scene without real-world ambient noise (*time code value: 19:04*). The loss of control and logic is supported by the music and leads to an illusion-reality.[95] There is not only a loss of sense of space but also a change of perspective. This change of the point-of-listening[96] into the inside of a cage creates an oppressive feeling.

[89] https://freetousesounds.com/
[90] https://www.zoom-na.com/products/field-video-recording/field-recording/zoom-h3-vr-handy-recorder
[91] Sonnenschein, Sound Design, 105.
[92] Görne, Sounddesign, 201.
[93] Forrester, "Auditory perception and sound as event: theorising sound imagery in psychology."
[94] Polotti, Sound to Sense - Sense to Sound, 431-32.
[95] Sonnenschein, Sound Design, 105.
[96] Beck, "Point-of-listening in radio plays."

3.3.4 SOUND EFFECTS

Lastly, recorded Foley sounds and sound effects from libraries were added to the project. A study by Emma Rodero[97] shows that sound effects are effective in stimulating the creation of mental images in the listener's mind and thereby enhance the listener's attention. In presented audio drama, Foley sound effects were performed by the author and audio files from sound libraries were combined in several layers to create independent sounds. To recreate reality and produce realistic hard effects, various small sound elements can be used.[98] According to Marshall McGee[99], layering is the most important tool for a sound designer when it comes to the creation of sound effects.

Table 3 provides information about which sound effects have been used in accordance to the timeline of the audio drama and whether those are recorded Foley sounds or sound effects from libraries. For more detailed information on the samples see chapter 4.5.

Time code	*Foley sounds*	*Sound effects (source)*
0:39	Keys, sunglasses, wallet	-
1:03	Keys, sunglasses, wallet	-
1:48	Pair of drumsticks	-
1:07	Shoes on gravel	-
4:39	-	Erykah Badu's song *Certainly* played in the radio. A radio loudspeaker was simulated with Logic Pro X's convolution plugin Space Designer[100]
4:43	-	Wine pouring sample (Tonebenders[101])
5:17	Kitchen sink, running water	-
6:20	-	Walking on gravel sample (MusicRadar[102])
9:13	Rustling jacket, zipper, keys	-
9:45	-	Siren: Layered sound effect (Logic Pro X sample library[103], Sonniss[104] and MusicRadar[105])

[97] Rodero, "See It on a Radio Story."
[98] Averese, Post Sound Design, 75.
[99] Marshall McGee, "Layering - the most important tool for a sound designer." YouTube video, 4:31, August 29, 2016, https://www.youtube.com/watch?v=0biAgn2ct0A.
[100] https://support.apple.com/kb/PH27798?locale=en_GB
[101] http://tonebenderspodcast.com/
[102] https://www.musicradar.com/
[103] https://www.apple.com/logic-pro/plugins-and-sounds/
[104] https://sonniss.com/
[105] https://www.musicradar.com/

Time code	*Foley sounds*	*Sound effects (source)*
10:02	Keys	Body sounds: Layered sound effect (Logic Pro X, Sonniss and MusicRadar)
10:17	-	Groan male sample (Logic Pro X sample library edited with Logic Pro X's Vocal Transformer[106])
10:41	-	Logic Pro X's Space Designer was used to convolve a groan male sample with birds, dog, lion and walrus samples. All samples were taken from the Logic Pro X sample library
11:29	-	Rustling leaves (Tonebenders) and a Foley chair sample (Audio Shade[107])
11:44	-	Inhale and exhale sample (Logic Pro X)
18:50 – 26:20 (surreal sound scene)	Eyes: Pop sounds with fingers and mouth	Various woosh, swell, drone, movement, water and waterdrop samples (Logic Pro X). Cage doors: layered sound effect (Logic Pro X, 99Sounds[108] and MusicRadar)
27:39	Hands grinding in gravel	-
28:15	Rustling jacket	-
28:40	Hands grinding in gravel	-
30:36	-	Drag away: Layered sound effect (Logic Pro X and Audio Shade)
31:21	Rustling jacket, Shoes on gravel	-
31:55	-	Rustling leaves, cracking branches: Layered sound effect (Logic Pro X sample library, Tonebenders, Sonniss and Samples From Mars[109])
32:51	-	Rustling leaves sample (Tonebenders)
33:41	-	Rustling leaves, cracking branches, crushing bones: Layered sound effect (Logic Pro X, Tonebenders, Sonniss and Samples From Mars)
34:37	-	Rustling leaves, cracking branches: Layered sound effect (Logic Pro X, Tonebenders, Sonniss and Samples From Mars)
35:08	-	Running on gravel sample (MusicRadar)

Table 3. Foley sounds and sound effects from libraries used throughout the audio drama

The exact timecode is given and whether the sound tracks used are Foley recordings and/or sound effects from libraries.

[106] https://support.apple.com/kb/PH27374?locale=en_GB
[107] https://www.audioshade.com/
[108] http://99sounds.org/
[109] https://samplesfrommars.com/

4 PRODUCTION ENVIRONMENT

The Institute of Electronic Music and Acoustics (IEM) at the University of Music and Performing Arts Graz carries out a lot of research in Ambisonics and also provides sound studios and audio equipment (for staff and students / for research as well as applied sound projects). Accordingly, the 3D audio drama "Spinnenbank" was produced in 5th order Ambisonics. The following chapter deals with the technical equipment used.

4.1 HARDWARE

4.1.1 WORKSTATION

For the production, a custom-made workstation from Z | TOOLS[110] was used, equipped with an 8-Core Intel Xeon E5 processor (up to 5 Ghz), 64 GB 2933 MHz DDR4 RAM, 1 TB SSD, 4 TB HD (engineered for 24/7 workloads) and a Radeon RX 570 4096 MB graphics card. The operating system used is macOS High Sierra[111] (version: 10.13.6).

4.1.2 AUDIO INTERFACES AND LISTENING ENVIRONMENT

An RME Fireface 802112 audio interface was used for the voice recording at an audio studio at the FH JOANNEUM University of Applied Sciences[113]. The Focusrite Scarlett 2i2 2nd Gen[114] audio interface was used for the playback via beyerdynamic DT-770 Pro 250 Ohm[115] headphones and a MADIface USB[116] audio interface was used for playback at the IEM CUBE[117] and the IEM Produktionsstudio[118].

[110] http://ztools.at/
[111] https://en.wikipedia.org/wiki/MacOS_High_Sierra
[112] http://www.rme-audio.de/en/products/fireface_802.php
[113] https://www.fh-joanneum.at/sound-design/master/en/my-studies/infrastructure/
[114] https://focusrite.com/usb-audio-interface/scarlett/scarlett-2i2
[115] https://global.beyerdynamic.com/dt-770-pro.html
[116] https://www.rme-audio.de/en/products/madiface_usb.php
[117] http://iem.at/~zmoelnig/publications/cube_en/
[118] https://iem.kug.ac.at/services/jahresberichte-wurde-2013-eingestellt/200001/produktionsstudio.html

4.1.3 MICROPHONES AND FIELD RECORDER

First-order Ambisonic microphones, namely the SoundField SPS200[119], Sennheiser Ambeo VR[120] and Oktava MK-4012[121] were used in combination with the Zoom H6 recorder[122] to perform spatial field recordings. Although of a lower spatial resolution, first-order recordings capture sound images of the real world in sense of environment and perspective.[123]
Comparisons of first-order Ambisonic microphone arrays are discussed in detail in various research papers[124] [125] [126].
The X/Y capsule supplied with the Zoom H6 recorder was utilized for Foley recording and voice was recorded using the AKG C414 XLS[127] condenser microphone and a pair of AKG C391 B[128] condenser microphones with a pop filter placed between the microphone and the narrator.

4.2 DIGITAL AUDIO WORKSTATION

The digital audio workstation (DAW) Logic Pro X[129] was used for the recording and mixing the narrator's voice as well as for audio editing and layering sound effects. Furthermore, proper 5.1 ambience sounds were selected from the DAWs sample library. Voice tracks were recorded at a 48 kHz sample rate and a 24 bit rate. All tracks, recorded as well as those created in Logic Pro X, were rendered at 44.1 kHz for further processing. The main project was mixed in the DAW REAPER[130], which supports up to 64 channels per track in order to create higher order Ambisonic projects (up to 7th order).[131] With a sample rate of 44.1 kHz and a 24 bit rate. The 5th order Ambisonic tracks were rendered as multi-channel files in *.wav* file format that comply with the AmbiX[132] format.

---

[119] https://www.soundfield.com/#/products/sps200
[120] https://en-us.sennheiser.com/microphone-3d-audio-ambeo-vr-mic
[121] http://www.oktava-shop.com/Small-and-medium-diaphragm-condenser-mics/Oktava-MK-012-01-12.html
[122] https://www.zoom-na.com/products/field-video-recording/field-recording/h6-handy-recorder
[123] Barrett, "Ambisonics and acousmatic space: a composer's framework for investigating spatial ontology."
[124] Kurz, Pfahler and Frank, "Comparison of first-order Ambisonic microphone arrays."
[125] Bates et al., "Comparing Ambisonic Microphones: Part 1"
[126] Bates et al., "Comparing Ambisonic Microphones - Part 2"
[127] https://www.akg.com/Microphones/Condenser%20Microphones/C414XLS.html
[128] https://www.akg.com/Microphones/Condenser%20Microphones/2442H00010.html
[129] https://www.apple.com/logic-pro/
[130] https://www.reaper.fm/
[131] IEM Plug-in Suite. "Can I use the plug-ins with every DAW?"
[132] Nachbar, "ambix - A Suggested Ambisonics Format."

REAPER templates for 3rd, 5th, and 7th order Ambisonic productions can be obtained from the IEM Plug-in Suite website[133]. This presented project was created using the 5th order Ambisonic production template and adapted to the project's specific requirements. *Figure 6* shows the routing structure of the author's REAPER session.

