PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

On the use of spherical microphone arrays in a classical musical recording scenario

Johann-Markus Batke

Hochschule Emden/Leer, Germany, Johann-Markus.Batke@hs-emden-leer.de

Abstract

The usage of spherical microphone arrays in virtual reality production environments have been seen to be steadily increasing. This is understood to be a natural progression since the microphone array‘s output is easily converted to the Higher Order Ambisonics (HOA) sound field representation.

However, for the recording of classical music, there is still much debate as to whether the application of spherical microphone arrays makes any sense at all. Depending on the targeted playback format, various established coincident single microphone setups, as well as spaced setups are preferred. Spaced microphone setups are particularly popular since they ensure a certain degree of decorrelation for the resulting loudspeaker signals. On the other side, microphone array recordings have a much higher degree of freedom with regard to spatial post processing.

This contribution investigates how advantages of both recording strategies can be combined. The recording of a string quartet is used as a well known recording scenario for classical music. A spaced AB recording serves as the basis. A microphone array in a central position within the quartet another one distant from the quartet are used to enhance the AB signals. Different target formats including stereo, surround as well as 3D audio using HOA are discussed.

Keywords: 3D Sound, Higher Order Ambisonics, Production tools

1 INTRODUCTION

Musical content production aims to reproduce a musical experience when the recorded content is played back on loudspeakers or headphones. The spatial information in the content is an important component to reach this goal, of course. Spatial information also plays a role for video production. The production of visual content for virtual reality (VR) or augmented reality (AR) often includes 3-dimensional (3D) sound field information. Here it is important that the spatial information is physically correct and matches the visual impression.

Also for music recording there is a growing demand in 3D audio representations, however the adaption to visual content is often not required. Traditionally channel based approaches like stereo or surround sound are in use. Channel based representations refer to a specific loudspeaker layout like stereo (2.0) or surround sound (e.g. 5.1). Recording music using microphones for such loudspeaker layouts leads to spatial microphone layouts. For classical music recordings often the tonmeister aims to assign one microphone to one channel without further processing. This is the popular concept of using a main microphone (or Hauptmikrofon) [11].

The audio aquisition for VR and AR is often done using microphone arrays. Unlike the music recording case, the output of a microphone array is always processed and adapted to the target playback system. Higher Order Ambisonics (HOA) is often chosen as an intermediate representation [14]. As this representation enables spatial modifications of the soundfield it is an interesting option the the music production case as well.

In the contribution we look at both approaches in a music recording szenario. The "Schunk string quartet" playing compositions of Michael Schunk was recorded using a AB pair of microphones as well as two Eigenmike microphone arrays. The first is approach is to use the AB recording with additional microphone arrays. The AB serves as main microphone, the desired result is a stereo recording. The purpose of the microphone arrays is the enhancement the existing recording used as spot microphones and room microphone respectively. In turn the second approach is to use the microphone array in the middle of the quartet as main microphone and to enhance it using the AB pair. The desired result is a 3D sound file, using HOA as a format.

2 TECHNICAL BACKGROUND

The delivery format for audio content often determines which recording approach is chosen. In general, audio content can be delivered in a channel based, a scene based or an object based representation. The audio standard MPEG-H reflects the existence of all this approaches [4]. In this contribution we look at channel based and scene based representations.

2.1 Channel based production

The channel based audio production is based on a main microphone. Well known main microphone settings for stereo loudspeaker layouts are XY, AB, ORTF, MS or Blumlein [10]. An coincident microphone setting like XY produces an intensity based stereo image. The spaced AB setting produces a stereo image containing mostly time-of-arrival information. A ORTF microphone setting produces both, intensity as well as time-of-arrival information. For 3D loudspeaker layouts microphone settings like ORTF-3D are suggested that refer to a 3D layout with 8 loudspeakers [12].

Spot microphones are usually added in classical recording scenarios to obtain a certain degree of freedom for modifications of the audio content. The spot microphones' signals need delay compensation that is dependent on the main microphone's position. The positioning of the spot in the resulting audio scene (stereo image, if 2D) is done using a panning tool that can address the targeted loudspeaker layout [7, 6].

Coincident microphone settings like MS and microphone arrays need further processing before the signals are played back on the loudspeaker. In this context, the entire processing chain is termed as beamforming [8].

Stereo playback delivers good localisation information when the recording was performed decently [13]. An AB recording delivers decorrelated signals resulting in good spatial imaging when played back on loudspeakers, however these signals need additional processing when played back on headphones as the precedence effect is relevant to loudspeaker production [9]. The spatial imaging of coincident recordings works nicely if monitored on head phones. As a drawback, such recordings may suffer from comb filtering when played back on loudspeakers.

2.2 Scene based production

Scene based production means the creation of a sound field descriptions using Higher Order Ambisonics (HOA) [14]. The recording of natural sound fields for HOA is accomplished using microphone arrays. For 1st order recordings, such arrays are commercially available since the 1970th like the Soundfield microphone by Soundfield or more recently the Ambeo microphone by Sennheiser among others. For higher orders, microphone arrays like OctoMic by Core Sound or the Eigenmike by mh acoustics are available.

