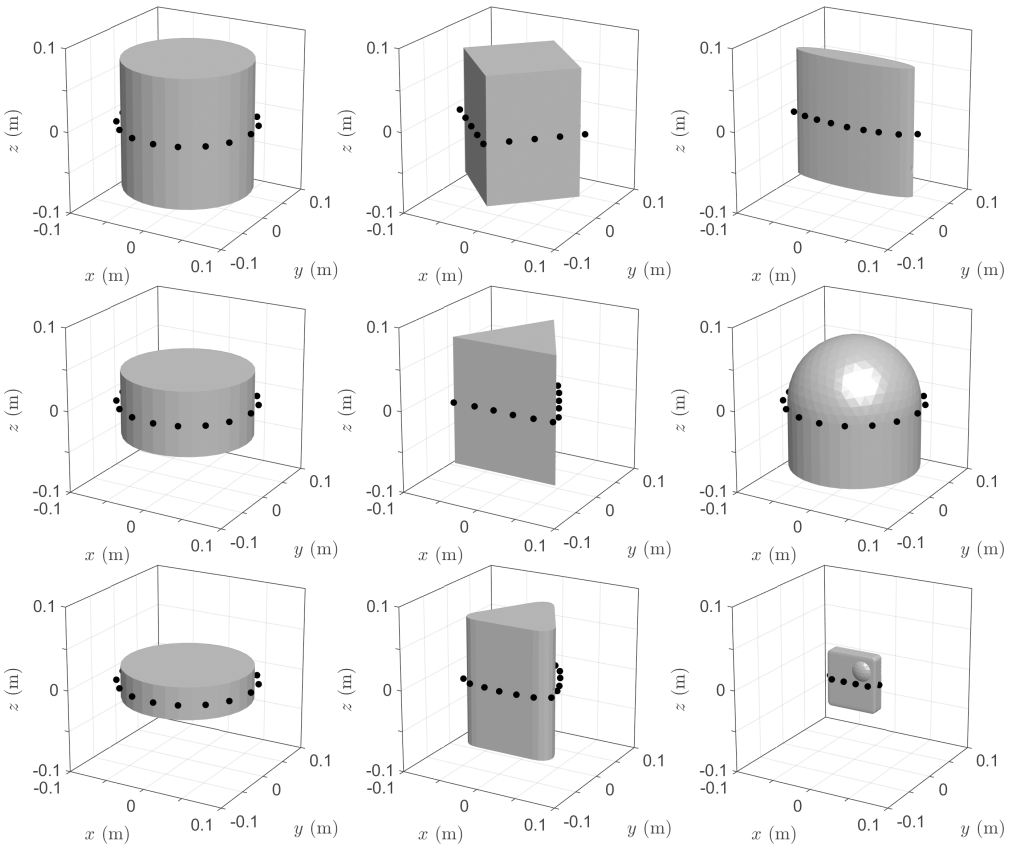
CHALMERS
UNIVERSITY OF TECHNOLOGY



The investigation of eXMA method with non-spherical scatters

Master's thesis in Sound and Vibration

ZIYI HU

Department of Architecture and Civil Engineering

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2022
www.chalmers.se

Master's thesis 2022

The investigation of eXMA method with non-spherical scatters

ZIYI HU

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Architecture and Civil Engineering
Division of Applied Acoustics
Audio Technology Group
Chalmers University of Technology
Gothenburg, Sweden 2022

The investigation of eXMA method with non-spherical scatters
ZIYI HU

Supervisor: Jens Ahrens, Department of Architecture and Civil Engineering
Examiner: Jens Ahrens, Department of Architecture and Civil Engineering

Master's Thesis 2022
Department of Architecture and Civil Engineering
Division of Applied Acoustics
Audio Technology Group
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: The baffle shapes that we considered in our study.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2022

The investigation of eXMA method with non-spherical scatters
ZIYI HU
Department of Architecture and Civil Engineering
Chalmers University of Technology

Abstract

The XMA was a recently presented higher-order ambisonic microphone array which is based on the spherical microphone array (SMA) and equatorial microphone array (EMA) but without a traditional spherical scattering body. Since it is compatible with the EMA, the XMAs can also be designed with the microphones placed on a circumferential contour around the scattering body, which is called the equatorial XMA (eXMA). Compared with the classical SMAs, the eXMA method reduced the required number of microphones significantly since it did not need the microphones to be distributed over the whole surface of the scatterer. The eXMA shows a good application prospect in spatial sound field recording especially when combined with the VR camera to produce a complete panoramic audio-visual experience from a first-person view. However, the eXMA has so far only been evaluated as a head-mounted array, i.e. with a human head as the baffle. The performance of eXMA with other shapes of scatterers are unknown.
In this work, we used the mesh2hrtf implementation of the boundary element method (BEM) to simulate eXMA calibration measurements for a variety of candidate scatterers including cylinders, cubics and some shapes that are inspired from real VR 360 cameras. We also deformed those shapes and moved up the microphone array to see the influence. Based on those simulations, we identify what spherical harmonic orders can be obtained with what accuracy for a set of convex scattering body geometries that are of relevance in the given context.
We demonstrate that the shape of the body is not very critical. The eXMA shows very robust performances with the different shapes of scatterers, some of them even have corners. Reducing the height of the scatterers or moving up the microphone array to the edge will increase the error but the accuracy is still acceptable. The main limitation is the size of the scatters that small bodies do not allow for extracting higher orders at low frequencies. Limitations of the simulation are discussed and at the end we also generate some spatial audio recordings based on the cuboid and the squashed cylinder scatterers.

Keywords: Spatial Audio, SMA, EMA, XMA, Ambisonic, Spherical Harmonics.

Acknowledgements

Thanks to my supervisor Jens Ahrens, his professional knowledge, patient instruction brought me to the high end world of the new ambisonic recording technology. And his optimistic and positive attitude really helped me a lot during the master thesis process.

Thanks to Hannes Helmholz, without his kindly help in equipment setting and software adjusting, I can not finish the recording process smoothly.

Thanks to Yun Zhou and Wenkang Liu, their good instrument playing in the video helped us to really feel the interesting ambisonic performance of our XMA method.

Thanks to my parents, thanks to their support for my study in Chalmers. And thanks to all the other people that helped me during my daily time in Sweden.

Due to the Covid, I only spent one and half years in Sweden. The time in Chalmes, in Gothenburg and in Sweden is quite impressive and important. I think I will really miss this period. And I hope I can meet you guys again soon.

Ziyi Hu, Changsha, 11 2022

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

AR	Augmented Reality
EMA	Equatorial Microphone Array
FOA	First-order Ambisonic
HOA	High-order Ambisonic
HRIR	Head Related Impulse Response
HRTF	Head Related Transfer Function
HRTF	Head Related Transfer Function
MS	Mid Side
ORTF	Office de Radiodiffusion Télévision Française
SH	Spherical Harmonics
SMA	Spherical Microphone Array
VR	Virtual Reality
XMA	X Microphone Array

Nomenclature

Below is the nomenclature of indices and parameters that have been used throughout this thesis.

Indices

m	Index for the degree of spherical harmonics
n	Index for the order of spherical harmonics
q	Index for the number of microphones
l	Index of horizontally propagating plane

Parameters

α	Azimuth
β	Colatitude
ω	The radian frequency
c	The speed of sound
L	The total number of horizontally propagating planes
M	The maximum resolvable SH degree
N	The maximum resolvable SH order
Q	The total number of the microphones
R	The radius of the spherical baffle

Coefficients and Functions

S^{surf}	The sound pressure field on the surface of the baffle
$\mathring{S}^{\text{surf}}_{n,m}$	The SH coefficient of the sound pressure on the surface of the baffle
$\breve{S}_{n,m}$	The SH coefficients of the incident sound field

$\mathring{H}_{n,m}^{\mathrm{L,R}}$	The user's left and right ear's SH coefficients of the head related transfer function
$\mathring{S}_{n,m}^{\mathrm{surf,pw}}$	The SH coefficients of plane wave sound field
$Y_{n,m}$	The Spherical Harmonics basis function
$P_n^{\lvert m\rvert}$	The associated Legendre function

Contents

List of Figures

1

Introduction

The videos for virtual reality (VR) and augmented reality (AR) are typically captured by a dedicated camera array and stitched together to form a panoramic video. People can look around when using VR or AR headsets. To make users also hear "around", we need to produce a head-tracked binaural audio. This audio Though this method sacrifices some spatial resolution on the elevation direction, it still provides some good advantages [2]. Firstly, the Equatorial Microphone Array (EMA) method reduced the number of required microphones and only distributed them around the equator of the spherical scattering object. With appropriate scatters, it can achieve the same spatial resolution as the Spherical Microphones Arrays (SMA) method but with much fewer microphones. Another important advancement of the EMA method is that it provides an SH can be captured with rigid SMA. The shape of the array and the number of microphones in SMA limited the application of this method. It is hard to be mounted onto AR headsets additionally to a camera array and impractical to create a head-mounted array since the microphones need to be distributed over the entire head. Based on the SMA method, Ahrens et al [7] proposed the EMA. Decomposition of the captured sound feld from microphones that are on a circumferential contour on the scattering object" [6], which means we can extend the EMA solution to non-spherical objects like the human head. The extensibility of this approach gave us a possibility to create a head-mounted array in practice. For arbitrarily- shaped compact scatterers, Ahrens et al [7] extend the SMA and EMA methods to sXMA and eXMA methods. The result shows that the accuracy of eXMA is only slightly lower than the EMA method with a spherical scatterer.

In the previous study, the eXMA method shows a potential applicability in non-spherical scatterers. So in this thesis, we will focus on how far the shape of a scatterer can deviate from a sphere. A metric to evaluate the accuracy of eXMA with arbitrarily-shaped scatterers also needs to be found. This metric will help us to confirm whether the given shape is suitable or not. After some interesting and suitable shapes have been found, we will build prototypes for them and apply the eXMA method to make some recordings. These nice recordings will show us how well this method works in the real world.

1.1 A brief introduction of spatial audio recording and presentation

The term spatial audio has a wide usage covering from the digital signal field to the psychology field but does not have a well recognized definition yet [3]. In this thesis, we mainly focus on the audio signal that contains spatial information.

1.1.1 Stereo and Surround Sound

The stereophony and surround sound presentation always use two and more audio channels and can provide more spatial information of the sound field compared to a single microphone recording [30]. There are basically four types of stereo recording configurations: Spaced pairs, Coincident pair, Near-coincident pair and Mid-side [18]. Fig 1.1 shows the typical microphone settings of these four types:
i) AB stereo recording using a spaced pair of omnidirectional microphones, the distance between two microphones is about 50cm. The distance makes a time delay between the two microphones and provides a very open and spacious sound [22].
ii) XY stereo recording is based on a coincident pair of directional- or bidirectional, angled microphones. These two microphones are placed so close that they can be considered at the same location. This provides a perfect position accuracy but less spatial compared with the AB type [22].
iii) ORTF (Office de Radiodiffusion Télévision Française) stereo recording is based on a near-coincident pair of cardioid microphones. The distance between two microphone heads is 17cm and the angle between them is 110 degrees. This distance mimics the space between two human ears and this angle emulates the shadow effect of the human head, which gives the listener a good effect both in position accuracy and spaciousness [22].
ix) Mid-side stereo recording is based on a cardioid microphone (Mid) at the center point to the stage that picks up mono sound and a bi-directional microphone (Side) placed down side the mid microphone to pick up signals from left and right side. The spatial information is mainly recorded by the side microphone. The position accuracy of MS setting is also perfect because the two microphones are at the same place [18].

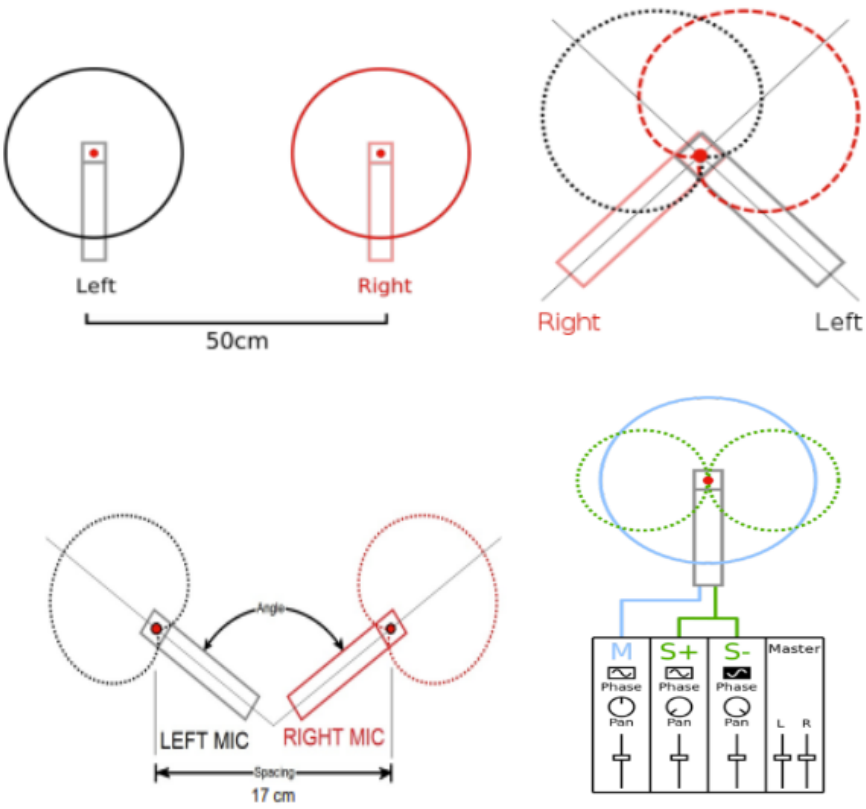


Figure 1.1: Four types of stereo recording configurations [4]. Top Left: AB stereo recording. Top Right: XY stereo recording. Bottom Left: ORTF (Office de Radiodiffusion Télévision Française) stereo recording. Bottom Right: Mid-side stereo recording.