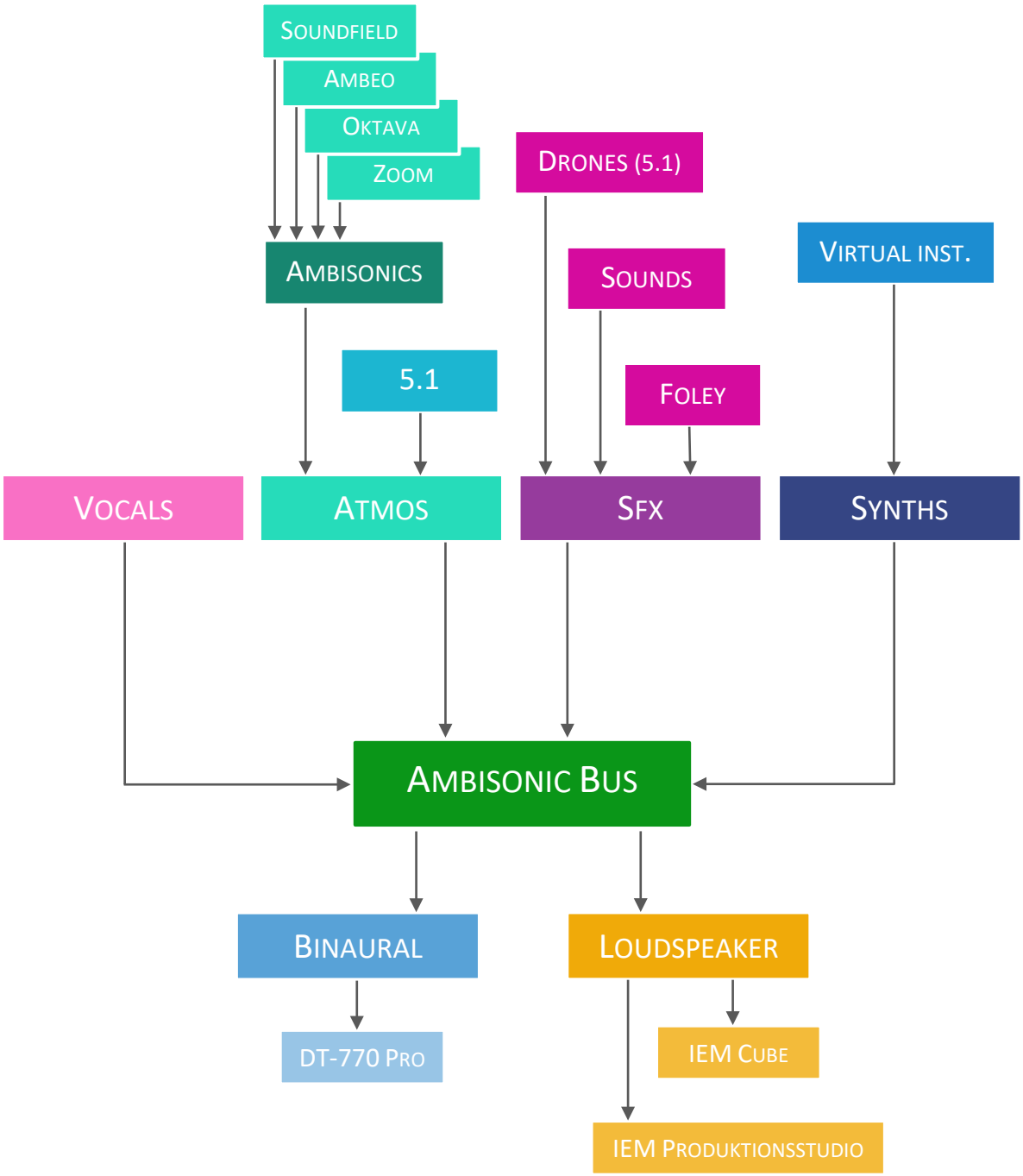


Fig. 6. REAPER routing structure

Individual audio tracks were grouped and routed to the Ambisonic Bus for binaural rendering or loudspeaker playback.

[133] https://plugins.iem.at/docs/tutorial_basicrouting/#step-5-test-your-routing-and-save-as-template

Aro[134] points out that time in professional production is too limited to perform different processing tasks in different software tools. Working with only a single software program may have the advantage of not having to jump back and forth between different applications as rendering audio files and embedding them in other software may take up valuable time. On the other hand, it may bear limitations in designing sounds, because different software tools may have different strengths for audio processing that may all be applicable in one project.

4.3 PLUG-INS

Logic Pro X plug-ins[135] were used to apply equalization, de-essing, dynamic processing and time-based processing of the voice track as well as some sound effects. In REAPER, plug-ins from the IEM plug-in Suite[136], the ambiX - Ambisonic plug-in suite[137] and the mcfx – multichannel audio plug-in suite[138] were utilized for Ambisonic encoding and decoding as well as equalization, dynamic processing, time-based processing and metering. Furthermore, the SoundField by RØDE plug-in[139], the AMBEO A-B Converter[140] and the SPARTA Array2SH141 plug-in was used to encode A-format recordings, starting out from the first-order Ambisonic microphones, into B-format. A demo version of the Harpex[142] plug-in was used to convert the 1st order recording to 3rd order Ambisonics and create a so-called upmix. These plug-ins were not used for the main project, but for comparison only, which is described in more detail in chapter 5. In addition, the ReaEQ[143] plugin provided by REAPER was used to apply equalization on a few tracks before they were encoded into the Ambisonic domain.

[134] Aro, “Surround Sound in Radio Drama.”
[135] https://www.apple.com/logic-pro/plugins-and-sounds/
[136] https://plugins.iem.at/
[137] http://www.matthiaskronlachner.com/?p=2015
[138] http://www.matthiaskronlachner.com/?p=1910
[139] https://en.rode.com/soundfieldplugin
[140] https://en-hk.sennheiser.com/ambeo-downloads
[141] http://research.spa.aalto.fi/projects/sparta_vsts/plugins.html
[142] https://harpex.net/
[143] https://www.reaper.fm/reaplugs/

4.3.1 IEM PLUG-IN SUITE

The **Stereo**Encoder is used to encode, foley sounds and sound effects, as well as drones and musical elements into the Ambisonic domain. Foley sounds and sound effects are placed in front with a stereo width of 90° too or, alternatively, placed around the listener, according to the taste of the author. On these tracks an automation of the azimuth, elevation and stereo width parameters is applied. Musical elements and drone sounds are placed either at the side or at the back, in order to avoid interference with the voice track and sound effects. The stereo width is partially automated.

The **Multi**Encoder is utilized to encode all multichannel ambience tracks (Ambisonics, 5.1) and 5.1 drone sounds into the Ambisonic domain. First-order Ambisonic microphone recordings were not encoded into B-format, but the individual capsule signals were encoded in 5th order Ambisonics in the direction of the microphone capsules (see *figure 9*). A track is created for each of the four capsule signals and routed as "Parent send" to a group track on which the **Multi**Encoder is inserted (see *figure 10*). The *left front* signal is panned 100% to the left, routed as parent channels 1-2 and placed with an azimuth of 45° at the **Multi**Encoder plug-in. The *right front* signal is panned 100% to the right, routed as parent channels 1-2 and placed with an azimuth of -45° at the **Multi**Encoder plug-in. The *left back* signal is panned 100% to the left, routed as parent channels 3-4 and placed with an azimuth of 135° at the **Multi**Encoder plug-in. The *right back* signal iswas panned 100% to the right, routed as parent channels 3-4 and placed with an azimuth of -135° at the **Multi**Encoder plug-in. Due to this routing, with hard panning to the left or right, each individual capsule signal reaches the group track as a mono signal (see *figure 11*). The elevation parameter of the individual tracks is determined differently, in accordance to the author's taste.

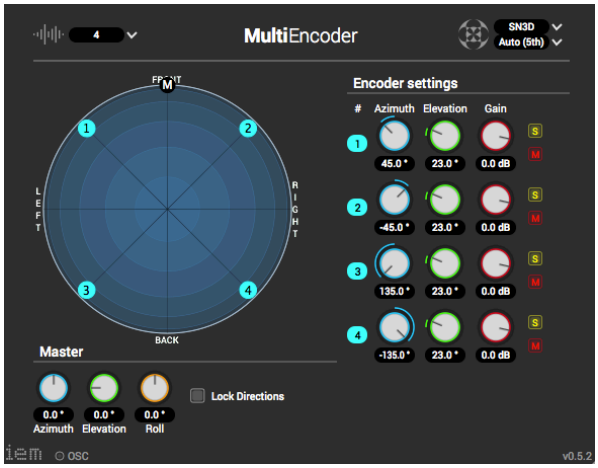


Fig. 7. Screenshot of the author's settings of the MultiEncoder, utilized to encode first-order Ambisonic microphone recordings to the 5th order Ambisonic domain.

Fig. 9. Screenshot of the author's REAPER track structure for individual capsule signals

Both left front and left back signal are each panned 100% to the left. The right front and the right back signal are each panned 100% to the right.

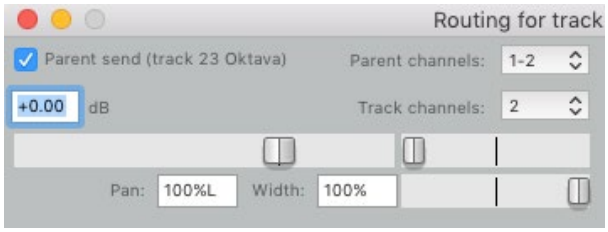


Fig. 8. Screenshot of the author's REAPER routing for parent channels

Both left front and right front signal are each routed as parent channels 1-2. The left back and the right back signal are each routed as parent channels 3-4.

The routing for the 5.1 ambience and drone tracks is done in accordance to the same principle as the Ambisonics tracks. However, 6 individual signals (L, C, R, Ls, Rs, LFE) are encoded in 5th order Ambisonics (see *figure 12*). The left (L) signal is panned 100% to the left, routed as parent channels 1-2 and placed with an azimuth of 45° in the **Multi**Encoder plug-in. The center (C) signal is panned 100% to the right, routed as parent channels 1-2 and placed with an azimuth of 0° in the **Multi**Encoder plug-in. For a selection of group tracks, i.e. thunder, dark drones, the center channel is muted to avoid getting too much signal from the front. The right (R) signal is panned 100% to the left, routed as parent channels 3-4 and placed with an azimuth of -45° in the **Multi**Encoder plug-in. The left side (Ls) signal is panned 100% to the right, routed as parent channels 3-4 and placed with an azimuth of 135° in the **Multi**Encoder plug-in. The right side (Rs) signal is panned 100% to the left, routed as parent channels 5-6 and placed with an azimuth of -135° in the **Multi**Encoder plug-in. The low-frequency effects (LFE) channel is panned 100% to the right, routed as parent channels 5-6 and placed with an azimuth of 0° and an elevation of -90° in the **Multi**Encoder plug-in. Due to this routing, with hard panning to the left or right, each individual 5.1 signal reaches the group track as a mono signal. The elevation parameter of the individual tracks, with the exception of the LFE, is determined differently, according to the taste of the author.

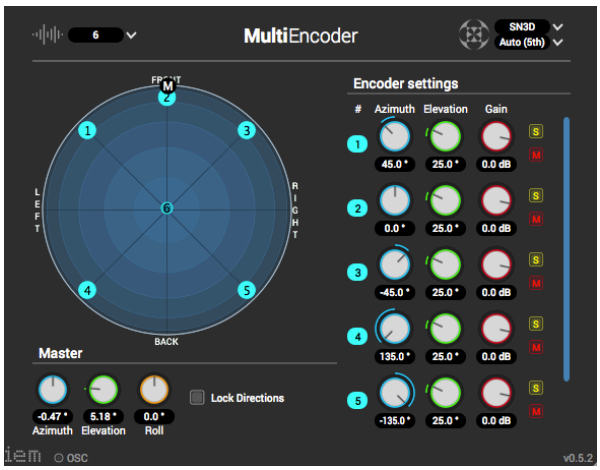


Fig. 10. Screenshot of the author's settings of the MultiEncoder, utilized to encode 5.1 ambience tracks to the 5th order Ambisonic domain.