The output of these arrays is converted to the HOA representation providing full 3D audio information. In the context of virtual environments like 360°videos or augmented reality scenarios the adaption of the audio scene to the visual scene is feasible using HOA modifications like rotation or spatial warping.

Spot microphones can be used for HOA based productions as well. Like in the channel based case the delay between main microphone and spot position needs compensation. The directional information is encoded by calculation a plane wave HOA representation. In combination with the HOA decoding step this is like panning for the channel based case, this time with a panning function defined by the HOA decoding [1].

The playback of a scene based recording to a number of loudspeakers or a pair of headphones is enabled by decoding the sound field. The spatial imaging depends on the loudspeaker layout or the binaural rendering [5].

To summarise, all spaced main microphone conecpts provide a limited degree of freedom for modifications, whereas array recordings provide various options for post processing like sound field rotation, spatial warping and beam steering.

3 PRACTICAL CONSIDERATIONS

In the string quartet recording investigated in this paper we have an AB pair of microphones and two Eigenmike microphone arrays available. The AB pair and one Eigenmike are positioned near to the string quartet and could be used as main microphone. To gain inside to specific advantages and drawbacks of each method, both options are investigated.

3.1 Material and methods

The recording took place on 3018-02-03 in the "Schweizer Kirche" in Emden. This church is medium sized and provides acoustically a good environment for classical music recordings. Figure 1 gives a view from outside and Figure 2 from inside to gain a general impression.

Figure 1. The "Schweizer Kirche" in Emden is the recording location for the recordings investigated in this paper. The church has a decent reverb appropriate to classical music. Photo by Matthias Süßen

As shown in Figure 2 the string quartet is seated in front of the apse. The AB pair of microphones is in front of the quartet, one microphone array (EM1) is surround by the musicians. The second microphone (EM2) is located apart in the middle aisle.

Figure 2. The string quartet is seated in front of the apse. The AB pair of microphones is in front of the quartet, the first microphone array is surround by the musicians.

The positioning of the microphones while recording is shown in Figure 3. Note that the pictures show a slightly different positioning.

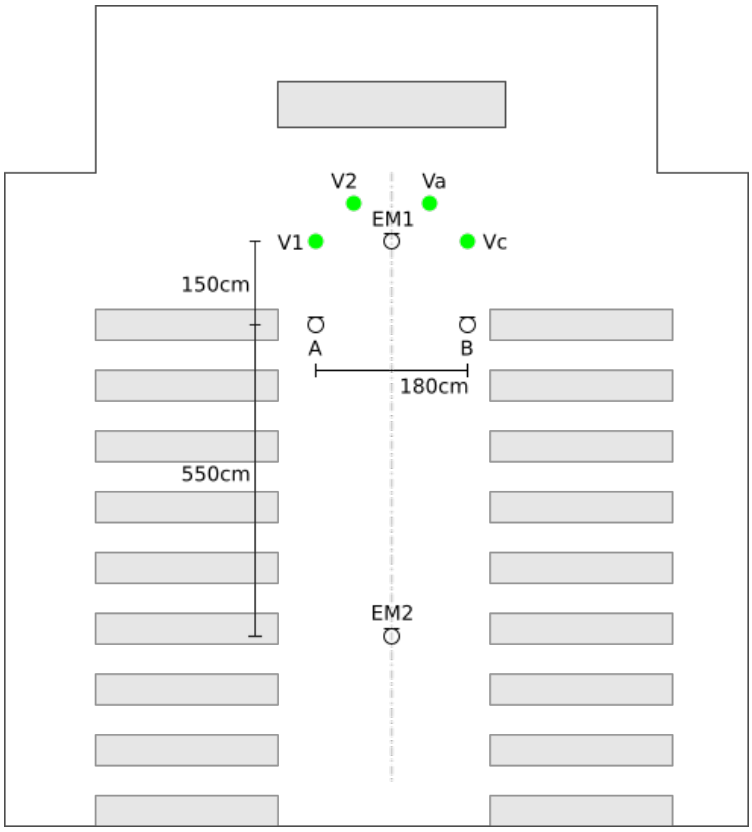

Figure 3. Positioning of the microphones used for recording: in front of the string quartet there is an AB pair of omnidirectional capsules. Inside of the quartet and apart the microphone arrays EM1 and EM2 are located.

The recording devices are summarised in Table 1. All soundcards where synchronised using the clock signal of the soundcard serving as a master. The recording software is Reaper by Cockos. The master slave settings for simultanious start/stop of all three Reaper instances needed for recording all microphones was dropped due to network issues, instead all programms were operated individually. The HOA processing is enabled by the "IEM plug-in suite" as provided by the "institute for electronic music" from Graz [2].

Table 1. Hardware and software used for recording and production.