Fig 1.2 shows the typical setup of the stereo present system. In stereophony, the spatial perception of the listener is controlled by increasing the amplitude or delaying the signal of one loudspeaker. When the sound from the two loudspeakers arrive at both ears at the same time with the same amplitude, it appears to originate from a point in the center of the two speakers (phantom source) [55]. Increasing the amplitude of one loudspeaker's signal will move the perceived location of the phantom source towards this loudspeaker, delaying the signal of one loudspeaker will move the perceived location of the phantom source away from this loudspeaker. These two phenomena are called amplitude panning and delay panning respectively [3].
How the time gap between the two arriving signals influenced people's perception of the stereo sound has been studied by several researchers:
i) When the time interval is smaller than 1ms, this difference can not be distinguished by the hearing system. This phenomenon that the superpositioned sound fields summed up at listeners' ears is referred as summing localization in psychoacoustic [3, 17, 48].
ii) When the time interval is between 1ms to 40ms, the precedence effect (Haas effect) takes place [3, 28]. In this time window, the perceived direction of the first arrived sound will not be influenced by the later sound that comes from a different direction, the later sound will not be perceived as an echo but as a room impression.

However, if the second-arriving sound is at least 15 dB louder than the first, the precedence effect breaks down [3, 32].
iii) When the interval is larger than 40ms, the later signals will be perceived as distinct echoes [17].

The surround sound system is an extension of the stereophony. Fig 1.2 Right shows the typical setup of the 5.0 surround system. The three added loudspeakers have different usage: the center loudspeaker is used to present the sound content that is supposed to stay at the center steadily. The left and right surrounded loudspeakers are used to increase the spatial perception by playing the decorrelated signals [30].

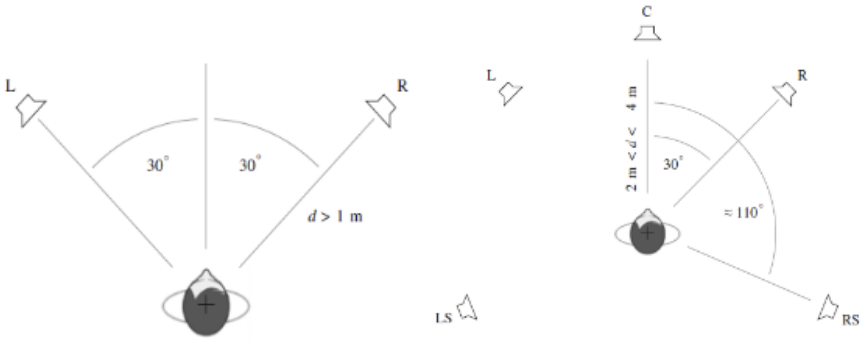


Figure 1.2: The typical setup of the stereo present system [3]. Left: the 2.0 surround system. Right: the 5.0 surround system.

1.1.2 Head-Related Transfer Functions and Binaural Recording

To create an audio scene with specific spatial attributes that makes the listener feel like in a real acoustic environment, the presentation system needs to imitate the acoustic properties of humans. As the sound propagates from the source to the listener, there are a lot of other humans' factors except the surrounding environment that will determine what we will hear. For example: the size and shape of our head, ears, ear canal, density of the head, size and shape of nasal and oral cavities [10, 29]. All of these human properties will transform the sound like boosting some frequencies and attenuating others and thus affect how we perceive it. These properties that characterize how an ear receives a sound from a point in space are called head-related transfer functions (HRTFs), the effect of HRTFs is shown in Fig 1.3. HRTFs for left and right ear $h_R(t)$ and $h_L(t)$ describe the filtering of a sound source x(t) before it is perceived at the left and right ears as $x_L(t)$ and $x_R(t)$ [29].

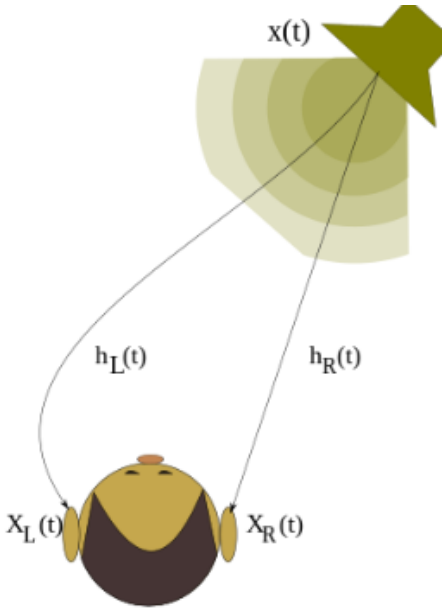


Figure 1.3: The effect of human head related functions [29].

As an extension of stereo recording, binaural recording can capture the HRTFs and accurately simulate human hearing through a calibrated headphone playing system. As shown in Fig 1.4, by setting the microphone at the position of the eardrum inside the ear channel of human [49] or a dummy head [31], this recording method can easily capture the natural sound alteration due to the interaction with human body and surround environment [30].

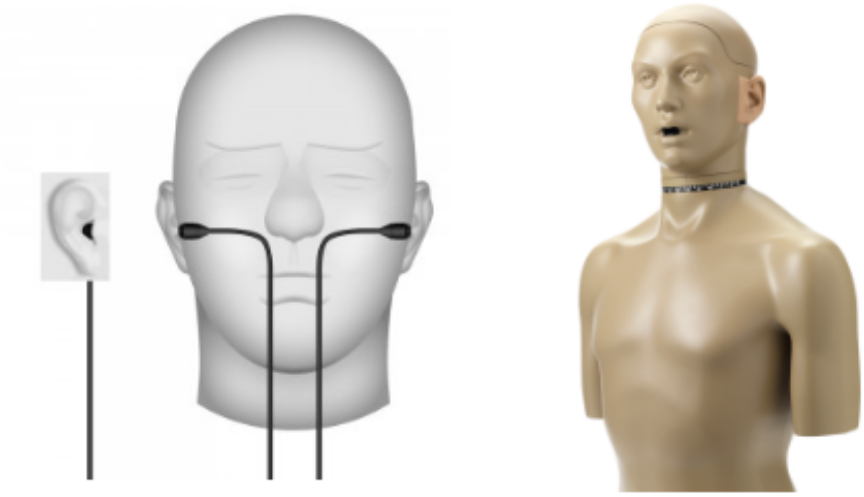


Figure 1.4: The Binaural Stereo recording [46] and the KEMAR dummy head [31]
.

With the help of head related impulse response (HRIR) which is the inverse Fourier transform of HRTF., we can simulate human hearing of an arbitrary sound source through a calibrated headphone playing system. People estimate the location of a

source comparing the cues received at both ears. The two cues may have time and intensity differences. By convolving the head related impulse response (HRIR) with the arbitrary sound source, we can convert the sound to what the listener will hear if the sound had been played at the source location with the listener's ear at the microphone location. And thus the listener can feel the virtual sound source with a proper playing set [42].

1.1.3 The Acoustic Curtain

To better capture and present spatial information, it is intuitive to apply more microphones and loudspeakers. Snow et.al [45] introduced the idea of the Acoustic Curtain in the 1930s which formed the basis of the sound field synthesis [3]. Fig 1.5 shows the concept of acoustic curtain. The sound is captured by a wall of microphones in the primary recording room. Each microphone is directly connected to a loudspeaker that is settled on the wall of the secondary reproduction room. The listener is assumed to hear sound coming from a virtual source that is located not in the primary recording room but just behind the curtain of loudspeakers. Moreover, the system was percepted satisfied even with only two or three microphones and loudspeakers. This can be explained by the hearing mechanism that similar to the Stereophony was triggered [3].

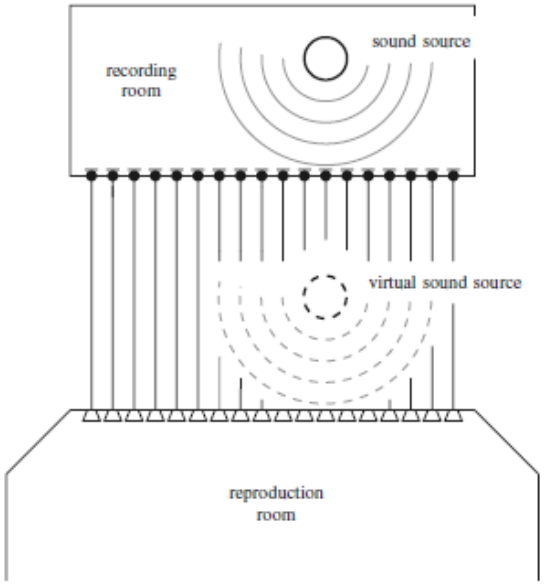


Figure 1.5: The concept of acoustic curtain [3].

The reproduced sound field is very realistic with a good matching between the microphones and loudspeakers [30,40]. But a perfectly matched record and playback system is not practical since the configuration of the system (position and number of microphones and loudspeakers) varys a lot with different room settings [30]. To avoid this problem, nowadays microphone recording signals will be processed rather than used directly. The processing part will analyze the signal and provide more

information about the sound field including the source position, the characteristics of the environment and also a clean sound source signal [40].

1.1.4 Ambisonics

The ambisonics technology was firstly developed in the 1970s by Felgett [24], Gerzon [25], and Craven [20]. Gerzon's work in the 1970s gave us what we call first-order Ambisonic recording and playback technology today [25]. We briefly introduce his playback setup here: at the beginning, an arrangement of loudspeakers as shown in Fig 1.6 was used to control the sound field inside the loudspeaker arrangement which was adapted from Gerzon [25]. Nowadays, the ambisonic format can be used to represent spatial audio information by spherical harmonic expansion coefficients and it allows for head-tracked binaural reproduction over headphones as well as for reproduction over loudspeaker arrays or even conventional stereo and surround setups [8].

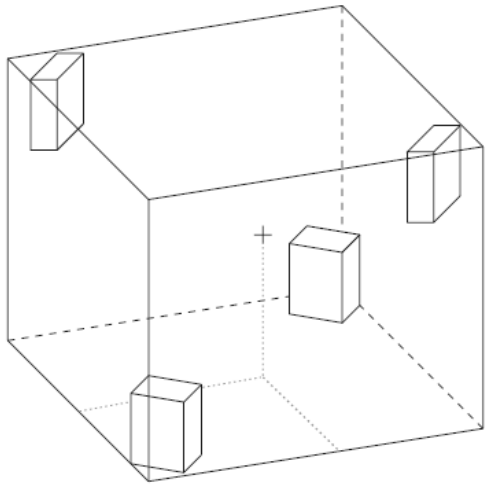


Figure 1.6: A typical ambisonic loudspeaker setup; the mark indicates the point in which the sound field is controlled [3].

1.1.4.1 First-order Ambisonic (FOA) recording

The first-order Ambisonic recording technology benefits a lot in VR and sound field recording [56]. The microphone setup of the first-order Ambisonic (FOA) recording system is similar to the MS stereo recording, As shown in Fig 1.7 left, a 2D first-order Abisonic recording system typically only needs one more figure-of-eight microphone compared with MS recording. The three signal channels of the 2D FOA format are called W, X and Y which correspond to the omnidirectional microphone and two figure-of-eight microphones that are aligned with the Cartesian axes [56]. Fig 1.7 right shows the 3D FOA format which introduces one more figure-of-eight microphone to pick up the signals aligned with the Z Cartesian axes.

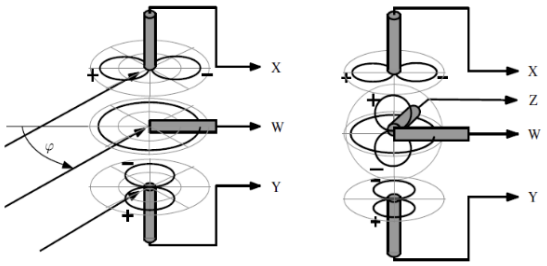


Figure 1.7: Two types of the first-order Ambisonic recording microphone configurations [56]. Left: the 2D FOA recording system. Right: the 3D FOA recording system.

The 3D first-order Ambisonic recording can also be realized by using 4 tetrahedral arranged cardioid microphones with the aiming directions FLU-FRD-BLD-BRU (front-left-up, front-right-down, back-left-down, back-right-up) [56], as shown in Fig 1.8. This kind of arrangement is widely used in practice, Fig 1.9 shows some ambisonic microphones in the commercial market, they are Sennheiser AMBEO VR Microphone [9], Core Sound TetraMic [19], and SoundField SPS200 [44]. Nowadays, the first-order Ambisonic recording technology's benefits of capturing the whole surrounding sound scene with only 4 compact microphones and easily permits rotation of the sound scene makes it popular in video recording, rendering and VR games [56].

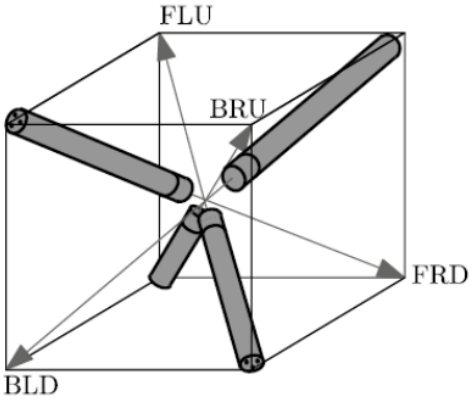


Figure 1.8: Tetrahedral array with four cardioids [56].