The **Multi**EQ is used for equalization, especially for low- and high-pass filtering of ambience tracks, foley sounds and sound effects, as well as for musical elements and drone sounds. An automation of different frequency bands is been applied throughout the timeline. Room impressions are created using the **Room**Encoder by placing audio signals, such as lights or a radio, in a virtual room. On some of those sound effects and musical elements, the **Omni**compressor is utilized to achieve a slight compression. The same plug-in is used on the Ambisonic bus as a kind of master bus compression. Reverberation is utilized with the **Fdn**Reverb on ambience tracks, foley sounds and sound effects. Wet/dry and bypass parameters are automated. Another plug-in used is the **Dual**Delay plug-in for some of the musical elements. The energy distribution on the sphere is visualized with help of the **Energy**Visualizer[144], which is inserted on the Ambisonic bus. The **Binaural**Decoder is used for rendering the Ambisonic input signal to a binaural headphone signal.[145] The **AllRA**Decoder and the **Simple**Decoder are used for decoding the Ambisonic input signal to loudspeaker signals.[146] Plug-in descriptions and guides can be found on the IEM Plug-in Suite website[147].

4.3.2 ambiX & mcfx PLUG-IN SUITE

The *ambix_encoder-o5* plug-in is used to encode voice tracks and some of the sound effects into the Ambisonic domain and apply an azimuth and elevation move to them. The main voice is placed in the front, with a partly automation of the azimuth and elevation parameters. In order to avoid noticeable gaps in the Ambisonics representation, the *ambix_widening* plug-in is utilized for source widening on a few multichannel ambience tracks (Ambisonics, 5.1) and musical elements. The *mcfx_meter* is inserted on the Ambisonic bus to monitor the signals sent to the individual channels and the *mxfc_gain_delay* is used to test the speaker setup in the IEM CUBE[148] and at the IEM Produktionsstudio[149]. More information on the ambiX - Ambisonic plug-in suite and the mcfx – multichannel audio plug-in suite can be found on Matthias Kronlachner's website[150] and in his publication on the plug-in suite[151].

[144] IEM Plug-in Suite. "Plug-in Descriptions."
[145] Ibid.
[146] Ibid.
[147] https://plugins.iem.at/
[148] http://iem.at/~zmoelnig/publications/cube_en/
[149] https://iem.kug.ac.at/services/jahresberichte-wurde-2013-eingestellt/200001/produktionsstudio.html
[150] http://www.matthiaskronlachner.com
[151] Kronlachner, "Plug-in Suite for Mastering the Production and Playback in Surround Sound and Ambisonics."

4.4 SOFTWARE INSTRUMENTS

Logic Pro X virtual instruments[152] are used for musical textures. *EXS24 mkII* is a software sampler based on audio recordings to emulate real instruments such as guitars, pianos, and drums.[153] *Alchemy* is a sample synthesizer which features additive, spectral, and granular synthesis as well as resynthesis, sampling and virtual analogue engines.[154] The virtual analogue synthesizer *ES2* combines subtractive synthesis with elements of frequency modulation (FM) and wavetable synthesis.[155] The *Sculpture* synthesizer creates sounds by simulating the physical properties of a vibrating string and enables creation of virtual models of an acoustic instrument, such as a violin or a cello. It is, however, not limited to the reproduction of real instruments.[156] *Retro Synth* is a flexible 16-voice synthesizer that provides analogue, sync, wavetable and FM synthesizer engines.[157] The Piano Roll Editor[158] was used to place MIDI notes on the timeline.

4.5 SOUNDS

Apart from the voice recordings, all field recordings and foley recordings, sounds are used from the Logic Pro X sample library[159], Free to Use Sounds[160], Samples From Mars[161], 99Sounds[162], Sonniss[163], Tonebenders[164], MusicRadar[165] and Audio Shade[166]. These were partly purchased by the author or are available royalty free. Sounds from the Logic Pro X sample library were selected in the DAW and bounced afterwards. For the selected 5.1 samples, 6 individual tracks (L, C, R, Ls, Rs, LFE) were rendered.

[152] https://www.apple.com/logic-pro/plugins-and-sounds/
[153] Apple. “Logic Pro X: EXS24 mkII overview.”
[154] Apple. “Logic Pro X: Alchemy overview.”
[155] Apple. “Logic Pro X: ES2 overview.”
[156] Apple. “Logic Pro X: Sculpture overview.”
[157] Apple. “Logic Pro X: Retro Synth overview.”
[158] https://support.apple.com/kb/PH13422?locale=en_GB
[159] https://www.apple.com/logic-pro/plugins-and-sounds/
[160] https://freetousesounds.com/
[161] https://samplesfrommars.com/
[162] http://99sounds.org/
[163] https://sonniss.com/
[164] http://tonebenderspodcast.com/
[165] https://www.musicradar.com/
[166] https://www.audioshade.com/

5 SELECTED SCENES

The following chapter describes selected scenes from the audio drama in more detail, relating to created soundscape. The 5 chosen scenes are representations of the ambience sound design and for each, several renderings were created. A render with and without the narrator's voice, musical elements and sound effects is presented for each scene to more clearly perceive the spatial impression. In order to be able to listen to field recordings taken with different first-order Ambisonic microphones, there is a render that only contains the spatial ambience recording. The file name is marked with *"_atmos"*. In the audio drama first-order Ambisonic recordings are often combined with 5.1 ambience tracks, these renderings are marked with *"_atmos_plus5point1"*.

For the audio drama the IEM **Multi**Encoder is utilized to encode all multichannel ambience tracks (Ambisonics, 5.1) and 5.1 drone sounds into the Ambisonic domain, for detailed information, refer to chapter 4.3.1. This method provides the advantage of encoding first-order Ambisonic recordings directly into 5th order Ambisonics. For the comparison of different Ambisonic plug-ins and their encoding in different orders the SoundField by RØDE plug-in, the AMBEO A-B Converter and the SPARTA Array2SH plug-in is used to encode A-format recordings, starting out from the first-order Ambisonic microphones, into B-format. These plugins are available free of charge, as is the IEM plug-in suite. A demo version of the Harpex plug-in is used to convert the 1st order recording to 3rd order Ambisonics and create a so-called upmix. This approach was chosen to be able to listen to the spatial field recordings in 1st, 3rd and 5th order and to identify sound characteristics of the respective plug-ins and their encoding. As the narrator's voice, 5.1 ambience tracks, sound effects and musical elements are encoded into the 5th order Ambisonic domain and spatial field recordings are partially encoded in lower resolution, a so-called hybrid solution[167] is created. Additionally, each scene has binaural renderings.

[167] Barrett, "The perception, evaluation and creative application of high order ambisonics in contemporary music practice."

The author is of the opinion that with higher encoding of field recordings the spatial depth increases. Adding 5.1 ambience tracks provides more spatial width. However, as (hearing) impressions are subjective, all renderings are provided on **USB stick** to be listened to and individually compared.

In the following chapters the selected scenes are presented according to above described principles. Information on length and exact time code in the radio play, as well as information about location in the audio drama and first-order Ambisonic microphones used for spatial field recordings is given, the corresponding tables (*table 4* to *8*) show the plug-ins that were used for encoding in different Ambisonic orders. To be able to compare the plug-ins, loudness is compensated. The folders where the respective audio examples are located are specified. Detailed information on the audio samples included on the **USB stick** can be found in appendix B.

Each folder and the included audio files are structured in the same fashion:

/selectedscenes/scene*"scene number"*/*"plug-in name"*/

*"scene number"*_*"plug-in name"*_all.wav

*"scene number"*_*"plug-in name"*_atmos.wav

*"scene number"*_*"plug-in name"*_ atmos_plus5point1.wav

binaural/*"scene number"*_*"plug-in name"*_ binaural_all.wav

binaural/*"scene number"*_*"plug-in name"*_ binaural_atmos.wav

binaural/*"scene number"*_*"plug-in name"*_ binaural_atmos_plus5point1.wav

SCENE 01

This field recording was taken in the woods. One hears chirping birds, insects, barking dogs, wind noises in the trees and cars in the distance. Additionally, a 5.1 ambience track with sounds from a park is used.

Length: 1.10 min

Time code: 1:02 – 2:12

Location in the audio drama: Suburban, close to a tram stop

Microphone: Oktava MK-4012[168]

Additional information: WAV files, 24bit, 44100Hz, 36 channels (Ambix ACN/SN3D) resp. 2 channels; if available, the corresponding microphone preset was used with the plugin; IEM **Multi**EQ was used for low- and high-pass filtering.

Plug-in	Ambisonics order	Folder
IEM MultiEncoder[169] version 0.5.3 -	5th order	**/scene01/multiencoder/**
Harpex[170] version 1.6	3rd order	**/scene01/harpex/**
SPARTA Array2SH[171] version 1.5.0	1st order	**/scene01/sparta/**
SoundField by RØDE[172] version 1.0.4	1st order	**/scene01/soundfield/**
AMBEO A-B Converter[173] version 1.2.3	1st order	**/scene01/ambeo/**

Table 4. Selected scenes, audio examples – Scene 01

Plug-ins used for encoding in different Ambisonic orders. The folders where the respective audio examples are located are specified.

[168] http://www.oktava-shop.com/Small-and-medium-diaphragm-condenser-mics/Oktava-MK-012-01-12.html
[169] https://plugins.iem.at/docs/plugindescriptions/#multiencoder
[170] https://harpex.net/
[171] http://research.spa.aalto.fi/projects/sparta_vsts/plugins.html
[172] https://en.rode.com/soundfieldplugin
[173] https://en-hk.sennheiser.com/ambeo-downloads

SCENE 02

This field recording was taken inside a tramway. Noises of the tram, a murmuring crowd, human noises and cars passing by can be heard. In addition, no 5.1 ambience track is used.

Length: 0.43 min

Time code: 6:45 – 7:28

Location in the audio drama: Inside a tramway

Microphone: SoundField SPS200[174]

Additional information: WAV files, 24bit, 44100Hz, 36 channels (Ambix ACN/SN3D) resp. 2 channels; if available, the corresponding microphone preset was used with the plugin; IEM **Multi**EQ was used for low- and high-pass filtering.

Plug-in	Ambisonics order	Folder
IEM MultiEncoder[175] version 0.5.3	5th order	**/scene02/iem_multiencoder/**
Harpex[176] version 1.6	3rd order	**/scene02/harpex/**
SPARTA Array2SH[177] version 1.5.0	1st order	**/scene02/sparta/**
SoundField by RØDE[178] version 1.0.4	1st order	**/scene02/soundfield/**
AMBEO A-B Converter[179] version 1.2.3	1st order	**/scene02/ambeo/**

Table 5. Selected scenes, audio examples – Scene 02

Plug-ins used for encoding in different Ambisonic orders. The folders where the respective audio examples are located are specified.

[174] https://www.soundfield.com/#/products/sps200
[175] https://plugins.iem.at/docs/plugindescriptions/#multiencoder
[176] https://harpex.net/
[177] http://research.spa.aalto.fi/projects/sparta_vsts/plugins.html
[178] https://en.rode.com/soundfieldplugin
[179] https://en-hk.sennheiser.com/ambeo-downloads

SCENE 03

This field recording was taken in the woods. One hears chirping birds and a rooster and cars in the distance. A second field recording is used taken inside a standing tramway with open doors, where one can hear the room tone of the tram, human sounds and another tramway passing by. It was completed using a 5.1 ambiance track with sounds from a park.

Length: 0.56 min

Time code: 8:29 – 9:25

Location in the audio drama: Suburban, near a tram stop

Microphone: Sennheiser Ambeo VR[180]

Additional information: WAV files, 24bit, 44100Hz, 36 channels (Ambix ACN/SN3D) resp. 2 channels; if available, the corresponding microphone preset was used with the plugin; IEM **Multi**EQ was used for low- and high-pass filtering.