#	*hardware*	*provider*
2	omnidirectional microphone MK 2	Schoeps
2	microphone array Eigenmike	mh acoustics
1	soundcard Fireface 800	RME
3	misc. notebooks	Apple
	software	
3	Digital audio workstation Reaper	Cockos
1	HOA VST plugins	iem Graz

3.2 AB as a main microphone

When the AB pair serves as main microphone, the quartet as well as a certain amount of reverberation is captured. The "perceived" listening position could be located in front of quartet and AB pair.

The signal of EM1 is then used as source for spot microphone. Four beams are directed to the individual players. The beams are calculated at order $N = 3$. The beams result in four signals that are treated like spot microphones. For the stereo production they are delayed and added to the main signal using the stereo panning tool as provided by Reaper.

The second array EM2 is used in the same manner as EM1. The obtained signals contain room reverberation that is mixed into the main audio image. For surround the EM2 is used as source for the surround loudspeakers.

If a HOA based scene of the AB recording is created, i.e. for headphone playback, the four individual signals can be employed as object sources that are positioned in the scene resulting by the AB recording. The positioning needs to be done according to the actual placement of the players. The AB pairs' signals can be used as main sound sources at the respective positions of the AB array supported by 4 individual sources. As the HOA scene is newly created, a sufficient high order can be chosen to ensure a sufficient spatial resolution.

3.3 Microphone array as main microphone

Using the microphone array EM1 as main microphone sets the listening position to it's location in the recording scenario, that is the center of the string quartet. The AB pair can be added to this audio scene as two additional plane wave sources [14]. This is accomplished using the stereo encoder plugin from the IEM plugin suite. It is practical to chose the same Ambisonics order as used for the microphone array encoding ($N = 3$). As the AB pair is made with omnidirectional capsules, also distance filtering can applied to the pressure signals. They would then appear as point sources in the resulting scene.

Another option is to translate the listening position of the microphone array to the frontal position as chosen in the former section. Translations of sound fields are computationally costly in terms of HOA, though [3].

4 SUMMARY

A string quartet recording was recorded using a spaced AB setting of microphone. Additional microphone arrays where used to capture direct sound and reverberation of the room. Two approaches were discussed: firstly the more traditional approach of music recording, that is the use of the AB pair as a main microphone. Secondly one microphone array was taken as main microphone, the sound field was supplemented by the AB pair.

References

[1] Johann-Markus Batke and Florian Keiler. "Using VBAP-derived panning functions for 3D ambisonics decoding". In: *Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics*. Paris, France, May 2010. URL: http://ambisonics10.ircam.fr/drupal/files/proceedings/presentations/O14_47.pdf.

[2] Institute of Electronic Music and Acoustics, eds. *IEM Plug-in Suite*. The IEM Plug-in Suite is a free and Open-Source audio plugin suite including Ambisonic plug-ins up to 7th order created by staff and students of the Institute of Electronic Music and Acoustics. Version v1.10.2. May 6, 2019. URL: https://plugins.iem.at/.

[3] Nail A. Gumerov and Ramani Duraiswami. *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions*. Amsterdam: Elsevier, 2004.

[4] Jürgen Herre et al. "MPEG-H Audio—The New Standard for Universal Spatial/3D Audio Coding". In: *J. Audio Eng. Soc* 62.12 (2015), pp. 821–830. URL: http://www.aes.org/e-lib/browse.cfm?elib=17556.

[5] ITU-R. *Multichannel sound technology in homeand broadcasting applications*. Report BS.2159-4. International Telecommunication Union, 2012. URL: https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-BS.2159-4-2012-PDF-E.pdf.

[6] Mark Poletti. "Robust Two-Dimensional Surround Sound Reproduction for Nonuniform Loudspeaker Layouts". In: *J. Audio Eng. Soc.* 55.7/8 (July 2007), pp. 598–610.

[7] Ville Pulkki. "Virtual Sound Source Positioning Using Vector Base Amplitude Panning". In: *J. Audio Eng. Soc.* 45.6 (June 1997), pp. 456–466.

[8] Boaz Rafaely. *Fundamentals of Spherical Array Processing*. Springer, 2015.

[9] Francis Rumsey. *Spatial Audio*. Focal Press. 2001.

[10] Francis Rumsey and Tim McCormick. *Sound and Recording*. Focal Press, 2009. ISBN: 0240521633, 978-0-240-52163-3.

[11] Helmut Wittek. *hauptmikrofon.de*. URL: https://hauptmikrofon.de.

[12] Helmut Wittek and Günther Theile. "Development and Application of a Stereophonic Multichannel Recording Technique for 3D Audio and VR". In: *Audio Engineering Society Convention 143*. Oct. 2017. URL: http://www.aes.org/e-lib/browse.cfm?elib=19266.

[13] Helmut Wittek and Günther Theile. "The recording angle - based on localisation curves". In: *Audio Engineering Society Convention 112*. May 2002.

[14] Franz Zotter and Matthias Frank. *Ambisonics*. Springer, 2019.