Figure 1.9: Some ambisonic microphones that applied a tetrahedral array with four cardioids in the commercial market. From left to the right: Sennheiser AMBEO [9], Core Sound TetraMic [19], and SoundField SPS200 [44].

1.1.4.2 High-order Ambisonic (HOA) recording

The design of HOA microphones is a turning point that opens an exciting experimental field in real 3D sound field recording [21]. The higher-order Ambisonic microphone arrays are always distributed over a rigid sphere with some radius r [8,56]. This kind of microphone array is called spherical microphone array (SMA). SMA now is commercially available, as shown in Fig 1.10, the em32 Eigenmike microphone array [23] and the B&K spherical microphone array [15]. Whereby this model even have a set of video cameras integrated so that the complete audio-visual data are recorded. Compact spherical microphone arrays can record sound events from the first-person perspective. We will introduce this kind of microphone array and its derivative in detail later.

Figure 1.10: Some commercially used spherical microphone arrays. Left: em32 Eigenmike microphone array with 32 channels [23]. Right: BK spherical microphone array with 36-50 channels [15].

1.2 Spherical microphone arrays (SMAs) and Equatorial microphone arrays (EMAs)

Spherical microphone arrays (SMAs) is a convenient solution for capturing and analyzing spatial sound fields and shows the property of sound incident angle independence [1, 35, 37, 54]. The typical setting of SMAs were shown in Fig 1.11 left, where plenty of microphones were uniformly distributed on the rigid spherical baffle. Spherical harmonics (SHs) were applied to represent the sound field which was captured by SMA and it facilitates the application in binaural rendering [38]. With the help of binaural rendering, we can compute the signals that would arise at the listener's ears when exposed in the captured sound field. The disadvantage of SMA is obvious: to realize the direction-independent spatial resolution, it requires a significant number of microphones to be placed on the surface of a rigid sphere.

However, since most sound scenes in real world have much greater change in azimuth than in elevation [16] and our listening system is also optimized for that [17], it is enough to restrict our consideration in the horizontal plane [2, 6, 7]. Though compared with SMA, the circular microphone arrays have a bad ability to discriminate sound in elevation, it is not necessary to provide the same resolution for all directions in practice [14]. And actually the listener's acuity of azimuthal is better than their evational acuity when they spatialized the spatial sound [14]. Under this context, a circular microphone array with appropriate scatters and much fewer microphones may achieve the same spatial resolution is possible [6].

The equatorial microphone array (EMA) was established by Ahrens et.al [6] in 2021. As a variant of SMA, the EMA also has a spherical scatter but the microphones are placed along the equator of the sphere rather than distributed over the surface. The audio quality with playback over headphone is very good when the SH order is N=5 or higher [8, 13, 34, 51]. The limitation of EMA is that it cannot produce the real ambisonic representation of the captured sound field but only a horizontal projection of the sound field. But it saves a significant number of microphones for a given SH order N: for SMA, the required number of microphones is (N+1)2 but for EMA, it only needs 2N+1 [8].The EMA makes it possible to create an array with the SH orders that is unacceptable for SMAs. Fig 1.11 shows a comparison between an 8th-order SMA with 110 microphones and 8th-order EMA with 17 microphones.

To better understand the meaning of SH order, we employed simple visual representations of the first few real spherical harmonics in Fig 1.12. We can consider the shape of each mode just as the directivity of the microphone array. The 0th mode is omnidirectional [8] and the shape of the 1st mode is very similar to the directivity of a figure of 8 microphone. With higher order, the spatial resolution will be higher, and to achieve this, we will need more microphones. It is enough to capture sufficient spatial information with only 4th or 5th order of microphone array.

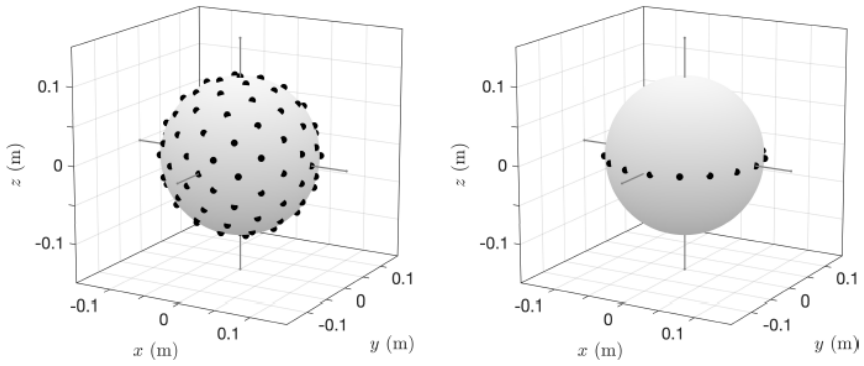


Figure 1.11: The comparison of microphone numbers between two methods [2]. Left: the 8th-order SMA with 110 microphones. Right: the 8th-order EMA with 17 microphones.

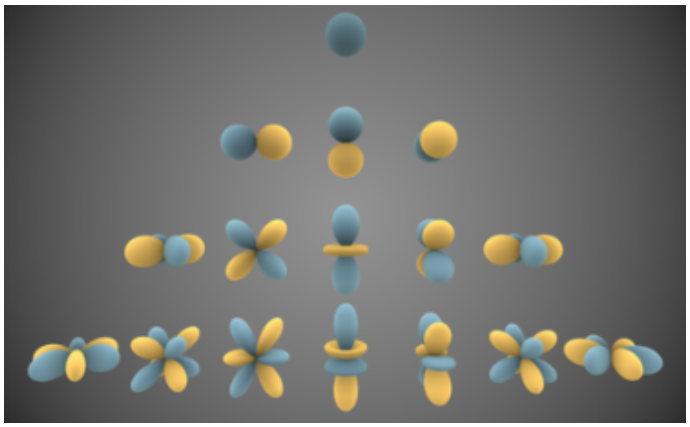


Figure 1.12: Visual representations of the first few real spherical harmonics, from the top to the bottom is the 0th order to the 3rd order [47].

1.3 The XMA

The SMA and EMA both have a good performance in spatial audio representation, but they are only workable with a rigid spherical scattering object and thus the usage of spatial microphone arrays is limited. For example, it is inconvenient to use an ambisonic microphone array with spherical baffle [8] on the wearable arrays. Combining the concepts of SMA and EMA, Arhens et.al developed the XMA that don't need a spherical scatter, the X means the arrays can be mounted on arbitrarily-shaped scatters [7]. In the concept of XMA, Arhens used the term sphere-like XMA (sXMA) to describe the microphone arrays that distributed over the surface of arbitrary scatter like Fig 1.13 left and the term equatorial XMA (eXMA) to the describe the microphone arrays that located along the equator of the arbitrarily-shaped scatter like Fig 1.13 right.

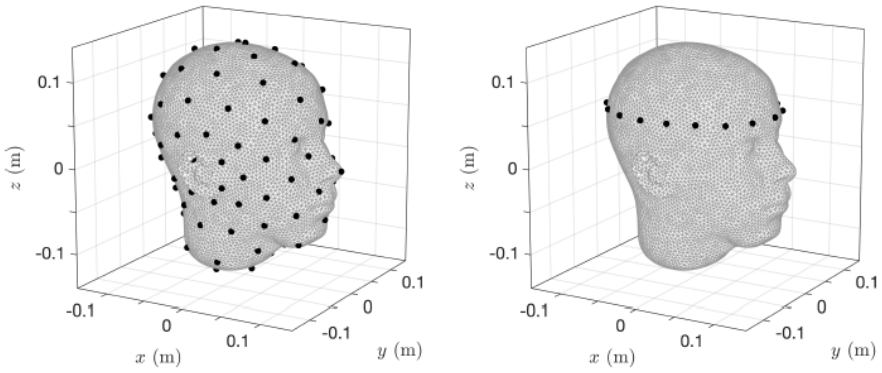


Figure 1.13: The microphone distribution of the sXMA (right) ang the eXMA (left) on the human head [6].

Nowadays, the XMA method has only been used as a head-mounted microphone array [6], which can be integrated into augmented reality (AR) glasses as shown in Fig 1.14. The combination of AR glasses and ambisonic microphone array makes the headset become a self-sufficient multimedia capture device [6]. Obviously, it is impractical to use sXMA in this condition since you cannot place the microphones all over your face and here in this paper, we will only talk about the eXMA.

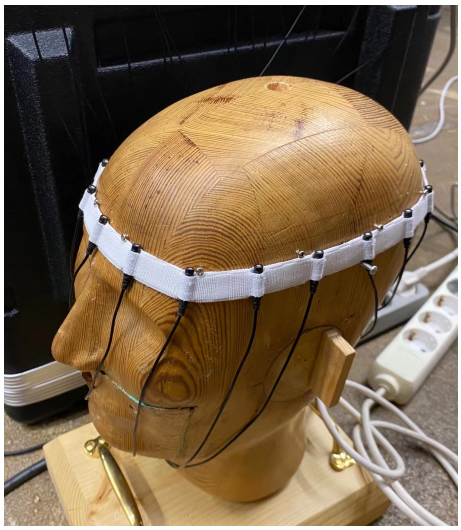


Figure 1.14: The eXMA head-mounted array.

Many literature research have demonstrated that the XMA method is robust against mismatching and displacement of the microphones, which means the effect of representation will not be affected too much in case the microphones are slightly moved. It was also shown that XMAs work pretty well for binaural reproduction when the size of the baffle is close to the human's head [7]. So that, in this paper we only apply baffles with the size very close or just little smaller than the human's head to the XMA. We evaluate the performance of baffles that depart more from a spherical shape, including cylinders, different prisms, commercial 360° VR camera like shapes and some other strange shapes. For the different form factors, we investigate via

numerical simulations which SH order can be reliably retrieved from the microphone signals at what frequency.

1.4 The cost of height variant

The most different between the EMA and XMA to the SMA could be the distribution of the microphones. In SMA the microphones are distributed all over the surface of the baffle, but the microphones of EMA and XMA are concentrated to the equatorial. This arrangement requires the incident sound field for EMA and XMA to be height invariant.

How much will it cost for the EMA and eXMA if the incident sound field is not height invariant? Will it influence the accuracy of the results significantly? In the previous study, Ahrens et.al [2, 6, 7] have compared the HRTFs with the binaural rendered output of the SMA, EMA and eXMA in different elevation degrees. In their work, the results of HRTFs, SMA, EMA and eXMA does not show any obvious difference when the elevation of incident sound is 0 degree. However, when they raise the elevation of the source, the results start to vary. The difference between HRTFs and SMA is still not significant. But the magnitude of EMA and eXMA deviated from the HRTFs obviously, especially when the sound comes from straight above (the elevation degree is 90). This deviation is audible but we are not sure how it will affect people perceptually. Fig 1.15 shows the comparison between HRTFs and SMA and EMA. The data is come from [6] For convenience, we will limit our consideration under the height invariant sound field. For VR, it is sufficient to restrict the considerations to the horizontal plane because this is the most common real-world scenario, and the human auditory system is optimized for it.

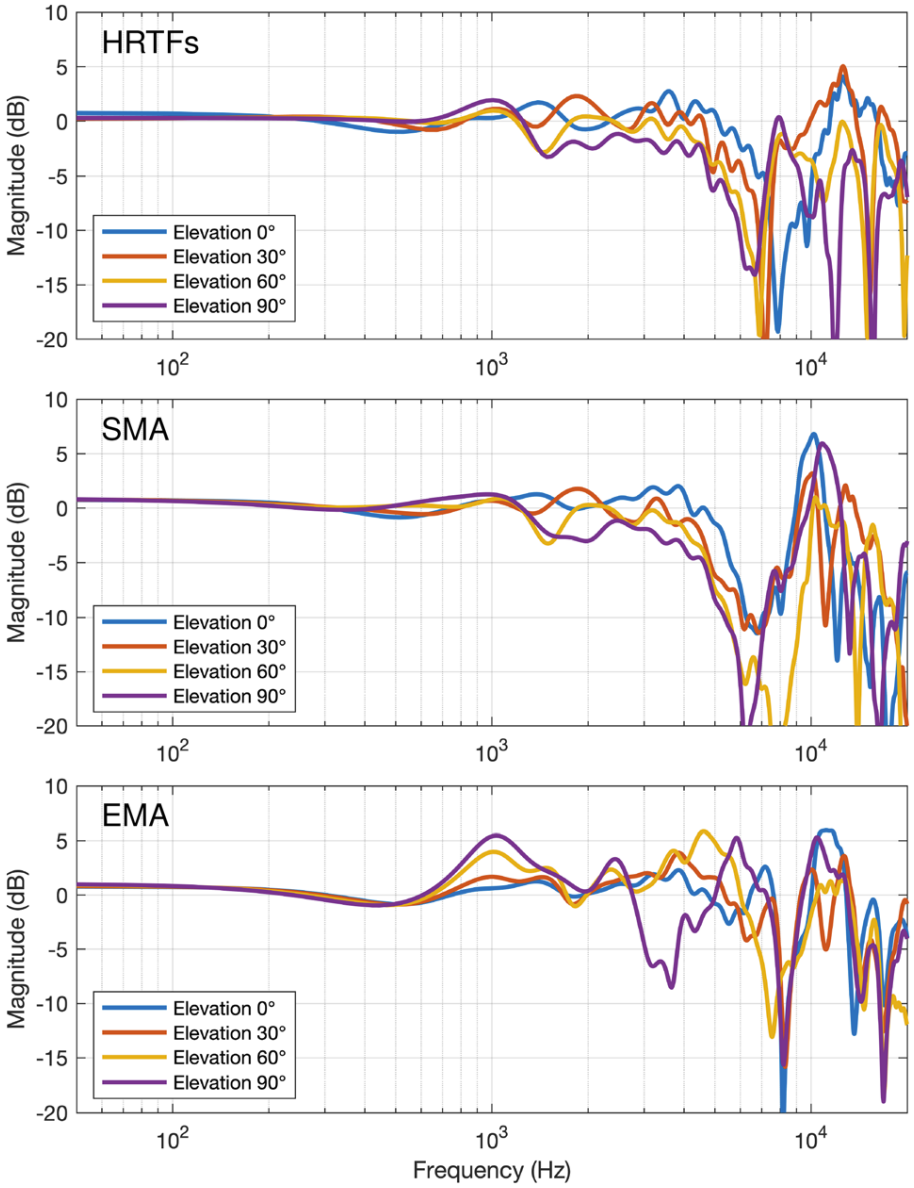


Figure 1.15: The comparison between HRTFs and SMA and EMA [6].