Plug-in	Ambisonics order	Folder
IEM MultiEncoder[181] version 0.5.3	5th order	**/scene03/iem_multiencoder/**
Harpex[182] version 1.6	3rd order	**/scene03/harpex/**
SPARTA Array2SH[183] version 1.5.0	1st order	**/scene03/sparta/**
SoundField by RØDE[184] version 1.0.4	1st order	**/scene03/soundfield/**
AMBEO A-B Converter[185] version 1.2.3	1st order	**/scene03/ambeo/**

Table 6. Selected scenes, audio examples – Scene 03

Plug-ins used for encoding in different Ambisonic orders. The folders where the respective audio examples are located are specified.

[180] https://en-us.sennheiser.com/microphone-3d-audio-ambeo-vr-mic
[181] https://plugins.iem.at/docs/plugindescriptions/#multiencoder
[182] https://harpex.net/
[183] http://research.spa.aalto.fi/projects/sparta_vsts/plugins.html
[184] https://en.rode.com/soundfieldplugin
[185] https://en-hk.sennheiser.com/ambeo-downloads

SCENE 04

This field recording was taken in the woods. One hears chirping birds, crows, insects, barking dogs, soft wind noises in the trees and cars and motorcycles in the distance. Additionally, a 5.1 ambience track with sounds from a park is used.

Length: 1.07 min

Time code: 11:28 – 12:35

Location in the audio drama: In a park

Microphone: SoundField SPS200[186]

Additional information: WAV files, 24bit, 44100Hz, 36 channels (Ambix ACN/SN3D) resp. 2 channels; if available, the corresponding microphone preset was used with the plugin; IEM **Multi**EQ was used for low- and high-pass filtering.

Plug-in	Ambisonics order	Folder
IEM MultiEncoder[187] version 0.5.3	5th order	**/scene04/iem_multiencoder/**
Harpex[188] version 1.6 -	3rd order	**/scene04/harpex/**
SPARTA Array2SH[189] version 1.5.0	1st order	**/scene04/sparta/**
SoundField by RØDE[190] version 1.0.4	1st order	**/scene04/soundfield/**
AMBEO A-B Converter[191] version 1.2.3	1st order	**/scene04/ambeo/**

Table 7. Selected scenes, audio examples – Scene 04

Plug-ins used for encoding in different Ambisonic orders. The folders where the respective audio examples are located are specified.

[186] https://www.soundfield.com/#/products/sps200
[187] https://plugins.iem.at/docs/plugindescriptions/#multiencoder
[188] https://harpex.net/
[189] http://research.spa.aalto.fi/projects/sparta_vsts/plugins.html
[190] https://en.rode.com/soundfieldplugin
[191] https://en-hk.sennheiser.com/ambeo-downloads

SCENE 05

This scene does not contain any field recordings. It is designed using different 5.1. ambience tracks, namely a room tone, room lights and two different rain and thunder samples. The IEM **Room**Encoder was used for placing lights above one's head and a radio on the right of a virtual room.

Length: 1.30 min

Time code: 4:42 – 6:12

Location in the audio drama: In the kitchen

Microphone: -

Additional information: WAV files, 24bit, 44100Hz, 36 channels (Ambix ACN/SN3D) resp. 2 channels;
IEM **Multi**EQ was used for low- and high-pass filtering;
the *ambix_widening* plug-in was utilized for source widening;

Plug-in	Ambisonics order	Folder
IEM MultiEncoder[192] version 0.5.3	5th order	**/scene05/iem_multiencoder/**
Harpex	3rd order	- (no FOA recording)
SPARTA Array2SH	1st order	- (no FOA recording)
SoundField by RØDE	1st order	- (no FOA recording)
AMBEO A-B Converter	1st order	- (no FOA recording)

Table 8. Selected scenes, audio examples – Scene 05

Plug-ins used for encoding in different Ambisonic orders. The folders where the respective audio examples are located are specified.

[192] https://plugins.iem.at/docs/plugindescriptions/#multiencoder

6 IMMERSIVE SOUND DESIGN SURVEY

Differences between stereo technologies and immersive media systems include a higher number of channels available, more technical flexibility and greater artistic freedom. A survey was conducted to obtain the opinion of experts in the field of immersive audio production on their setup and work stages. The replies of the experts from Austria, Germany, Switzerland, Australia and Ireland are summarized in this chapter and may serve as an addition or comparison to the author's approach on the sound design process. The questionnaire used can be found in appendix A.

First, consent of the interviewees for participation in a questionnaire about immersive sound design was obtained. The consent form asked for permission to use and publish comments and quotations to enhance the understanding of the topic as part of this thesis. Furthermore, the interviewees could choose whether they wanted to remain anonymous or if the researcher could explicitly name them and their company in direct quotation.

Participants have been informed that taking part in this study is completely voluntary, that they may stop participating at any time and that they may decide not to answer any specific question. They have been informed that confidentiality of the research data will be subject to standard data use policies at the EU (Data Protection Policy) and that identifiable information will be shared for educational purpose with authorized university staff only.

6.1 AUDIO PRODUCTION SETUP

The workstations used by the respondents include a variety of Mac and Windows systems, as each system operates different software's. The DAWs used vary from Pro Tools[193], REAPER[194], Nuendo[195] and Logic Pro X[196] (see *figure 13*). Media programming environments and stand-alone applications used include Max MSP[197] and SpatialSound Wave[198] as well as game engines like Unity[199], Unreal Engine[200] and audio middleware applications such as FMOD[201] and Wwise[202].

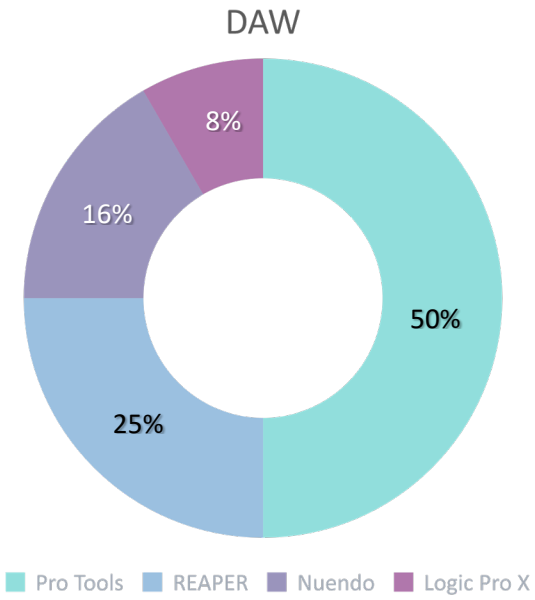


Fig. 11 - Distribution of DAW's used

[193] https://www.avid.com/pro-tools
[194] https://www.reaper.fm/
[195] https://new.steinberg.net/nuendo/
[196] https://www.apple.com/logic-pro/
[197] https://cycling74.com/products/max/
[198] https://www.idmt.fraunhofer.de/en/institute/projects-products/spatialsound-wave.html
[199] https://unity.com/
[200] https://www.unrealengine.com/en-US/
[201] https://www.fmod.com/
[202] https://www.audiokinetic.com/products/wwise/

The plug-ins used for spatialisation are highly diverse. They comprise of the IEM Plug-in Suite[203], Audio Ease 360pan suite[204], Blue Ripple Sound[205], Facebook 360 Spatial Workstation[206], dearVR[207], ambiX - Ambisonic plug-in suite[208], SPARTA[209], Sound Particles[210], binci[211], Gaudio[212], WigWare[213], Anymix Pro[214], and Spanner[215].
Audio interfaces used include manufactures like RME[216], Steinberg[217], Avid[218], MOTU[219], PreSonus[220] and NTP[221].

The experts surveyed work with 5.1, 7.1, 9.1, 7.1.4, 10.2.5 or octophonic loudspeaker setups with studio monitors from manufactures like Genelec[222] and Tannoy[223], as well as headphones from beyerdynamic[224] and Sennheiser[225]. Microphones used by the respondents include Sennheiser Ambeo VR[226], Soundfield ST450[227], Røde NT-SF1[228], Sennheiser MKH 8000[229], Core Sound TetraMic[230], em32 Eigenmike®[231], SCHOEPS ORTF-3D[232] and several binaural microphones including some self-made as well as custom-made Ambisonics and multichannel solutions, which are illustrated in *figure 14*.

[203] https://plugins.iem.at/
[204] https://www.audioease.com/360/
[205] https://www.blueripplesound.com/
[206] https://facebook360.fb.com/spatial-workstation/
[207] https://www.dearvr.com/
[208] http://www.matthiaskronlachner.com/?p=2015
[209] http://research.spa.aalto.fi/projects/sparta_vsts/
[210] https://www.soundparticles.com/
[211] https://binci.eu/
[212] https://www.gaudiolab.com/
[213] https://www.brucewiggins.co.uk/?page_id=78
[214] http://www.iosono-sound.com/vstaax-plug-ins/
[215] http://thecargocult.nz/spanner.shtml
[216] http://www.rme-audio.de/en/
[217] https://www.steinberg.net/en/products/audio_interfaces/ur_series/start.html
[218] https://www.avid.com/en/products/pro-tools-hd-omni
[219] http://www.motu.com/
[220] https://www.presonus.com/products/Audio-Interfaces
[221] http://www.ntp.dk/PRODUCTS/Interfaces---Routers
[222] https://www.genelec.com/
[223] https://www.tannoy.com/
[224] https://global.beyerdynamic.com/
[225] https://en-us.sennheiser.com/
[226] https://en-us.sennheiser.com/microphone-3d-audio-ambeo-vr-mic
[227] https://www.soundfield.com/#/products/st450mk2
[228] https://en.rode.com/nt-sf1
[229] https://en-us.sennheiser.com/mzd-8000
[230] http://www.core-sound.com/TetraMic/1.php
[231] https://mhacoustics.com/products
[232] https://schoeps.de/en/products/surround-3d/ortf-3d.html

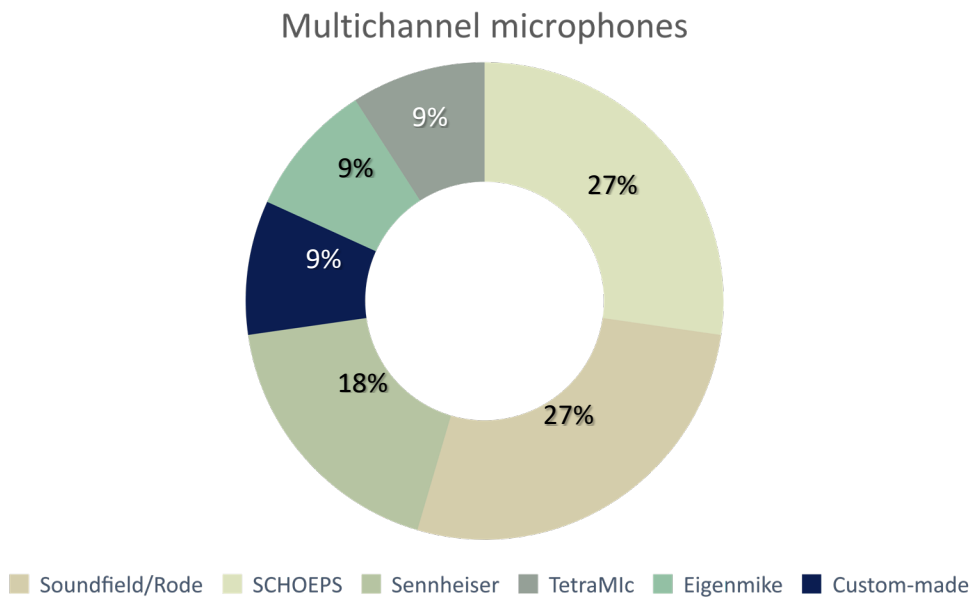


Fig. 12 - Distribution of multichannel microphones used

When being asked about what field recording devices they use, brands such as Zoom[233], Sound Devices[234] and Zaxcom[235] are named (see *figure 15*). The majority of the experts surveyed like to work with self-recorded audio files, but they also draw on sound libraries and make use of mono, stereo, binaural, Ambisonics, 4.0, 8.0 and 9.1 audio samples.