2

Theory

In this section, we will first briefly revisit the spherical microphone arrays (SMAs) and the equatorial microphone arrays (EMAs) since these two methods formed the fundamentation of XMAs. Then we will introduce the basic theory of XMAs and how we investigate the influence of baffle's shape on eXMAs.

2.1 Basic principle of different type of arrays

2.1.1 Spherical microphone arrays (SMAs)

As shown in Fig 1.10, conventional SMAs employ pressure microphones distributed over a rigid spherical baffle. The sound pressure field $S^{\text{surf}}(\beta, \alpha, R, \omega)$ on the surface of the baffle is given by

$$S^{\text{surf}}(\beta, \alpha, R, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \mathring{S}_{n,m}^{\text{surf}}(R, \omega) Y_{n,m}(\beta, \alpha) \tag{2.1}$$

where the baffle is centered at the coordinate origin, R is the radius of the spherical baffle, β and α are the colatitude and azimuth, $\omega = 2\pi f$ is the radian frequency in rad/s. The order n denotes the maximum order that can be extracted by the microphone array to n<N. One speaks this an Nth order decomposition, in SMAs, the required microphone number for Nth order is $(N+1)^2$. $Y_{n,m}(\beta, \alpha)$ are the SH basis functions,

$$Y_n^m(\beta, \alpha) = (-1)^m \sqrt{\frac{2n+1}{4\pi} \frac{(n-|m|)!}{(n+|m|)!}} P_n^{|m|}(\cos\beta) \mathrm{e}^{\mathrm{i}m\alpha} \tag{2.2}$$

the $P_n^{|m|}(\cos\beta)$ are the associated Legendre functions.

According to eq 2.1, the SH coefficient of the sound pressure on the surface of the spherical baffle $\mathring{S}_{n,m}^{\text{surf}}(R, \omega)$ is given by

$$\mathring{S}_{n,m}^{\text{surf}}(R, \omega) = \oint_O S^{\text{surf}}(\beta, \alpha, R, \omega) Y_{n,m}(\beta, \alpha)^* \, \mathrm{d}\Omega \tag{2.3}$$

where the $*$ denoted complex conjugation. The sound pressure field $S^{\text{surf}}(\beta, \alpha, R, \omega)$ here is actually the signal of the sound field captured by the microphones located all over the surface of the rigid sphere.

In practice, the integration in eq 2.3 will be approximated by summing the signals captured by the distributed microphones,

$$\mathring{S}_{n,m}^{\text{surf}}(R,\omega) = \sum_{q=1}^{Q} w_q S^{\text{surf}}(\beta_q, \alpha_q, \omega)\, Y_{n,m}(\beta_q, \alpha_q)^* \tag{2.4}$$

whereby w_q are the quadrature weights of the microphone locations, the q denotes the number of the microphones from No.1 to No.Q.

The next step is to calculate the SH coefficients of the incident sound field $\breve{S}_{n,m}(\omega)$. It is the target to realize the representation, it represents the captured sound field with the effect of the scatterer removed [6].

$$\breve{S}_{n,m}(\omega) = \mathring{S}_{n,m}^{\text{surf}}(R,\omega) b_n^{-1}(R,\omega) \tag{2.5}$$

with,

$$b_n(R,\omega) = -\frac{\mathrm{i}}{\left(\omega\frac{R}{c}\right)^2} \frac{1}{h_n'^{(2)}\left(\omega\frac{R}{c}\right)} \tag{2.6}$$

$h_n'^{(2)}$ denotes the derivative of the nth order spherical Hankel function of second kind.

The next step is to calculate the binaural rendering signal with the help of the user's left and right ear's SH coefficients of the head related transfer function (HRTF) $\mathring{H}_{n,m}^{\text{L,R}}(\omega)$ [5]

$$B(\omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{1}{4\pi \mathrm{i}^{-n}} \breve{S}_{n,m}(\omega)\, \mathring{H}_{n,m}(\omega) \tag{2.7}$$

The signal $B_{\text{L,R}}(\omega)$ is actually what the listener's left and right ears will hear when exposed in the captured sound field.

2.1.2 Equatorial Microphone Array (EMA)

The EMA use the same spherical baffle as the SMA but only place microphones on the equator of the sphere [2, 6]. This setup cannot represent sound fields from arbitrary directions, for example, it cannot differentiate two sound fields that are mirrored on the horizontal plane. The method to compute SH coefficients in EMA is not the same as in SMA.The SH coefficients in EMA $\mathring{S}_m^{\text{surf}}(\pi/2, R, \omega)$ is calculated by integrating the sound pressure with a exponential along the equator as

$$\mathring{S}_m^{\text{surf}}(\pi/2, R, \omega) = \frac{1}{2\pi} \int_0^{2\pi} S^{\text{surf}}(\pi/2, \alpha, R, \omega) \mathrm{e}^{-\mathrm{i}m\alpha} \mathrm{d}\alpha \tag{2.8}$$

Here we call the $\mathring{S}_m^{\text{surf}}(\pi/2, R, \omega)$ as the circular harmonic (CH) coefficient of the sound pressure field on the surface S^{surf}. Similar to the SMA, the integral will be approximated by summing up the signal captured by the microphones located on the equator. And the reconstruction also has a maximum order m limit that requires m<N. The minimum required microphone number for Nth order EMA is 2N+1.

The sound pressure field on the surface S^{surf} can then be calculated by

$$S^{\text{surf}}(\pi/2, \alpha, R, \omega) = \sum_{m=-\infty}^{\infty} \mathring{S}_m^{\text{surf}}(\pi/2, R, \omega) \mathrm{e}^{\mathrm{i}m\alpha} \tag{2.9}$$

So now we need to find out how to calculate the SH coefficients of the incident sound field $\breve{S}_{n,m}(\omega)$ fo EMA. Firstly, we combine the eq 2.1 and eq 2.5, change the order of the summations to get

$$\begin{aligned} &S^{\text{surf}}(\pi/2, \alpha, R, \omega) \\ &\quad = \textstyle\sum_{m=-\infty}^{\infty} \sum_{n=|m|}^{\infty} \breve{S}_{n,m}(\omega) b_n \left(\omega \frac{R}{c}, R\right) Y_n^m(\pi/2, 0) \mathrm{e}^{\mathrm{i}m\alpha} \end{aligned} \tag{2.10}$$

Comparing eq 2.9 and eq 2.10 we can find the relationship between CH coefficients $\mathring{S}_m^{\text{surf}}(\pi/2, R, \omega)$ and SH coefficients $\breve{S}_{n,m}(\omega)$,

$$\mathring{S}_m^{\text{surf}}(R, \omega) = \sum_{n=|m|}^{\infty} \breve{S}_{n,m}(\omega) b_n \left(\omega \frac{R}{c}, R\right) Y_n^m(\pi/2, 0) \tag{2.11}$$

Since the EMA cannot recognize the height differential, a height invariant sound field is assumed. The SH coefficients of incident sound field for a plane wave with unit amplitude is given by [27]

$$\breve{S}_{n,m}(\omega) = 4\pi \mathrm{i}^{-n} Y_n^m(\pi/2, \theta)^* \tag{2.12}$$

insert eq 2.12 to eq 2.11 we get

$$\mathring{S}_m^{\text{surf}}(R, \omega) = \mathrm{e}^{-\mathrm{i}m\theta} \sum_{n=|m|}^{\infty} 4\pi \mathrm{i}^{-n} b_n \left(\omega \frac{R}{c}, R\right) \left[Y_n^m(\pi/2, 0)\right]^2 \tag{2.13}$$

Rabenstein et al. [43] put forward that any height invariant incident sound field can be present by a set of plane waves. Inspired by Rabenstein's method of using an infinite number of plane waves with a unique complex weight to represent a sound field, Arhens et al. [2] reformed eq 2.13 by employing a sum over an infinite number of plane waves, and each plane wave having a special complex amplitude X as

$$\begin{aligned} \mathring{S}_m^{\text{surf}}(R, \omega) = &\underbrace{\sum_{l=1}^{\infty} X_l(\omega) \mathrm{e}^{-\mathrm{i}m\theta_l}}_{= \mathring{\breve{S}}_m(\omega)} \\ &\times \sum_{n=|m|}^{\infty} 4\pi \mathrm{i}^{-n} b_n \left(\omega \frac{R}{c}, R\right) \left[Y_n^m(\pi/2, 0)\right]^2 \end{aligned} \tag{2.14}$$

The important unknown quantity $\mathring{\breve{S}}_m(\omega)$ is given by

$$\mathring{\breve{S}}_m(\omega)) = \frac{\mathring{S}_m^{\text{surf}}(R, \omega)}{\sum_{n'=|m|}^{\infty} 4\pi \mathrm{i}^{-n'} b_{n'} \left(\omega \frac{R}{c}, R\right) \left[Y_{n'}^m(\pi/2, 0)\right]^2} \tag{2.15}$$

Then, the desired SH coefficients of the incident sound field $\breve{S}_{n,m}(\omega)$ can be calculated. Compare 2.11 and 2.15 we can easily find that

$$\breve{S}_{n,m}(\omega) = \overset{\circ}{\overline{S}}_m\ (\omega) 4\pi \mathrm{i}^{-n} Y_n^m(\pi/2, 0) \tag{2.16}$$

from this equation, we can then easily compute the ear signals via eq 2.7.

2.1.3 The XMA

Briefly summarize the processing in SMA and EMA: to calculated the signal that arises at a given ear of the listener if she/he is exposed to the captured sound field, the key is to get the SH coefficients of the sound pressure on the surface by the linear combination of the microphone signal. The solution of XMA is inspired by the characteristic of SMA and EMA that each channel of the ambisonic signal is a linear combination of the microphone signals and the weights can be derived analytically [8]. However, it is not possible to get an analytical solution of the incident sound field by observing the sound field on the surface of an arbitrary scatter. The calculation of the SH coefficients $\mathring{S}_{n,m}$ by the linear combination of microphone signals S^{surf} is

$$\mathring{S}_{n,m}^{\mathrm{surf}}\ (R, \omega) = \sum_{q=1}^{Q} \chi_{n,m}^{(q)}(\omega) S^{\mathrm{surf}}\ (\vec{x}_q, \omega) \tag{2.17}$$

the χ is a frequency depend complex weight of the microphone's signals. We can assume that the weight χ contains the information of the array like the size and the shape of the scatter and the position of the microphones. Once the set of χ is conducted and the whole set of the microphone array is unchanged. It can be straightforwardly applied in eq 2.17 to the microphone signals captured in an arbitrarily incident sound field and then obtain their SH coefficients $\mathring{S}_{n,m}$.

But how to conduct the χ is a problem since in eq 2.17 we have two unknown but only one equation.The idea is to make one of the unknowns become a known quantity by the calibration measurement. To conduct the calibration measurement, the XMA should be exposed to sound fields of which the corresponding SH coefficients $\mathring{S}_{n,m}$ are known. For example, a plane wave propagating sound field. The SH coefficients $\mathring{S}_{n,m}^{\mathrm{surf,pw}}$ of plane wave sound field is given by

$$\mathring{S}_{n,m}^{\mathrm{surf,pw}}\ (\omega) = 4\pi \mathrm{i}^{-n} Y_{n,m}(\phi, \theta)^* b_n(R, \omega) \tag{2.18}$$

The baffle of XMAs does not need to be acoustically rigid or have analytically describable acoustic impedance. The unknown acoustic properties of the baffle are taken into account inherently during the calibration.

The χ can be obtained by a least-squares fit according to eq 2.17, which requires at least Q+1 different sound fields of microphone signals S^{surf} and known SH coefficients $\mathring{S}_{n,m}^{\mathrm{surf,pw}}$. As shown in Fig 2.1, in practical, the calibration measurement will take place in an anechoic chamber. If a human head sized XMA is positioned at a distance of more than 1m from the loudspeaker, the incident sound field from the

loudspeaker in the free field can be approximated by a plane wave. This measurement needs a sufficient amount of incidence directions.

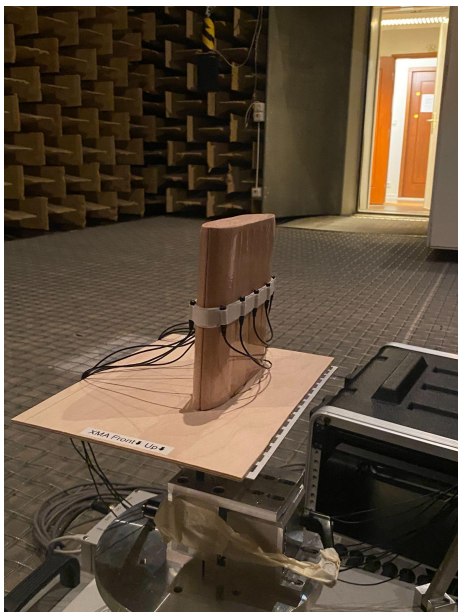


Figure 2.1: The calibration measurement of XMA in anechoic chamber.

Once we obtained the weight χ of the XMA, we can easily calculate the SH coefficients by eq 2.17 with the method of least squares-fit. And then apply it in the binaural rendering and get the signal in ear just as what we did in SMA and EMA.

2.2 Spatial aliasing

The spatial aliasing is an inevitable problem that will lead to audible impairment in the sound field recording and binaural signal synthesizing process. In this part we will briefly introduce it. As we all know, in digital signal processing,

As we all know, in digital signal processing, when we are sampling a set of continuous-time signals with a fixed sampling rate, the high frequency parts of the signal that above the Nyquist frequency cannot be deduced and reconstructed reliably. The signal part that is higher than the Nyquist frequency will be distorted and aliased to lower frequency components [41].

Similar to what people do in digital processing, the SMA or EMA method are both discretely capturing the real space continuous sound fields [34]. Analogous to sampling continuous signals, high spatial modes cannot be deduced reliably when we are sampling the real world space continuous sound field at discrete positions with a limited number of microphones. The high mode components will be mirrored into lower modes and result in spatial aliasing [34]. The captured sound information higher than the spatial aliasing frequency will be distorted and full of mistakes which we must neglect.

The spatial aliasing frequency f_A can be estimated by [39]

$$f_A = \frac{Nc}{2\pi r} \tag{2.19}$$

where the c is the speed of sound, and the N is the maximum resolvable SH order. The r means the radius of the rigid baffle. When lower than the spatial frequency, the spatial aliasing artifacts are of a very small magnitude that can be ignored. However, when above this frequency, it will increasing rapidly that makes this part full of error [34, 36].

3
Methods

In this paper, we used the boundary element method (BEM) which is implemented by mesh2hrtf from [52,53] to simulate eXMA calibration measurements for a number of candidate baffles. We thereby simulated the microphone signals due to sound originating from 101 point sources at different locations in the horizontal plane at a distance of 3m. The graphic of the point sources' positions is shown in Fig 3.1. The XMAs that we simulated will be settled at the center of the circle, and mounted with 17 microphones. And each time we will only simulate for one microphone and repeat this process 17 times to finish the whole test for one XMA. The 17 microphones will produce 8th ambisonic order except the GoPro MAX 360 shape which we only employed 10 microphones and produced 4th ambisonic order. The microphone number is comparatively high since binaural reproduction only needs a handful of microphones to produce a good result [2]. We put those high numbers of microphones for a good performance to make the interpretation of the data easier

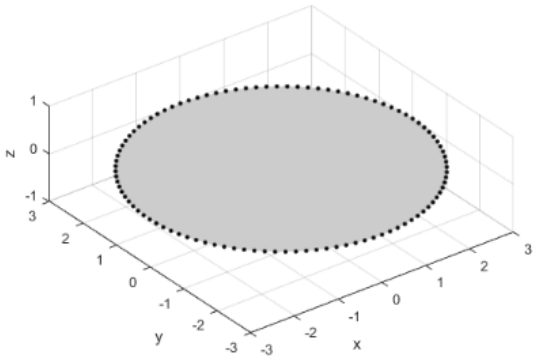


Figure 3.1: The 101 point sources at different locations in the horizontal plane at a distance of 3m.

To evaluate the performance of the XMA on our special shapes, we employed the normalized calibration error of the SH coefficients E(ω), which is defined as

$$E(\omega) = 20\log_{10}\frac{1}{L}\left|\sum_{l=1}^{L}\frac{\hat{\mathring{S}}_{n,m}^{(l)}(\omega) - \mathring{S}_{n,m}^{(l)}(\omega)}{\mathring{S}_{n,m}^{(l)}(\omega)}\right| \tag{3.1}$$

where $\mathring{S}$ are the correct known SH coefficients of the incident plane sound field. $\hat{\mathring{S}}$ are the SH coefficients that are computed by multiplying the XMA filter $X_{n,m}$ and the microphone signal. The $X_{n,m}$ were calculated by the least-squares fit from

eq 2.17. l is the index of a total of L horizontally propagating plane waves for which calibration data are available. We use L = 101 throughout this paper as mentioned before. The normalized error E(ω) is actually a comparison between the real correct coefficients and the SH coefficients calculated by our method. It is a very important parameter that can directly reveal the performance of the eXMA method when the scatter is non-spherical.

The top plot in Fig 3.2 shows the magnitude of the transfer function of the filters $X_{n,m}$. This filter converts the microphone signals of the eXMA into an SH representation. The bottom plot in Fig 3.2 depicts the normalized calibration error of the SH coefficients E(ω).The data are from previous study [2]. In the next part, we will mainly use the plots like Fig 3.2 bottom to show how we will demonstrate our results and analysis.

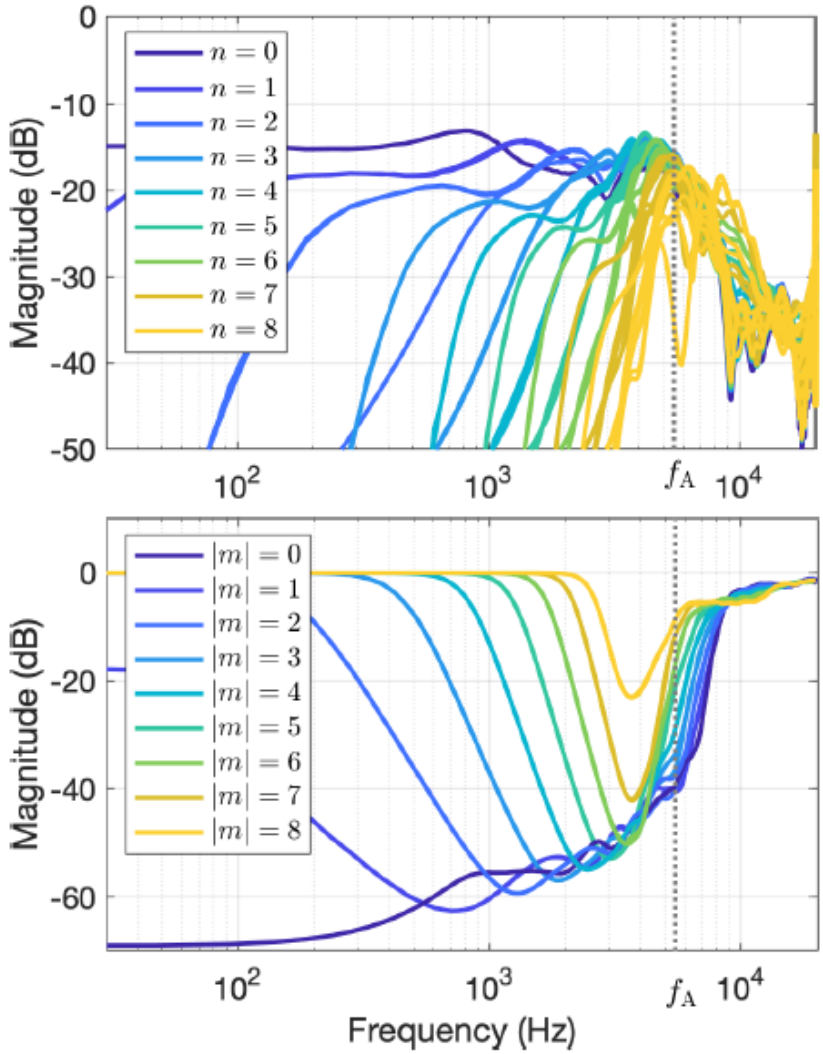


Figure 3.2: Top: $20 \log_{10} |X_{n,m}|$ for a selected microphone of the eXMA depicted in Fig 1.13 right. Bottom: Normalized calibration error E(ω) of that same eXMA [8].

The Fig 2.4 we used to demonstrate the result actually includes several important

principles that are valid in all simulation results that we will show in the result section:

i) In Fig 3.2 top, each color has a number of lines that are very close to each other. Actually, a given color line can mask all lines with the same color that depict the data for all azimuthal modes m that correspond to the indicated order n. This is very obviously in the lower nth order. Similarly, each line in Fig 3.2 bottom masks the lines for all other orders n that correspond to the indicated azimuthal mode m.

ii) At the left side of Fig 3.2, the magnitude of error is denoted in decibel. An error in 0 dB means the magnitude of the error is at the same level of the data itself. And all the calculated results with such a big error should be ignored, the corresponding SH coefficients are corrupted. We proposed to set -40 dB as the threshold of the error to determine whether the corresponding SH coefficient is reliable or not. An error in -40 dB means the magnitude of error is about 0.01 of the data itself, which is a quite conservative choice.

iii) In Fig 3.2 bottom, each mode m has a frequency range with a quite low magnitude of error in which it can be extracted with high accuracy. However, towards to lower frequency, the finite aperture will limit the accuracy of the result. At the lower frequency range, the wavelength of the sound becomes much longer than the aperture, at that point, the microphone array cannot extract the spatial information as well as before. This phenomenon also happens when we shrunk the size of the baffle and thus produced a smaller aperture array. At very low frequency, only the error of the 0th mode remains at low altitude. And thus only the omnidirectional information is useful as we mentioned before. It is impossible for us to identify the sound source direction with the omnidirectional information.

ix) For higher m mode, the error at the low-frequency end of the sweet point is bigger than the lower m mode. However, since those lower m modes have provided us enough spatial information to make us "feel the ambisonic", the lack of higher modes information will not result in a very severe impairment [34]. Another interesting and important phenomena is that for most modes at low frequency range in Fig 3.2 bottom, we found the normalized error $\mathrm{E}(\omega)$ is very high and even close to 0 dB. However, the big error did not cause a serious consequence. When we look to the Fig 3.2 top, we found that the corresponding filters $X_{n,m}$ at the low frequency range actually attenuate the microphone signals. The magnitude of the filters in this high normalized error frequency range are very small. This indicates that the mode might cannot be extracted from the microphone signals at this frequency range, and thus the information is missed in the ambisonic representation. In other words, this high error part may not result from the corruption of the array but from the missing information. In this low frequency range, we lost the higher m mode's information but what we have here is all correct. It is more preferable compared to obtaining corrupted modes at this range because at least what we got in the process is correct, we are not getting the wrong message but lost some spatial information.

x) In the range of above the spatial aliasing frequency f_A, the high magnitude of error denotes that most of the spatial information is not correct even though the SH coefficient exhibits energy. This high error frequency range part due to the spatial aliasing will be ignored. The information from higher modal orders appears into lower modal orders in the spatial aliasing range and result in spatial ambiguities [34]. However, contrary to intuition, this does not cause a significant perceptual impairment if at all [34]. Actually, the spatial aliasing is not all of harm. Compared with not have spatial aliasing at all, it is more beneficial to have them higher order available since they can help reduce unwanted angle dependencies in the reproduced signals [8].

4

Results

The data we present in this part are all based on the array response to sound incidence from horizontal directions. We did not simulate the response from a non-horizontally incidented sound field.

We found that the results of XMA with different shapes of baffles are very similar to EMAs and head-mounted XMAs [2] qualitatively and quantitatively. The XMA actually outputs a horizontal projection of the captured sound field [8]. When in the frequency range that is below the spatial frequency of the array, the output with the employed binaural rendering can be very close to the right answer: only a few dB deviated from the original correct one. When the frequency is above the spatial aliasing frequency of the array, the deviation will become larger, but the perceptual hearing impairment is not significant as mentioned before [34].

We considered and simulated several shapes of baffles who are typical in SMA and EMA and transformed them to find the potential of XMAs. We also simulated the shape of commercially used VR cameras to see the possibility of the application in VR recording in the future. We will discuss some important and interesting observations on a conceptual level so that we can apply the concepts to some other shape in future investigation.

4.1 No Baffle

Not using any baffle at the center of the microphone array is actually not an option in this thesis and we did not simulate this situation. However, an open microphone array in SMAs has been investigated before: in this case, the open array will meet numerical ill conditions at frequencies which correspond to the nodal values of the spatial spherical modes [11]. In other words, the certain SH modes can not be extracted at certain frequencies. At these frequencies, ambiguities arise and result in excessive noise and we can observe several peaks in the lower order of the $E(\omega)$ for no-baffle array.

Some methods have been proposed to mitigate the ambiguities, for example, using cardioid microphone arrays or applying two or more layers of arrays for no baffle condition by switching between the layers depending on frequency and mode [11].

Actually, these kinds of open microphone arrays do have some advantages [11]. At

low frequencies, a large solid sphere is needed for good performance. The realization and handling of large solid spheres might not be practical and thus the open microphone array could be preferable to rigid-sphere arrays. In addition, the spherical baffle might reflect back sound into the measurement region, therefore modifying the measured sound field. However, this type of microphone array never reached widespread use [8].

4.2 The height of the Baffles

We use a set of spherical baffles and vertically stand cylinders to illustrate the effect of the height of the baffles on the accuracy of the extracted SH modes. The number of the microphone is 17 to produce the 8th ambisonic order reproduction. All microphones are settled at the middle of the baffle. The shapes of baffles and the microphone positions are illustrated in Fig 4.1.