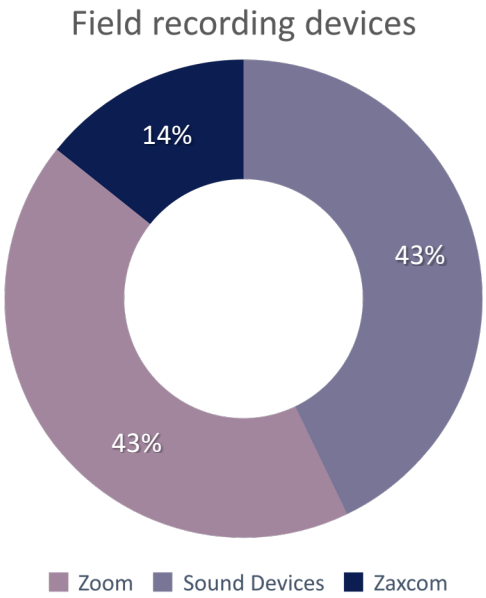


Fig. 13 - Distribution of field recording devices used

[233] https://www.zoom.co.jp/products/field-video-recording/field-recording
[234] https://www.sounddevices.com/
[235] https://zaxcom.com/products/nomad/

6.2 KNOWLEDGE IN STEREO PRODUCTION

All interviewed audio experts agree that it was possible, if not necessary, to transfer their knowledge of stereo production into the form of spatial audio production. Especially in relation to the fundamentals of working with DAWs, plug-ins, audio samples, etc. Enda Bates and Martin Rieger explain, that spatial audio is more complex than stereo. Understanding of auditory perception and the sound's impact on vision in 360° videos is crucial to be able to create the illusion of spatial sound. In immersive audio production, audible mistakes may occur easily as there are more factors that need to be considered for the immersive setting as opposed to stereo.

6.3 EXPERIENCE IN SPATIAL AUDIO

The respondents have been working in the audio industry from 10 to 37 years, have work experience in multichannel audio of 9 to 19 years and have started working with 3D audio between 4 to 9 years ago. The participants point out that the term spatial audio can be used as an umbrella term for every immersive audio format including binaural, 5.1, Ambisonics, VR, etc. However, some refer to 360° video using terms such as 360° audio, spatial audio or VR sound explicitly and use the term 3D audio for loudspeaker setups only.

6.4 IMMERSIVE MEDIA AND AUDIO FORMATS

Formats that the experts produce for include 360° videos, (interactive) VR, AR, games, cinematic VR productions (360° films), Ambisonics, Dolby Digital[236], Dolby Atmos[237], Auro-3D[238], SpatialSound Wave[239], as well as music work for octophonic arrays (see *figure 16*).

[236] https://www.dolby.com/us/en/technologies/dolby-digital.html
[237] https://www.dolby.com/us/en/brands/dolby-atmos.html
[238] https://www.auro-3d.com/
[239] https://www.idmt.fraunhofer.de/en/institute/projects-products/spatialsound-wave.html

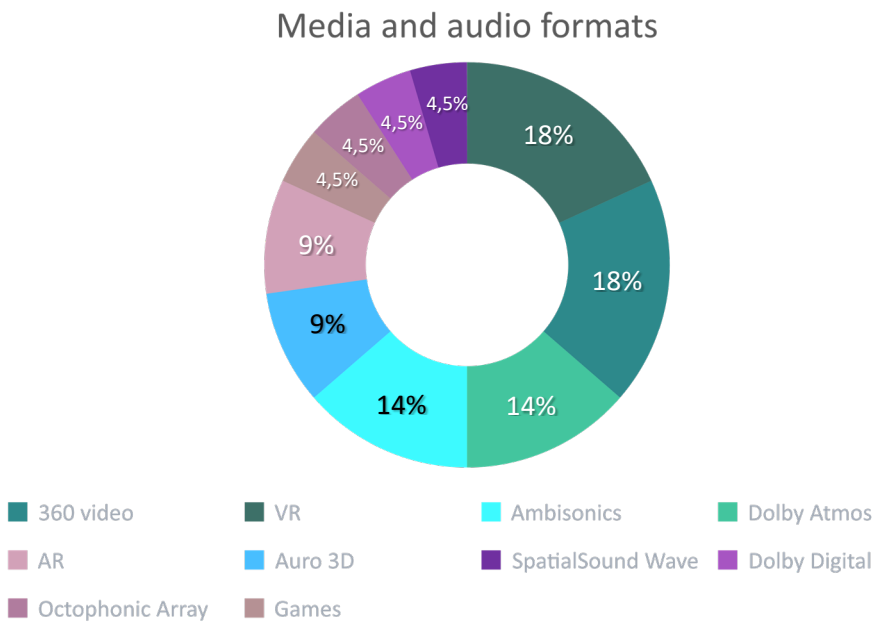


Fig. 14 - Distribution of Immersive media and audio formats in which productions are carried out

6.5 SPATIAL AUDIO PRODUCTION WORKFLOW

Experts point out that it is important to define the delivery platform first, and then adjust the spatial audio needs to suit the final delivery format. Methods to get started could be either to collect sounds needed and work with discrete sources first, or, alternatively, to start with setting up spatial field and multichannel recordings. Another way to start the project is to select basic ambience sounds from sample libraries, to define the given space, and then continue by adding dynamic layers. Also, it may be possible to get an AAF[240] with synchronized lavalier microphones and then start the process from scratch by syncing multichannel recordings, going on to the editing and restauration and lastly adding additional recordings of objects and ambiences, eventually combining them with sounds from libraries if needed.

The storyline determines the placement of the sounds. Working on 360° videos and VR content, a starting point is to place sounds on visible sources. In addition, head tracking of auditory objects can be implemented. Enda Bates from Trinity College Dublin, who wrote his PhD thesis on "The Composition & Performance of Spatial Music"[241], describes to commonly

[240] http://avid.force.com/pkb/articles/en_US/compatibility/en336549
[241] http://endabates.net/PhD%20Spatial%20Music.html

think in gestures, referring to a specific position and movement. When it comes to strategy for spatial positioning, Stefan Zaradic from IAN.solutions[242] is guided by his own excitement.

Almost none of the respondents uses sketches, neither on electronic devices nor on paper, to define the spatial arrangement. If anything, a sketch could be drawn to mark hidden microphones which have been used on a film set. Enda Bates points out that visual layouts do not really match with how humans actually perceive spatial sound. A circle drawn on paper can be seen all at once, opposed to a circular sonic trajectory, which takes time to occur and has to be consciously followed by the listener.

The spatial placement of dialogue, monologue or narration depends on the context of the scene. Usually, leading voice is placed in the center or, in very few cases, between center and left channel or center and right channel. Narration or voice-over without a visual component should be head-locked, even mono and non-spatialized produce may be applicable. Dialogue in 360° videos and VR content should be placed to match the position of the speaker. The same method applies for a narrator appearing on screen. Depending on the character's action, the voice could also be head-tracked.

The experts surveyed point out that well-made (multichannel) field recordings are the foundation to transport listeners to a new acoustic space. Thereon, a spatial bed can be built. An illusory world does not necessarily have to be realistic or naturalistic. The key is to create an environment that is believable. Time-based processing, including Ambisonic impulse responses, editing and equalisation can be used to design those believable ambience sounds. The decorrelation of an audio signal by granulation and its spatial distribution can be useful for soundscape composition.

Timbre and loudness depend on the project's purpose as there are no standards defined yet. A broad recommendation is not to make it too loud and not too quiet. Loudness has an impact on perceived distance and should be deployed carefully, especially when producing audio for 360° video and VR. Reducing high frequencies can be associated with an increase in distance. Knowing about the technical implementation of the production is essential for all

[242] https://www.ian.solutions/

respondents, as it indicates the delivery specification and therefor the limitations of the delivery device. The selection of the tools used depend on those factors and should result in a well-balanced listening experience.

Opinions of the surveyed experts differ widely in regard to the use of equalization, dynamic processing and reverberation. Some prefer to use several different approaches when working in spatial audio as opposed to their approach in a traditional stereo mix approach. Equalisation and dynamics are being used similar to film, although dynamics are limited when producing for headphone playback. Some experts prefer to utilize equalisation and dynamics for mono, stereo or multichannel tracks only, but point out that it might work for Ambisonics as well. Another possibility to apply compressors or limiters is to use it sparingly during the audio mixing process alternatives are may be the use of clip gain or only faders. In 360° videos, reverberation should match the real space. At times it might be difficult to recreate acoustically accurate spaces and a combination of several convolution reverbs may be useful to overcome these difficulties. However, using reverberation may not always be necessary, especially in spatial music pieces for loudspeakers.

The respondents agree that theoretical research may lead to results that can be acoustically perceived by the listener. Damian Candusso[243] holds a PhD in spatial sound and mentions that his practise is based on his research. Enda Bates points out that most of his work is based on determining what the listener actually perceives and it is therefore no surprise that spatial perception by the listener is taken into account during the production process.

There are several ways to interpret the term immersion. The right balance between sound sources and their placement in a virtual space can help to create a credible environment. Yet if the illusory sound world is not interesting, the listener will most likely be bored and not be immersed, regardless of the production technique. Stefan Zaradic points out that emotions triggered in the listener can intensify the immersion. While Enda Bates states to use either head tracking or personalized head-related transfer functions (HRTF) to achieve the technical level of immersion. Adding early reflections and reverb can also help with externalization in

[243] http://www.damiancandusso.com/

this regard. On the other hand, Martin Rieger from VRTonung[244] explains, that sounds which have no visual representation, including background music, may interfere with immersion.

6.6 FUTURE OF 3D AUDIO

The experts surveyed point out that one day, 3D audio will most likely become a standard format, just as stereo productions are nowadays. Unlike any other medium, spatial audio is a functional component of VR, AR, XR and 360° video and enhances the experiential value. In other words, it is a necessity rather than an additional feature for those formats. Visual components can be missed by the end user, but an audio element cannot as easily be ignored. Respondents also point out that 3D audio in movies is an acceptable feature but does not influence the perceived experience as much as 3D audio in other formats. However, they do not quite agree whether the future of 3D sound is actually now or if it is still far away.

[244] https://www.vrtonung.de/

7 DISCUSSION

Although only a small number of experts in the field of 3D audio was surveyed, the range of immersive media and audio formats in which the experts produce is still diverse. Some of the respondents have a long experience in the field of audio in general and in 3D audio in particular. The survey among audio experts has shown that there is a variety of understandings concerning the 3D audio topic. These are not always used equally among the interviewees and the same topics are often explained with different grasp. One could conclude that although there is a pool of terms used in 3D audio, there seems to be no unified standard yet.

As mentioned in the introduction, the concept of immersion can be interpreted differently as well. The feeling of immersion can be achieved by appealing to different senses, and furthermore the technical level of immersion needs to be taken into account in regard to create immersive experiences. Regard to present audio production, obtaining the opinion of listeners would be necessary to conclude whether the audio drama would be regarded to be immersive.

Regarding the production setup, macOS and Microsoft Windows systems are used almost equally or in parallel by the respondents. The macOS provides specific software, such as Logic Pro X that were used by the author, in accordance with surveyed experts, Pro Tools is the most frequently used DAW among them. This is probably due to the fact that this DAW is still referred to as the industry standard, although REAPER supports up to 64 channels per track in order to create higher order Ambisonic projects (up to 7^{th} order)[245] and is available for a reasonable price. However, if large projects with many channels per track are created and many Ambisonics plug-ins are used at the same time, it is computation-intensive and requires a high-end workstation.

If interactive content is produced for VR, it is inevitable that a sound designer will also acquire knowledge of game engines or at least middleware applications. However, due to the limited time on this project, the author did not include 360° video or VR.

[245] IEM Plug-in Suite. “Can I use the plug-ins with every DAW?”

There has already been a collaboration on a previous project with the language artist John Sauter, who contributed to this project with his text and the narrator's voice. The author provided the sound concept and entire audio production. Throughout the process there was a constant exchange between the two involved and feedback was given on a regular basis.

The plug-ins and microphones used show a great variety of experimentation on the part of the audio experts. The author used the technical means at his disposal and mostly used plugins of the IEM-Plugin-Suite, as it contains all plug-ins to perform Ambisonic productions and the personal support is provided by the Institute of Electronic Music and Acoustics at the University of Music and Performing Arts Graz. Even if it is not important for the function of plug-ins, the author is of the opinion that visually more appealing interfaces increase the pleasure of working with them. According to the author, the SoundField by RØDE plug-in, for example, is both visually more appealing and more clearly arranged in terms of functions than the AMBEO A-B Converter. In order to be able to give targeted suggestions for improvements for single plug-ins, the author would have to deal with them in more detail.

For field recordings, it was important to the author to gather usable recordings with as little equipment as possible. This was achieved with first-order Ambisonics (FOA) microphones and the 4-channel field recorder including 48V phantom power. Larger multi-channel rigs can deliver higher quality recordings, but the effort required to use them increases as well, as logistics become more extensive. The challenge with the FOA microphones was to record little or no noise, which could not always be prevented, and the noise was then removed during the editing process. To avoid people realising that they are being recorded and consequently changing their behaviour, the field recorder and the microphone were hidden in a laptop bag. Only the head of the microphone protruded from the bag and a black sock was rinsed over it. Of course, this is not an ideal microphone placement, but the intended purpose to capture realistic recordings was fulfilled.

In order to produce 3D audio for loudspeakers, the appropriate speaker setup needs to be available, which can be of greater effort in contrast to stereo or 5.1 setups. The author is pleased that he was able to access the fully equipped IEM studios during the production phase.

The question remains as to why the interviewed experts prefer to use their own recordings or combine them with sounds from libraries as they did not specify to use only libraries. The reasons could be that there are not enough or no available sounds of adequate quality or that the audio experts like to preserve their independence with self-made recordings. The author made field recordings for the same reason, which also allowed him to test different FOA microphones in the process.

The expert's opinion, on the importance of knowing about what format to produce for in advance is consistent with the author's opinion. The same is true for the statement that the story should always be in the foreground. Emotions triggered at the right time can put the listener directly into the (sound) scene. The author likes to start a project with ambience sounds to create the desired space and the environment, as do some of the interviewees as well. The author's approach can be described as a path from big to small, from ambient sound to musical elements to sound effects. First the big picture was created, and later individual details were discussed.

As the survey has shown, no audio experts use sketches to define the sound scene in advance. The author marked possible sound cues in the text and as it turned out later in the process, far too many words were marked and consequently the corresponding sound idea was not used in the audio drama. It was only by listening to the speech track that it became clear where the space for sounds was available.

Since there is still no standard for the loudness of Ambisonic productions yet, the author suggests to depend on one's auditory perception. Macedo[246] likes to divide the artistic work into two modes: the working mode, which covers recording, processing and editing, and the listening mode, which changes the focus of the mind to the aural results. In different phases of the composition, this dual approach means listening as often as possible and leaving the ear as the final judge in the process of deciding what the final work will sound like.
The author shares the opinion of some respondents that equalization, dynamic processing and reverberation should be used sparingly.

[246] Macedo, "Phenomenology, Spatial Music and the Composer."

As the audio drama is told by a narrator, the author decided to position the voice classically in the center. Among the audio experts surveyed, everyone would proceed similarly, as the narrator does not move in space. The author is of the opinion that the narrator's voice does not sound as benevolent after encoding in Ambisonics as it does in the Logic Pro X stereo bounce, however, no improvement to that issue could be achieved. Despite the microphones catching disturbance noises coming from the microphone stand or rustling of clothes during voice recordings, these were tolerated in order to not hinder the narrator in his speech flow and to make the recordings livelier.

Experience in stereo production is definitely an advantage for 3D audio production. The author concludes that small mistakes might reoccur eventually as there is more to consider in a 3D audio setting as opposed to e.g. stereo. Especially given the fact that Ambisonics or 5.1. ambience sounds increase the number of tracks considerably and such large projects require a lot of overview and concentration.
The author is of the opinion that 3D audio will gain even more influence and find its way into the listeners living room, either through sound bars or headphone applications.

8 CONCLUSION

Intense immersion is not only achieved by production techniques. Nonetheless, 3D audio offers the advantage of placing ambience sounds, music elements and sound effects in space and creating spatial impressions. Furthermore, it offers the advantage of technical immersion through a higher number of speakers or e.g. head tracking and personalized HRTF's for headphones. The story always contributes to immerse oneself in an experience and also mono/stereo productions can create immersion.
In order not to endanger the immersion of sound space creation, one should consider to create a believable environment. Therefore, sound quality for e.g. field recordings, voice recordings, samples and software instruments are as important as trying to avoid errors in production. These can be manifold and range from unwanted noises in field, voice or music recordings to routing errors in DAW's, especially when working with higher order Ambisonic projects. Although there might be more room for additional sounds in a 3D audio environment than in a limited stereo image, even in 3D audio, less is sometimes more.

As a sound designer one faces both artistic and technical challenges. Especially with 3D audio one has to master the technology to be artistically free. Technical instructions can be passed on, but the artistic freedom lies with the creator. Taste and aesthetics are always in the eye (in this case the ear) of the observer and preferences can be of great variety. The accompanying spatial audio piece turned out to be more of a technical realization than an artistic achievement.

As the work phase for this master thesis was limited in time, some tasks remain to be done in future. A questionnaire regarding the audio drama and selected scenes could be designed for a listening test, to be carried out in different listening situations with the aim of obtaining the opinion of listeners to conclude whether the audio drama would be regarded to be immersive.

To gain more artistic freedom, the author could deal with the used plug-ins in more detail. One could try to cut down on the use of the REAPER project to save channels, whereas for higher quality field recordings larger multi-channel rigs could be tested.

Technical ear training as well as more detailed theoretical knowledge in the field of psychoacoustics, auditory perception and spatial hearing and the perception of spatial sound can be helpful in producing immersive spatial audio in future projects.

REFERENCES

LITERATURE

Aro, Eero. “Surround Sound in Radio Drama.” Paper presented at the AES 22nd UK Conference, Cambridge, UK, April 11-12, 2007.

Averese, John. Post Sound Design: The Art and Craft of Audio Post Production for the Moving Image. New York; London: Bloomsbury Academic, 2017.

Baalman, Marije A.J. “Spatial Composition Techniques and Sound Spatialisation Technologies.” Organised Sound, Volume 15, Issue 3 (2010): 209–218. Accessed March 5, 2019. https://doi.org/10.1017/S1355771810000245.

Barrett, Natasha. “The perception, evaluation and creative application of high order ambisonics in contemporary music practice.” IRCAM Composer in Research Report 2012. Accessed April 3, 2019. https://www.natashabarrett.org/Barrett_IRCAM_Report_v2.pdf.

Barrett, Natasha. “Ambisonics and acousmatic space: a composer’s framework for investigating spatial ontology.” In Proceedings of the Sixth Electroacoustic Music Studies Network Conference, Shanghai, China, June 21-24, 2010.

Barrett, Natasha. “Spatio-musical composition strategies.” Organised Sound, Volume 7, Issue 3 (2002): 313-323. Accessed March 20, 2019. https://doi.org/10.1017/S1355771802003114.

Bates, Enda et al. “Comparing Ambisonic Microphones - Part 2.” Paper presented at the 142nd AES Convention, Berlin, Germany, May 20-23, 2017.

Bates, Enda et al. “Comparing Ambisonic Microphones: Part 1.” Paper presented at the AES Conference on Sound Field Control, Guildford, UK, July 18-20, 2016.

Blauert, Jens. Spatial Hearing: The Psychophysics of Human Sound Localization. Cambridge, Mass: MIT Press, 1997.

Blauert Jens and Lindemann Werner. "Auditory spaciousness: Some further psychoacoustic analyses." The Journal of the Acoustical Society of America 80, Issue 2 (1986): 533–542. Accessed September 3, 2019. https://doi.org/10.1121/1.394048.

Bregman, Albert. "Auditory Scene Analysis: The Perceptual Organization of Sound." The Journal of the Acoustical Society of America 95, Issue 2 (1994). Accessed March 19, 2019. https://doi.org/10.1121/1.408434.

Corey, Jason. "Technical ear training: Tools and practical methods." The Journal of the Acoustical Society of America 133, Issue 5 (2013). Accessed March 19, 2019. https://doi.org/10.1121/1.4805967.

Droumeva, Milena. "Understanding immersive audio: A historical and socio-cultural exploration of auditory displays." In Proceedings of ICAD 05-Eleventh Meeting of the International Conference on Auditory Display, Limerick, Ireland, July 6-9, 2005.

Eaton, Callum and Lee, Hyunkook. "Quantifying Factors of Auditory Immersion in Virtual Reality." Paper presented at the AES International Conference on Immersive and Interactive Audio, York, UK, March 27-29, 2019.

Elfitri, Ikhwana and Luthfi, Amirul. "Reviews on Technology and Standard of Spatial Audio Coding." Organised Sound, Volume 6, Number 1 (2017): 209–218. Accessed July 16, 2019. https://doi.org/10.25077/jnte.v6n1.372.2017.

Frank, Matthias, Zotter, Franz and Sontacchi, Alois. "Producing 3D Audio in Ambisonics." Paper presented at the 57th AES International Conference, Hollywood, USA, March 6-8, 2015.

Frank, Matthias. "How to make Ambisonics sound good." Paper presented at the 7th Forum Acusticum, Krakow, Poland, September 7-12, 2014.

Kendall, Gary S. “Spatial perception and cognition in multichannel audio for electroacoustic music.” Organised Sound, Volume 15, Issue 3 (2010): 228–238. Accessed September 3, 2019. https://doi.org/10.1017/S1355771810000336.