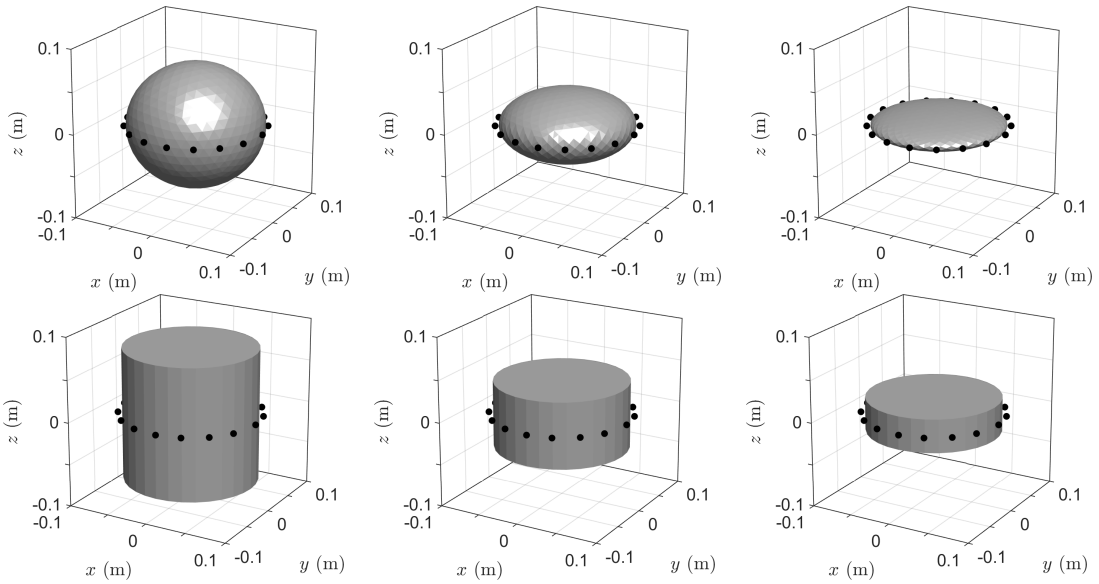


Figure 4.1: The shapes and microphone positions are of different heights. The top row from left to right: i) Sphere with radius r=78mm but squashed in the z direction, the height h=1.8r. ii) Sphere with radius r=78mm, h=0.9r. iii) Sphere with radius r=78mm, h=0.2r. The bottom row from left to right: i) Cylinder with radius r=78mm, h=2r. ii) Cylinder with radius r=78mm, but squashed at the z direction, h=r. iii) Cylinder with radius r=78mm, h=0.5r.

The results of each shapes are shown below. Fig 4.2 shows the E(ω) for the spheres that have a height h =1.8r, h=0.9r and h=0.2r. Fig 4.3 shows the E(ω) for the cylinders that have a height h=2r, h=r and h=0.5r. The radius r=78mm for all of them, which is close to the radius of a human's head.

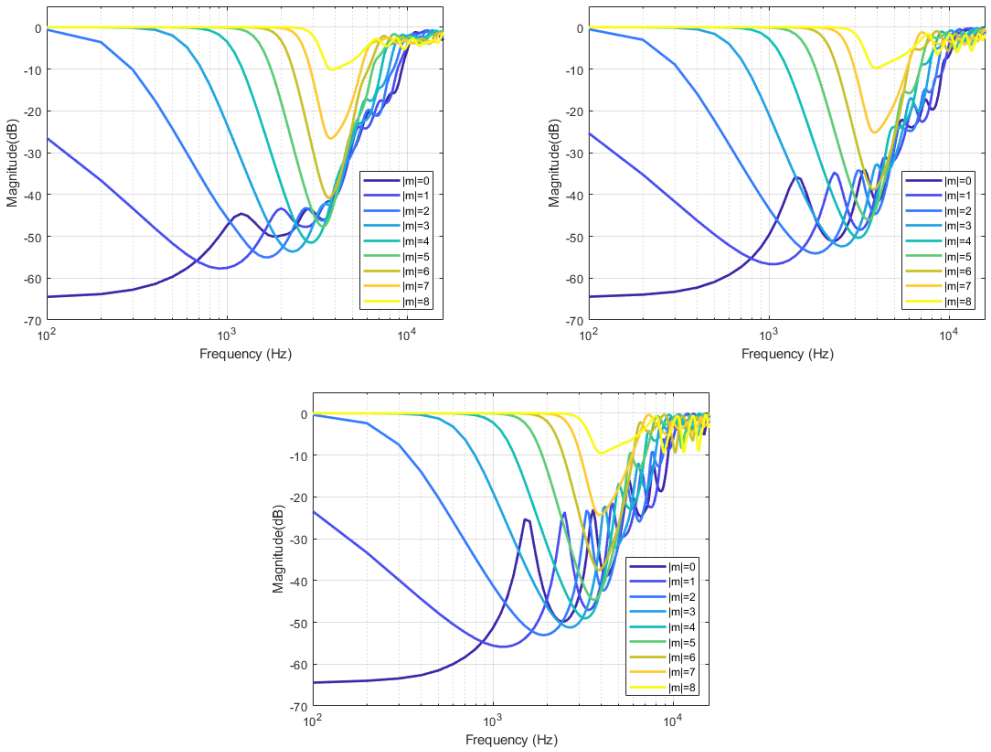


Figure 4.2: E(ω) of a Sphere with radius r = 78mm and different height h, cf. Fig 4.1 top row). Top Left: h = 1.8r. Top Right: h = 0.9r. Bottom:h=0.2r.

Comparing these two sets of results, we found that the accuracy that is provided by a sphere baffle in XMA method is very close to the results of the cylinder; the advantage is not significant. For both sphere and cylinder shape of baffles, lower the height of the baffle will result in more and higher error peaks in different mth order at different frequencies. As we mentioned in section 4.1, the high error peaks in lower order are very common to be observed in open arrays. This lower height of the baffle approach actually makes the XMA's properties more close to an open array.

Both sphere and cylinder baffles with the height of about half radius have more pronounced open-array-like peaks in the error compared to sphere and cylinder with height close to radius.

If we longer the cylinder, the accuracy of the XMA method will become better and lower down the error, but this advantage will not be provided above a certain height. Fig 4.4 shows the E(ω) for the cylinder with h=3r and h=4r. We can see that the error peak in the 0th mode around 1 kHz for h=2r and h=r which is quite obvious in Fig 4.3 now disappeared, but the difference between h=3r and h=4r is not very significant.

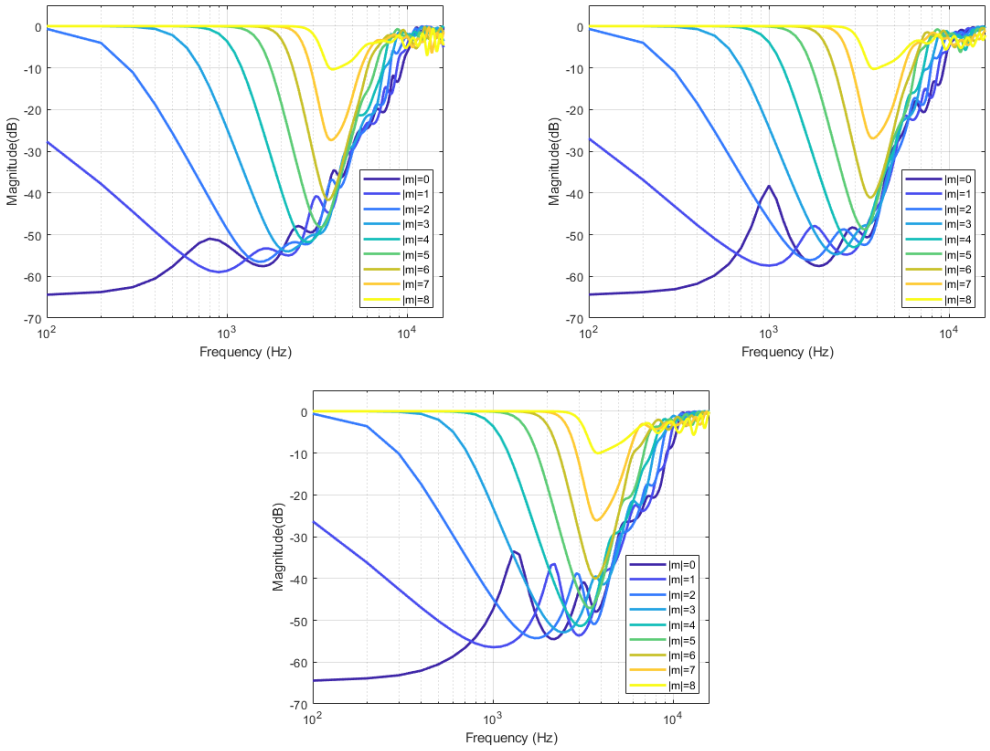


Figure 4.3: E(ω) of a cylinder with radius r = 78mm and different height h, cf. Fig 4.2 bottom row the middle and the left).Top Left: h = 2r. Top Right: h = r. Bottom:h=0.5r.

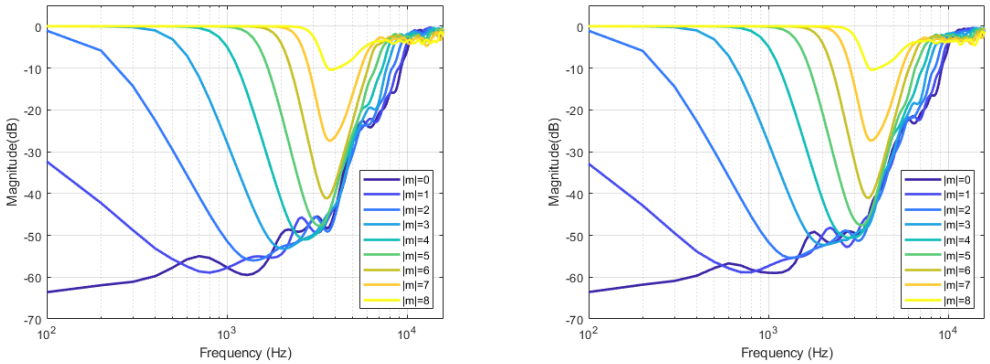


Figure 4.4: E(ω) of a cylinder with radius r = 78mm and different height h. Left: h = 3r. Right: h=4r.

Figure 4.5: Photograph of a Vuze 360 camera [50] (left) and a Live Planet 360 camera [33].

The shape of squashed sphere and cylinder are similar to commercial VR cameras Vuze 360 [50] and Live Planet [33], but the size is about 30% bigger than the cameras. The shape of these two commercial VR cameras are shown in Fig 4.5 The performance of the Live Planet 360 camera is similar to the result of the cylinder baffle with height h=r but without the hump in the 0th order at 1 kHz [8]. These results indicate the XMA's potential to produce ambisonic sound reproduction when integrated with VR cameras who have an appropriate shape and size.

4.3 Shape of the Cross-Section

Since the circular can be considered as a polygon with an infinite number of sides, and in the software Blender, it is actually a polygon with a lot of sides, (typically with 36 or 48 sides) to approximate a circular. It is very intuitive for us to see what will happen when we change the shape of the cross-section of the cylinder baffle from circular to octagon, hexagon, square and even triangular.

Fig 4.6 shows three baffles with different shapes of cross-section from square to triangular and triangular with smoothed corners. The microphone arrays are located at the middle of the baffle. The number of the microphone is 17 to extract up to 8th order of SH coefficients. The radius of the smoothed corners is 1mm. They are all derived from the cylinder with the radius r=78mm and height h=2r. We only changed the number of the sides in Blender so the diagonal length of each cross section is still the same as the circle's diameter d=2r. We omit the plots and results of octagon and hexagon shaped cross sections. And the E(ω) of these two shapes of cross section are not surprisingly very close to the result of the original cylinder.

Fig 4.7 shows the normalized E(ω) result of the square section and the triangular section with sharp corners. To our surprise, the magnitude of error does not become bigger but even smaller compared with the same height cylinder. The peak in 0th m mode even lower down. Rounding the triangular cross-section to vanish the corner and make the shape smoother as shown in Fig 4.6 bottom does have a benefit but not significant.

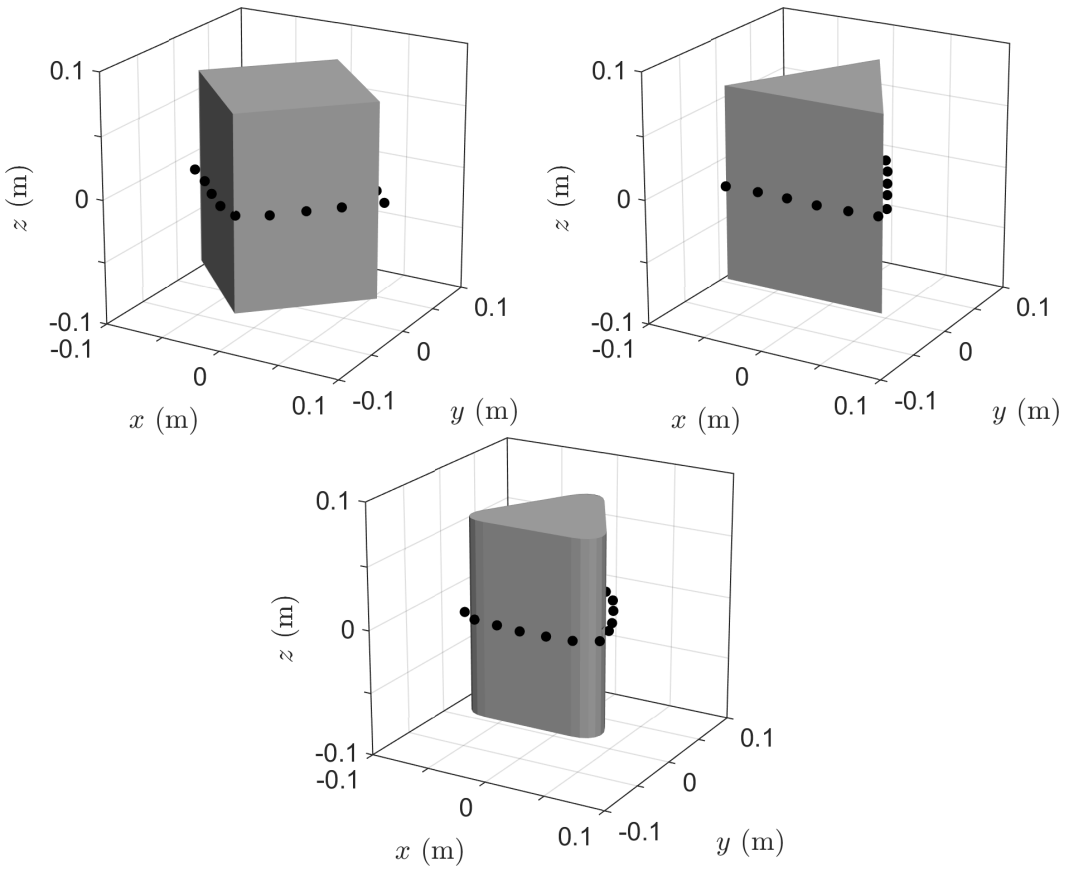


Figure 4.6: The shapes and microphone positions of different cross sections. From left to right: i) Square section with diagonal length l=2r=156mm the height h=2r. ii) Triangular section with the same diagonal length and height. iii) Triangular section with the same diagonal length, height and smoothed corner.

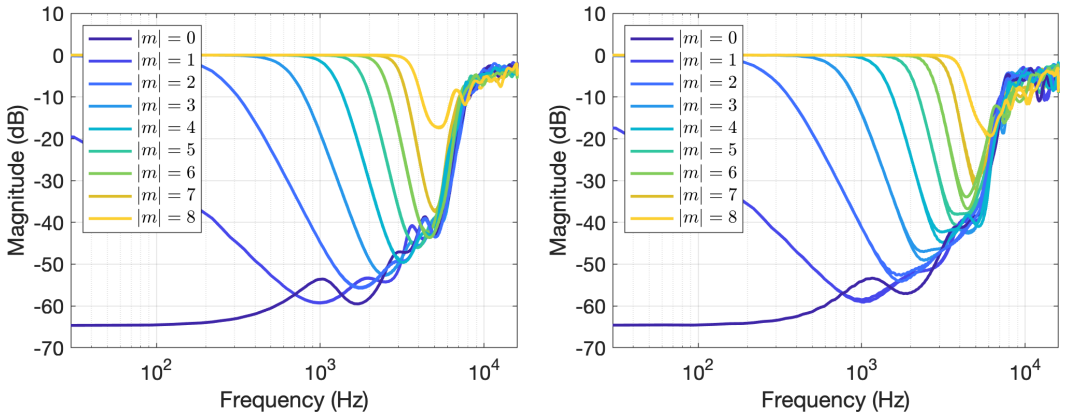


Figure 4.7: E(ω) of different shapes of cross sections. Left: Square section. Right: Triangular section.

Changing the cross-section of the baffles does not influence the accuracy of the XMA method significantly. The shape of the cross section is not critical and even the corners are possible. The relatively low magnitude of error $E(\omega)$ of these different shapes of cross-section denotes the good adaptability of the XMA method for different shapes of baffles. And it seems that the height of the baffle has more influence on the accuracy of the SH extraction when compared with the result in section 4.2.

4.4 Position of the Microphone Array

In all examples we presented before, the microphone array was mounted exactly in the middle of the baffle in terms of the height of the baffle. The distance from the microphones to the both sides of the baffle is equal. However, it is not practical to always mount the array in the middle of the baffle in daily use. For example, some VR cameras may have put the lens at the middle of the camera body and if we still mounting the microphone array at the middle, the version of the camera will be blocked.

In the previous study, the head mounted XMA showed good performance and the array was also not mounted at the middle of a human head but at the forehead. But how the position of the microphone array will influence the accuracy with which given SH modes can be extracted are not cleared. In this section, we chose the cylinder baffle that has shown before with the height h=2r and moved the microphone array from the middle to the top edge of the baffle. Fig 4.8 shows three different positions of the microphone array. The distance from the upper edge is 40mm, 20mm and 3mm, the last one can be considered as at the edge. The number of the microphone is still 17 to extract up to 8th order of SH coefficients. Since the shape of the baffles are symmetry, there is no difference between moving the microphone array up or down.

The reason for choosing a cylinder rather than a sphere is that when moving up the microphone array mounted on the cylinder, the radius of the array will not be shrunken and thus the aperture of the array is fixed. For a spherical baffle, the microphone array's aperture will become smaller: the radius of the cross section is smaller and the microphone array needs to be placed close to the surface of the baffle. The radius of the microphone array or the aperture of the array is actually a very important parameter that influences the performance of the XMA and we will introduce this in the next section.

We naturally predict that moving up the microphone array may cause a much worse error. When the array is closer to the edge, the property of the system might be close to an open array and thus increase the error. However, the result is quite different from what we predict. Fig 4.9 shows the $E(\omega)$ of the cylinder baffle with three different microphone array's positions that had been shown in Fig 4.8. Compared with the $E(\omega)$ of the cylinder that the microphone array was positioned at the

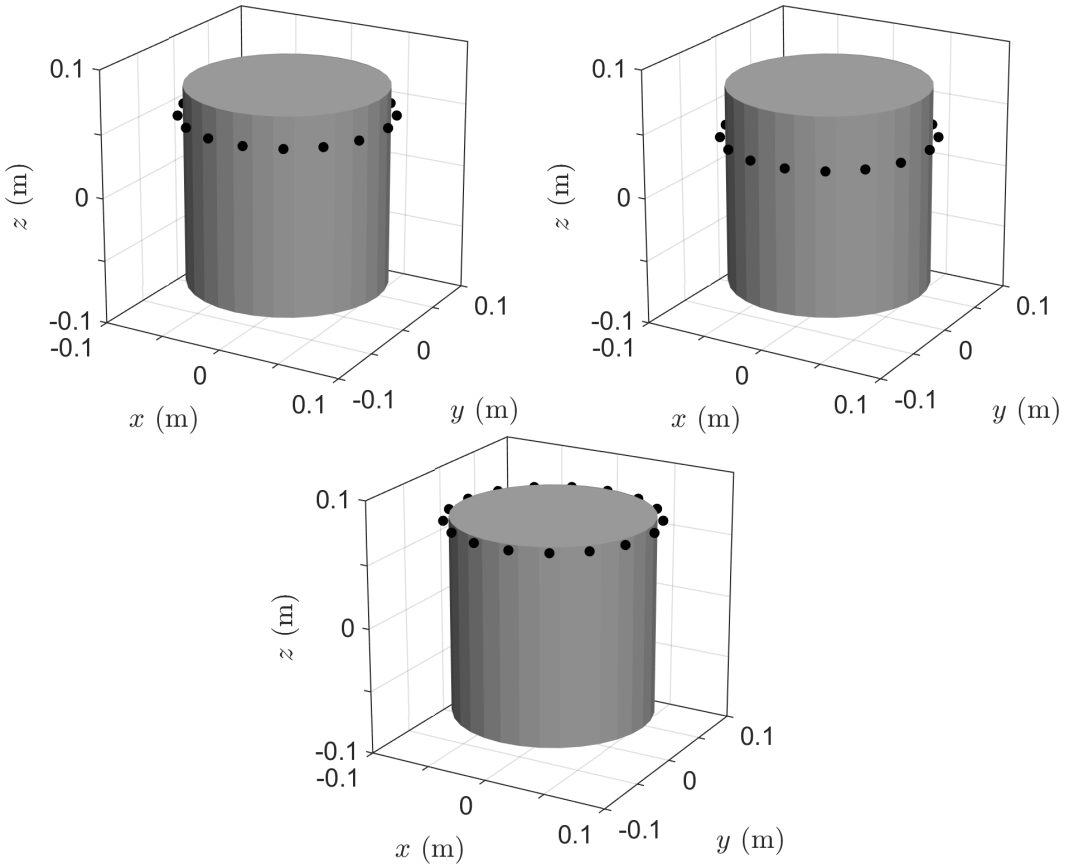


Figure 4.8: The shapes and move-up microphone positions of the cylinder baffles. The radius of the cylinder r=78mm, height h=2r. Top Left: the microphone array is 40mm below the top edge of the cylinder baffle. Top Right: the microphone array is 20mm below the top edge of the cylinder baffle. Bottom: the microphone array is positioned at the upper edge of the cylinder baffle.

middle in Fig 4.3 top left, moving up the microphone array does not change the result significantly. For both the 40mm and 20mm below the upper edge, the error amplitude for lower order is still smaller than -40 dB and the peak's value is even lower than at the middle. It seems that moving up the array a little will not lead to lose much spatial information.

However, when we move up the microphone array very close to the upper edge that is only 3mm below the upper edge, the phenomenon that we predicted before comes out. Fig 4.9 bottom shows the E(ω) of the cylinder that the microphone array is very close to the upper edge. The error amplitude of lower orders raised up and more peaks appeared that is similar to what we will see in an open array. However, the amplitude of error is still moderate and not higher than -30 dB, which means the error is smaller than 3% of the data itself. It is likely that a cylinder baffle with microphone array positioned at the upper edge can still produce a useful ambisonic sound field.

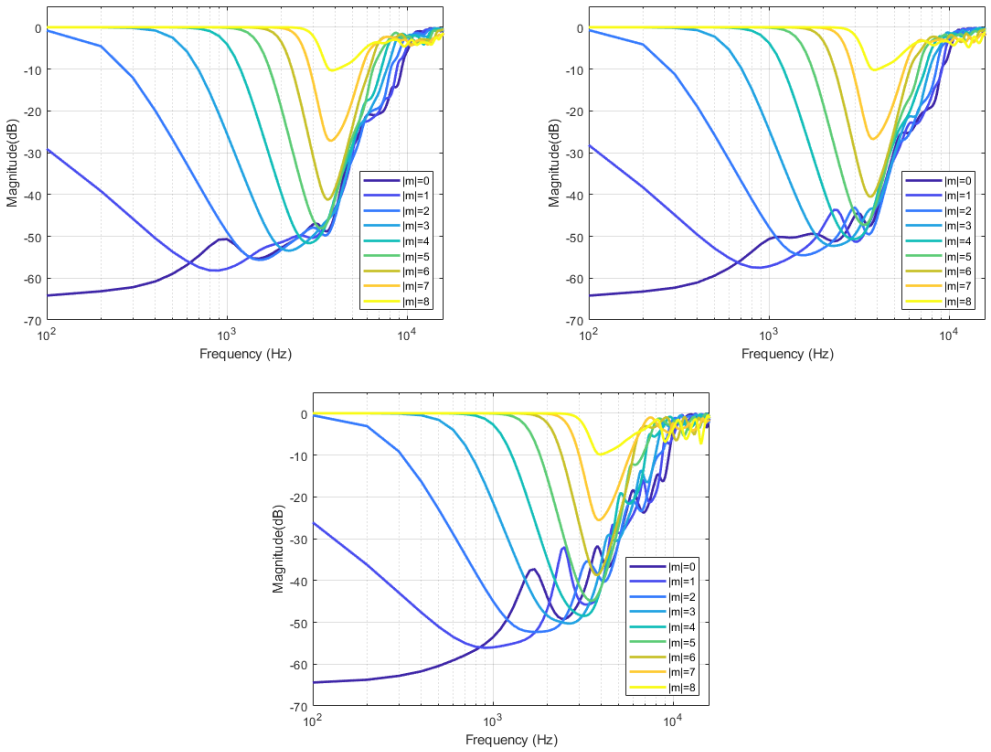


Figure 4.9: The E(ω) of cylinder baffles for different positions of microphone arrays. Top Left: the microphone array is 40mm below the top edge of the cylinder baffle. Top Right: the microphone array is 20mm below the top edge of the cylinder baffle. Bottom: the microphone array is positioned at the upper edge of the cylinder baffle.

Based on the results, the position of the microphone array is not a very critical factor. We can safely place the microphone array on different positions of the baffle rather than make it in the middle. This character makes the XMA method more practical in VR recording so that we can mount the array flexibly on the camera. However, placing it at the edge is not recommended and may cause a little worse accuracy.

4.5 Squash the sideways of the shape

In section 4.2, we have shown that squashing the baffle in the longitudinal way will make the E(ω) similar to an open array. But we are not clear about what will happen if we squash the baffle sideways. In this section, we simulated two cylinder baffles with the same height as the cylinder in Fig 4.1 down left and squashed it in the radial direction. Fig 4.10 shows the two radial squashed cylinder baffles. The number of the microphone is 17 to extract up to 8th order of SH coefficients and the array is at the middle of the baffle.

The squashed cylinder that the thickness is 80% of the height produced an error E(ω)

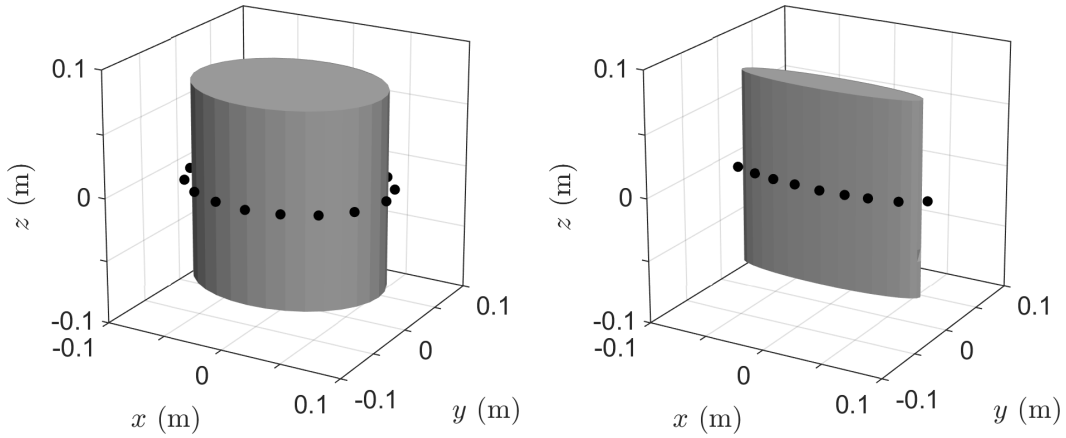


Figure 4.10: The shapes and microphone positions of the cylinder that squashed sideways. The original radius of the cylinder r=78mm, height h=2r. Left: the thickest part of the squashed cylinder is 80% of the height. Right: the thickest part of the squashed cylinder is 20% of the height.

that is very similar to the original cylinder. We therefore omit to present the data here. Fig 4.11 shows the error E() of the squashed cylinder that the thickest part is about 20% of the height. Even though the aperture of the array on the extremely squashed cylinder is smaller than the original cylinder, the accuracy of this shape is still quite good. The error magnitude of 0th and 1st order is lower than -40 dB in a broad frequency range just as the original cylinder. This means in this lower frequency range, rather than only have the omnidirectional information presented by the 0th order, we can still have some other spatial information available. It is still possible for human beings to get the ambisonic feeling.