Görne, Thomas. Sounddesign: Klang, Wahrnehmung, Emotion. München: Hanser, 2017.

Hahn, Ephraim. “Musical Emotions Evoked by 3D Audio.” Paper presented at the AES International Conference on Spatial Reproduction, Tokyo, Japan, August 6 – 9, 2018.

Herre, Jürgen et al. “MPEG-H Audio - The New Standard for Universal Spatial / 3D Audio Coding.” Journal of the Audio Engineering Society, Volume 62, Issue 12 (2014):821-830. Accessed July 16, 2019. https://doi.org/10.17743/jaes.2014.0049.

Kronlachner, Matthias. “Plug-in Suite for Mastering the Production and Playback in Surround Sound and Ambisonics.” in Gold-Awarded Contribution to AES Student Design Competition at the 136th AES Convention, Berlin, Germany, April 26-29, 2014.

Kurz, Eric, Pfahler, Felix and Frank, Matthias. “Comparison of first-order Ambisonic microphone arrays.” In Proceedings of ICSA 2015 (International Conference on Spatial Audio 2015), Graz, Austria, September 18-20, 2015.

Lawrence, Rob. “Evaluating Immersive Experiences in Audio for Composition.” In Proceedings of ICSA 2014, Erlangen, Germany, February 21-23, 2014, 82-90.

Lennox, Peter, Myatt, Toni and Vaughan, John M. “From Surround to True 3-D.” In Proceedings of the AES 16th International Conference: Spatial Sound Reproduction, Rovaniemi, Finland, April 10-12, 1999.

Macedo, Frederico. “Phenomenology, Spatial Music and the Composer: prelude to a phenomenology of space in electroacoustic music.” In Proceedings of the International Computer Music Conference 2011, Huddersfield, UK, July 31-August 5, 2011.

McGregor, Iain and Cunningham, Stuart. “Comparative Evaluation of Radio and Audio Logo Sound Designs.” Journal of the Audio Engineering Society Vol. 63, No. 11 (2015): 876-888. Accessed March 24, 2019. http://dx.doi.org/10.17743/jaes.2015.0076.

Møller, Henrik. “Fundamentals of Binaural Technology.” Applied Acoustics, Volume 36, Issues 3-4 (1992):171-218. Accessed September 3, 2019. https://doi.org/10.1016/0003-682X(92)90046-U.

Moore, Brian C. J. An Introduction to the Psychology of Hearing. Sixth edition. Leiden: Brill, 2013.

Morimoto, Masayuki. “Auditory spaciousness and envelopment.” In Proceedings of the 13th International Congress on Acoustics (ICA), Belgrade, Yugoslavia, 1989.

Nachbar, Christian et al. “ambix - A Suggested Ambisonics Format.” Paper presented at the Ambisonics Symposium 2011, Lexington, USA, July 2-3, 2011.

Noisternig, Markus et al. “A 3D Ambisonic based Binaural Sound Reproduction System.” Proceedings of the 24th AES International Conference on Multichannel Audio, Banff, Canada, June 26-28, 2003.

Parncutt, Richard. “Psychoacoustics and cognition for musicians.” In Sound musicianship: Understanding the crafts of music, edited by Andrew R. Brown, 76-87. Cambridge: Cambridge Scholars Publishing, 2012.

Pike, Christopher W. “Evaluating the Perceived Quality of Binaural Technology.” PhD diss., University of York, 2019.

Rodero, Emma. “See It on a Radio Story: Sound Effects and Shots to Evoked Imagery and Attention on Audio Fiction.” Communication Research, 39 (2012): 458-479. Accessed April 4, 2019. https://doi.org/10.1177/0093650210386947.

Rodero, Emma. “Stimulating the Imagination in a Radio Story: The Role of Presentation Structure and the Degree of Involvement of the Listener.” Journal of Radio & Audio Media, Volume 19, Issue 1 (2011): 45-60. Accessed April 8, 2019. https://doi.org/10.1080/19376529.2012.667022.

Roginska, Agnieszka and Geluso, Paul, eds. Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio. New York; London: Routledge, Taylor & Francis Group, 2018.

Romanov, Michael. “Comparison of Amplitude Panning Approaches on ITU BS.2051 Loudspeaker Layouts with Height.” Master's thesis, University of Music and Performing Arts Graz and Graz University of Technology, 2018.

Sonnenschein, David. Sound design: The expressive power of music, voice, and sound effects in cinema. Studio City, California: Michael Wiese Productions, 2001.

Sontacchi, Alois. “Dreidimensionale Schallfeldreproduktion für Lautsprecher- und Kopfhöreranwendungen.” PhD diss., Graz University of Technology, 2003.

Spors, Sascha et al. “Spatial Sound With Loudspeakers and Its Perception: A Review of the Current State.” Proceedings of the IEEE, Volume 101, Issue 9 (2013): 1920 – 1938. Accessed April 2, 2019. https://doi.org/10.1109/JPROC.2013.2264784.

Theile, Günther, Wittek, Helmut and Reisinger, Markus. “Wellenfeldsynthese-Verfahren: Ein Weg für neue Möglichkeiten der räumlichen Tongestaltung.” In Proceedings of the 22th Tonmeistertagung – VDT International Audio Convention, Hannover, Germany, November 22-25, 2002.

Zhang, Chenyan, Perkis, Andrew and Arndt, Sebastian. “Spatial Immersion versus Emotional Immersion, Which is More Immersive?” Paper presented at the 9th International Conference on Quality of Multimedia Experience (QoMEX 2017), Erfurt, Germany, May 31–June 2, 2017.

Zotter, Franz and Frank, Matthias. Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality. Cham: Springer, 2019. Accessed May 5, 2019. https://link.springer.com/book/10.1007/978-3-030-17207-7.

Zotter, Franz and Frank, Matthias. "All-Round Ambisonic Panning and Decoding." Journal of the Audio Engineering Society, Volume 60, Issue 10 (2012): 807-820. Accessed July 16, 2019. http://www.aes.org/e-lib/browse.cfm?elib=16554.

Zotter, Franz, Frank, Matthias and Sontacchi, Alois. "The Virtual T-Design Ambisonics-Rig Using VBAP." Paper presented at the 1st European Congress on Sound and Vibration, Ljubljana, Slovenia, September 15-18, 2010.

ONLINE SOURCES

WEBSITES/DOCUMENTS:

Apple. "Logic Pro X: Alchemy overview." Accessed June 23, 2019. https://support.apple.com/kb/PH27371?locale=en_GB

Apple. "Logic Pro X: ES2 overview." Accessed June 23, 2019. https://support.apple.com/kb/PH27429?locale=en_GB.

Apple. "Logic Pro X: EXS24 mkII overview." Accessed June 23, 2019. https://support.apple.com/kb/PH27714?locale=en_GB.

Apple. "Logic Pro X: Retro Synth overview." Accessed June 23, 2019. https://support.apple.com/kb/PH27625?locale=en_GB.

Apple. "Logic Pro X: Sculpture overview." Accessed June 23, 2019. https://support.apple.com/kb/PH27202?locale=en_GB.

Beck, Alan. “Point-of-listening in radio plays.” Sound Journal. Accessed April 8, 2019. https://www.kent.ac.uk/arts/sound-journal/beck981.html.

Bolles, Monica. “Defining Immersive (in the age of Everything Immersive).” Accessed August 6, 2019. https://www.thedometour.com/single-post/2019/06/20/Defining-Immersive-in-the-age-of-Everything-Immersive.

Cambridge Dictionary, s.v. “immersive.”, accessed August 6, 2019, https://dictionary.cambridge.org/dictionary/english/immersive.

Dyar, Hugh H. “Stuck In Spheres: Immersive Storytelling and 360 Video.” Accessed August 6, 2019. http://stuckinspheres.com/sis/concept/onimmersionandstorytelling/.

Ferster, Bill. “The Magical Number Seven, Plus Or Minus Two.” eLearning Design and Development. Accessed April 13, 2019. https://elearningindustry.com/magical-number-seven-plus-minus-two-memory-affects-the-perception-of-information.

Forrester, Michael A. “Auditory perception and sound as event: theorising sound imagery in psychology.” Sound Journal. Accessed April 8, 2019. https://www.kent.ac.uk/arts/sound-journal/forrester001.html.

IEM Plug-in Suite. “Can I use the plug-ins with every DAW?” Accessed May 9, 2019. https://plugins.iem.at/faq/#can-i-use-the-plug-ins-with-every-daw.

IEM Plug-in Suite. “Plug-in Descriptions.” Accessed June 24, 2019. https://plugins.iem.at/docs/plugindescriptions/#dualdelay.

International Telecommunication Union (ITU). “Multichannel sound technology in home and broadcasting applications.” Accessed May 14, 2019. https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-BS.2159-4-2012-PDF-E.pdf

Lessiter, Jane et al. “Evaluating Immersive User Experience and Audience Impact.” Digital Catapult, June 28, 2018. Accessed March 30, 2019. https://www.digicatapult.org.uk/news-and-views/publication/audience-immersive-report/.

Oxford Dictionary, s.v. “immersive.”, accessed August 6, 2019, https://en.oxforddictionaries.com/definition/immersive.

Wikipedia, s.v. “Immersion (virtual reality).”, accessed August 6, 2019, https://en.m.wikipedia.org/wiki/Immersion_(virtual_reality).

YOUTUBE VIDEOS:

Marshall McGee, “Layering - the most important tool for a sound designer.” YouTube video, 4:31, August 29, 2016, https://www.youtube.com/watch?v=0biAgn2ct0A.

SCHOEPS Mikrofone, “Florian Camerer (ORF): "Die Kirche, das Dorf und 3D-Audio - was gehört wohin?".” YouTube video, 1:00:46, April 18, 2018, https://www.youtube.com/watch?v=osZ842Zaj5Q.

LIST OF FIGURES

LIST OF TABLES

APPENDIX A - SURVEY

Immersive Sound Design Questionnaire

What does your audio production setup look like? In terms of…

- Workstation specifications (processor, RAM, hard drive(s), operating system)?
- DAW, media programming environments, stand-alone applications?
- Plug-ins (especially related to spatialisation)?
- Audio interface?
- Speaker setup/headphones?
- Field recorder & microphones?
- Audio samples: self-recorded or from libraries?
- Audio samples: mono/stereo files or surround sound files/3D audio files (5.1., ambisonics,…)?

How long have you been working in the audio production business? When did you start with spatial audio?
What terms would you use to describe your activities in this area? (e.g. 3D audio, immersive audio, 360 sound, ...)

What types of immersive media and audio format(s) are you producing for?

What does your spatial audio production workflow look like?

- In terms of methodology: What is your starting point? And what steps do you take from there? (set a frame and “fill out” or “step by step”)
- What is your strategy for spatial placement?
- What placement of sound would you avoid?
- Do you use sketches, either on electronic devices or on paper, to define the spatial arrangement?
- Where do you place dialogue/monologue/narration (if any)?
- Design possibilities/tools for ambience sounds?

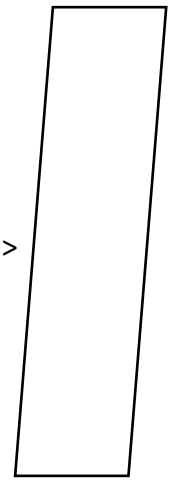
- What should be considered in terms of timbre and loudness?
- Do you think about the technical implementation during production?
- How do you use equalization, dynamic processing, reverberation, etc.?
- Do you use theoretical research in a way that leads to results which can be perceived aurally by the listener?
- Do you consider spatial perception during production?
- How many sound cues do you use in one sound scene? And how many of them are moving?
- What is your approach to immerse the listener in the story?
- What should be considered to not endanger the immersion of a sound space creation?

Was it possible for you to transfer your knowledge of stereo production into this form of audio production or did you have to start from scratch?

Where do you think the 3D audio journey will go in the future?

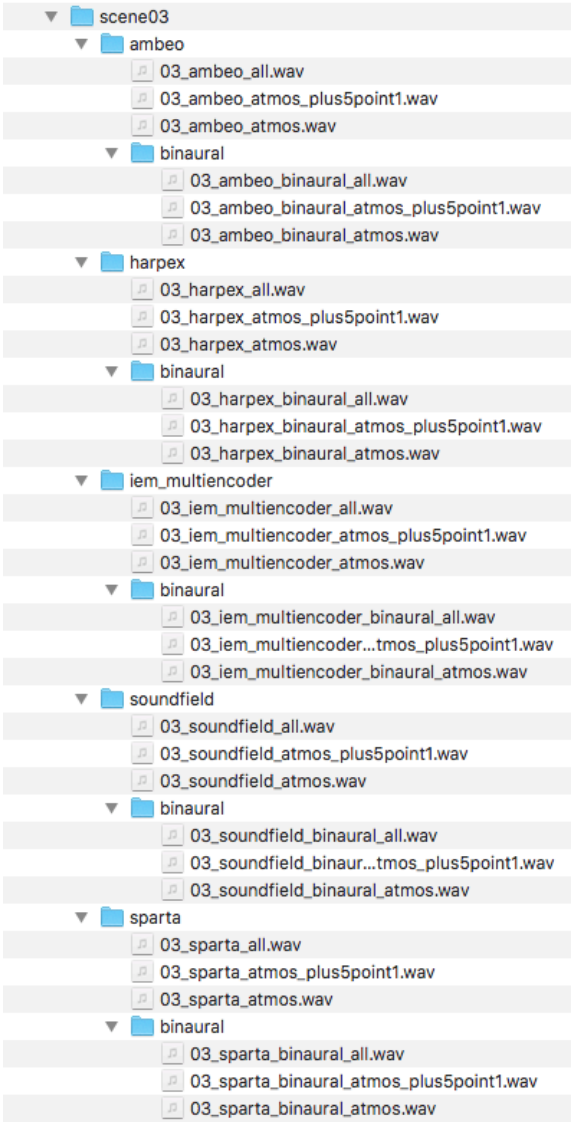
APPENDIX B - AUDIO EXAMPLES

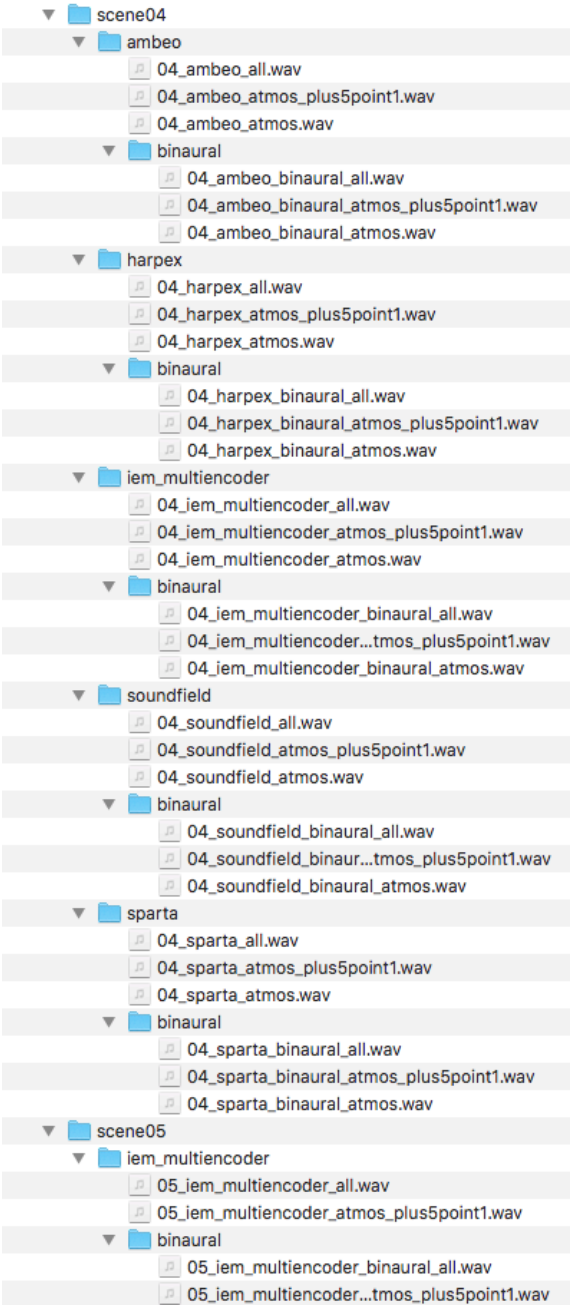
Audio samples can be found on this **USB stick** - - - - - - - - - - - - - - - - >



FILES INCLUDED:

- _Spinnenbank
  - spinnenbank_36channel.wav
  - spinnenbank_binaural.wav
- selectedscenes
  - scene01
    - ambeo
      - 01_ambeo_all.wav
      - 01_ambeo_atmos_plus5point1.wav
      - 01_ambeo_atmos.wav
      - binaural
        - 01_ambeo_binaural_all.wav
        - 01_ambeo_binaural_atmos_plus5point1.wav
        - 01_ambeo_binaural_atmos.wav
    - harpex
      - 01_harpex_all.wav
      - 01_harpex_atmos_plus5point1.wav
      - 01_harpex_atmos.wav
      - binaural
        - 01_harpex_binaural_all.wav
        - 01_harpex_binaural_atmos_plus5point1.wav
        - 01_harpex_binaural_atmos.wav
    - iem_multiencoder
      - 01_iem_multiencoderall.wav
      - 01_iem_multiencoderatmos_plus5point1.wav
      - 01_iem_multiencoderatmos.wav
      - binaural
        - 01_iem_multiencoderbinaural_all.wav
        - 01_iem_multiencoder...tmos_plus5point1.wav
        - 01_iem_multiencoderbinaural_atmos.wav
    - soundfield
      - 01_soundfield_all.wav
      - 01_soundfield_atmos_plus5point1.wav
      - 01_soundfield_atmos.wav
      - binaural
        - 01_soundfield_all.wav
        - 01_soundfield_atmos_plus5point1.wav
        - 01_soundfield_atmos.wav
    - sparta
      - 01_sparta_all.wav
      - 01_sparta_atmos_plus5point1.wav
      - 01_sparta_atmos.wav
      - binaural
        - 01_sparta_binaural_all.wav
        - 01_sparta_binaural_atmos_plus5point1.wav
        - 01_sparta_binaural_atmos.wav
  - scene02
    - ambeo
      - 02_ambeo_all.wav
      - 02_ambeo_atmos.wav
      - binaural
        - 02_ambeo_binaural_all.wav
        - 02_ambeo_binaural_atmos.wav
    - harpex
      - 02_harpex_all.wav
      - 02_harpex_atmos.wav
      - binaural
        - 02_harpex_binaural_all.wav
        - 02_harpex_binaural_atmos.wav
    - iem_multiencoder
      - 02_iem_multiencoder_all.wav
      - 02_iem_multiencoder_atmos.wav
      - binaural
        - 02_iem_multiencoder_binaural_all.wav
        - 02_iem_multiencoder_binaural_atmos.wav
    - soundfield
      - 02_soundfield_all.wav
      - 02_soundfield_atmos.wav
      - binaural
        - 02_soundfield_binaural_all.wav
        - 02_soundfield_binaural_atmos.wav
    - sparta
      - 02_sparta_all.wav
      - 02_sparta_atmos.wav
      - binaural
        - 02_sparta_binaural_all.wav
        - 02_sparta_binaural_atmos.wav


scene03
ambeo
03_ambeo_all.wav
03_ambeo_atmos_plus5point1.wav
03_ambeo_atmos.wav
binaural
03_ambeo_binaural_all.wav
03_ambeo_binaural_atmos_plus5point1.wav
03_ambeo_binaural_atmos.wav
harpex
03_harpex_all.wav
03_harpex_atmos_plus5point1.wav
03_harpex_atmos.wav
binaural
03_harpex_binaural_all.wav
03_harpex_binaural_atmos_plus5point1.wav
03_harpex_binaural_atmos.wav
iem_multiencoder
03_iem_multiencoder_all.wav
03_iem_multiencoder_atmos_plus5point1.wav
03_iem_multiencoder_atmos.wav
binaural
03_iem_multiencoder_binaural_all.wav
03_iem_multiencoder...tmos_plus5point1.wav
03_iem_multiencoder_binaural_atmos.wav
soundfield
03_soundfield_all.wav
03_soundfield_atmos_plus5point1.wav
03_soundfield_atmos.wav
binaural
03_soundfield_binaural_all.wav
03_soundfield_binaur...tmos_plus5point1.wav
03_soundfield_binaural_atmos.wav
sparta
03_sparta_all.wav
03_sparta_atmos_plus5point1.wav
03_sparta_atmos.wav
binaural
03_sparta_binaural_all.wav
03_sparta_binaural_atmos_plus5point1.wav
03_sparta_binaural_atmos.wav


scene04
ambeo
04_ambeo_all.wav
04_ambeo_atmos_plus5point1.wav
04_ambeo_atmos.wav
binaural
04_ambeo_binaural_all.wav
04_ambeo_binaural_atmos_plus5point1.wav
04_ambeo_binaural_atmos.wav
harpex
04_harpex_all.wav
04_harpex_atmos_plus5point1.wav
04_harpex_atmos.wav
binaural
04_harpex_binaural_all.wav
04_harpex_binaural_atmos_plus5point1.wav
04_harpex_binaural_atmos.wav
iem_multiencoder
04_iem_multiencoder_all.wav
04_iem_multiencoder_atmos_plus5point1.wav
04_iem_multiencoder_atmos.wav
binaural
04_iem_multiencoder_binaural_all.wav
04_iem_multiencoder...tmos_plus5point1.wav
04_iem_multiencoder_binaural_atmos.wav
soundfield
04_soundfield_all.wav
04_soundfield_atmos_plus5point1.wav
04_soundfield_atmos.wav
binaural
04_soundfield_binaural_all.wav
04_soundfield_binaur...tmos_plus5point1.wav
04_soundfield_binaural_atmos.wav
sparta
04_sparta_all.wav
04_sparta_atmos_plus5point1.wav
04_sparta_atmos.wav
binaural
04_sparta_binaural_all.wav
04_sparta_binaural_atmos_plus5point1.wav
04_sparta_binaural_atmos.wav
scene05
iem_multiencoder
05_iem_multiencoder_all.wav
05_iem_multiencoder_atmos_plus5point1.wav
binaural
05_iem_multiencoder_binaural_all.wav
05_iem_multiencoder...tmos_plus5point1.wav