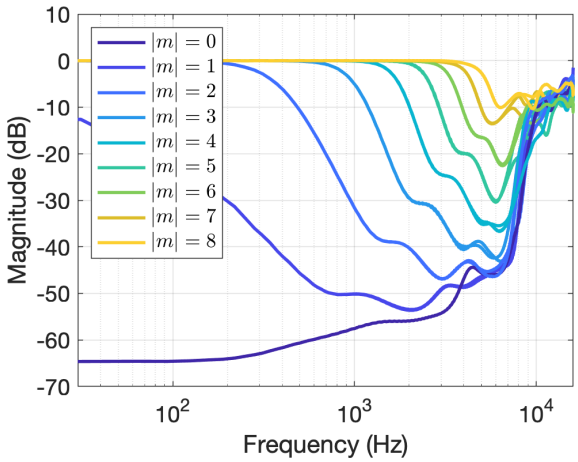


Figure 4.11: The E(ω) of cylinder baffles is extremely squashed, the thickest part of the squashed cylinder is 20% of the height.

However, the Fig 4.11 shows a new interesting phenomenon: except the 0th order, all the other lines have a steady platform that makes the error remain higher than the acceptable error threshold for a longer frequency range. Take the 2nd order as

an example, the platform shows up at about the -40 dB and remains for about 500 Hz range, which makes the valid range of the 2nd mode shorter.

If we look back to the other shapes we have shown before, we can see that the error of the triangular cross section baffle in Fig 4.7 right also has a similar platform. This negative effect may be caused by the shrinkage of the array's aperture since both of the two baffles are significantly smaller than others in dimension. The short spatial dimension in the sideway squashed cylinder increased the E(ω) moderately but still acceptable. It seems that the aperture of the array is long enough in one of the Cartesian dimensions to maintain an acceptable accuracy. It was also shown in some research that even a thin plate can be sufficient of a baffle, but not with equatorial microphone layouts [12].

4.6 Special Shapes

In this section we employed two kinds of special shapes to further identify the characters and limitations that the XMA method may arise. The first set of shapes is a combination of the shapes we simulated before: a cylinder with a dome on the top that looks like a mushroom. The second shape is a small cuboid-like baffle that is inspired by the GoPro Max commercial VR camera [26].

4.6.1 The "mushroom" Baffles

Fig 4.12 shows the special "mushroom" baffle which is generated by connecting a cylinder and a dome.The cylinder and dome in the original "mushroom" on the Fig 4.12 top left has the same size as the cylinder and sphere we simulated before. The "mushroom" on the top right has a smaller cylinder that the radius of it is 80% of the original one. The "mushroom" on the bottom has a smaller cylinder that the radius of it is 50% of the original one. The 17 microphones are positioned 3mm higher than the joints of these two shapes.

Fig 4.13 shows the normalized error E(ω) of the two "mushrooms".The original "mushroom" as shown in Fig 4.12 top left produces an almost same error E(ω) as a regular cylinder and we omit to present the data here. However, the normalized error E(ω) of the dome with a smaller cylinder increases obviously, as shown in Fig 4.13 left. The error hump looks very similar to the open array again. This is intuitive since the microphone array is located very close to the edge of the two shapes even though the downside of the dome is not really an open area which still has quite a big baffle.

Combining these two results with the result in section 4.4 of moving the microphone array to the top edge of the cylinder, all of them indicate the negative effect of the edge on the XMA method. Although the increased error is relatively moderate and the accuracy is acceptable, setting the microphone array as far away from the edge as possible is still necessary.

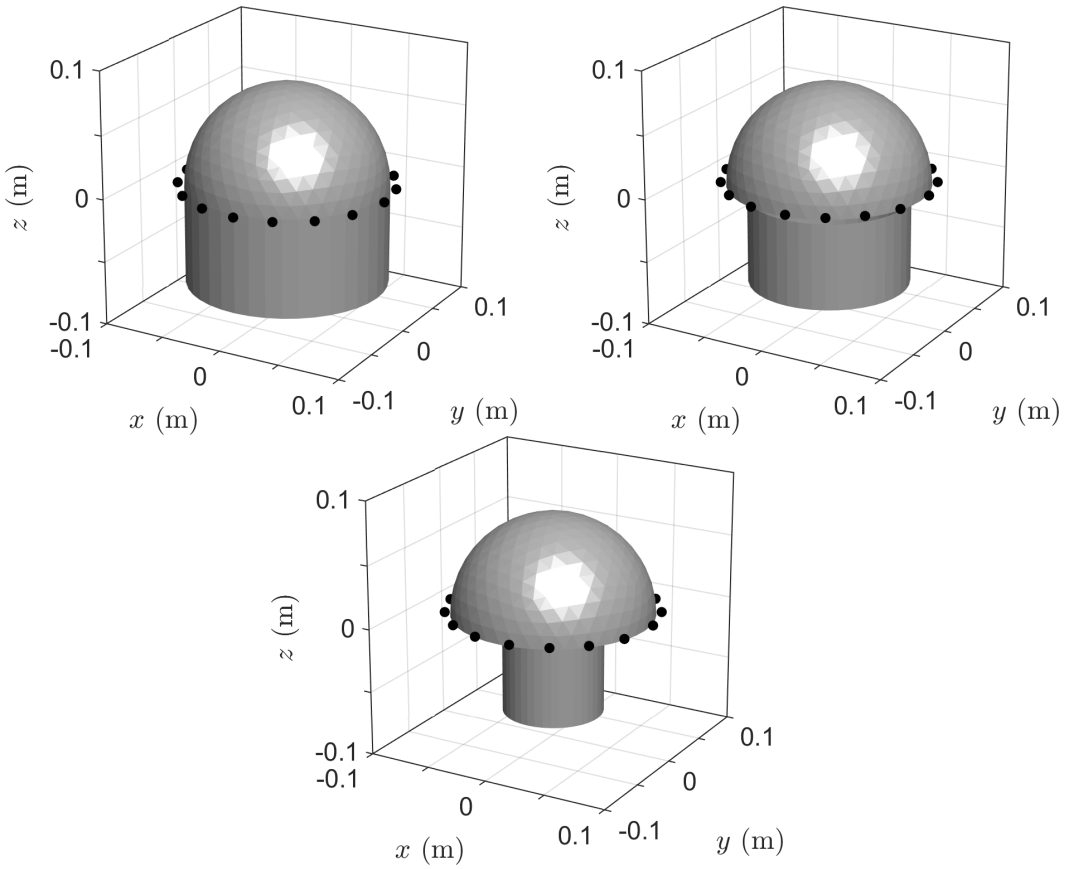


Figure 4.12: The shapes and microphone positions of the "mushroom". The original radius of the cylinder and the dome r=78mm, height of the cylinder and dome are both h=r. Top Left: the cylinder and the dome have the same radius and height. Top Right: the cylinder and the dome have the same height, but the radius of the cylinder is 80% of the dome. Bottom: the cylinder and the dome have the same height, but the radius of the cylinder is 50% of the dome.

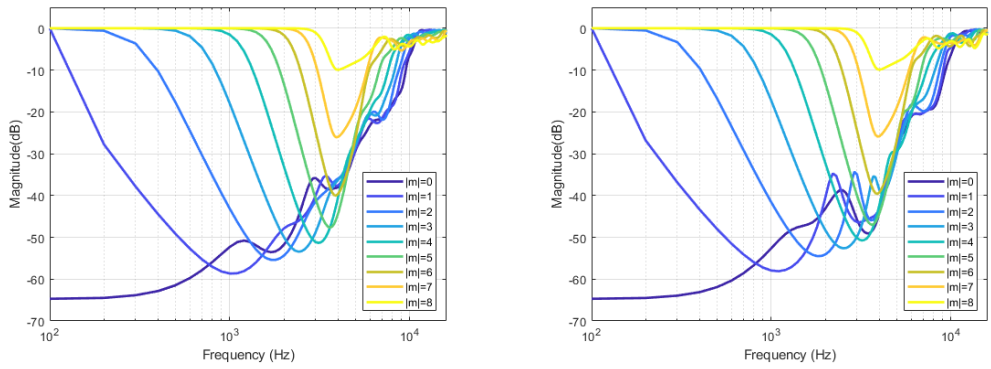


Figure 4.13: The E(ω) of the "mushroom" baffles. Left: the radius of the cylinder is 80% of the dome. Right: the radius of the cylinder is 50% of the dome.

4.6.2 The "GoPro Max" Baffles

This cuboid-like baffle is inspired by the GoPro MAX commercial 360 camera [26], both of the model and the GoPro are shown in Fig 4.14. The size of the baffle is 64mm in length,16mm in width and 69mm in high. The hump on the baffle represents the lens of the camera.

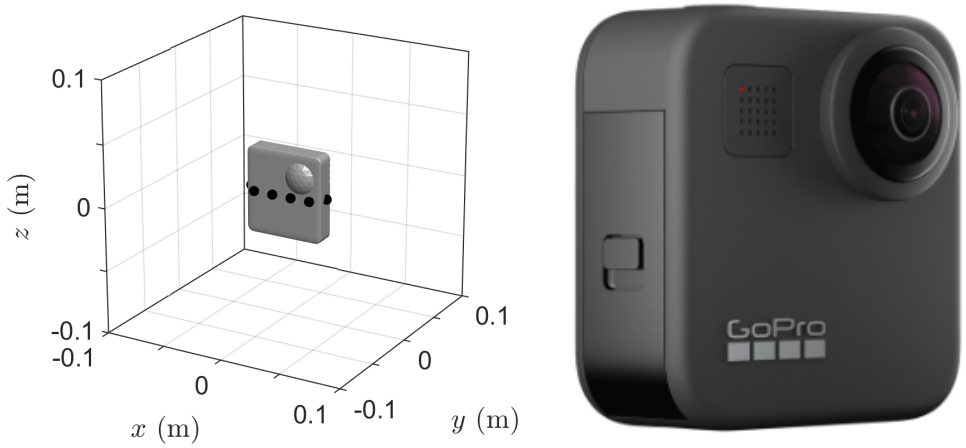


Figure 4.14: Left: the shape and microphone positions of the "GoPro Max" inspired baffle. Right: the "GoPro Max" 360 camera [26].

The microphone array is not settled in the middle but a little bit down to keep a safe distance from the lens. We also think that this distance can mitigate the negative effect of the joint edge between the lens and the body of the camera. Compared with all the baffles we simulated before, this cuboid-like one is significantly smaller in dimension. Putting 17 microphones on this much smaller body is too crowded and not practical, we employ a smaller number of microphones that only 10 microphones are placed on to extract up to 4th order of SH coefficients.

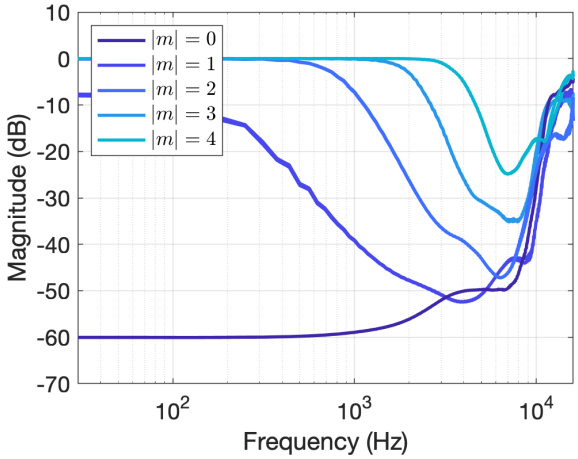


Figure 4.15: The E(ω) of the"GoPro Max" inspired baffle.

Fig 4.15 shows the error E(ω) of the cuboid-like baffle which is similar to the result of the extremely squashed cylinder in section 4.5. Because of the smaller size, the

spatial aliasing frequency is much higher than other baffles we simulated before. We can see the steady platform in the 2nd and 3rd mode clearly. The error of 0th mode is acceptable with a very high accuracy at very low frequencies. It seems that one dimension of this baffle is sufficiently large to maintain an acceptable performance just like the extremely sideway squashed cylinder.

But the error of the 1st mode is higher than the -40 dB threshold until 1 kHz, which means this mode will not be valid until 1 kHz and above. And the other higher order modes will be available only in a much higher frequency range. This actually means any sorts of spatial information is only available above 1 kHz. In other words, under the 1kHz range, we can only hear an omni-directional sound source without any directional information. With the frequency raised, the sound then becomes directional. We are not sure how it will influence our perception and this topic might be evaluated in future work.

4.7 Recording for two shapes

To demonstrate the good performance of our XMA on those non-spherical baffles, we made two recordings based on Fig 4.6 top left, the square cross section baffle and Fig 4.10 right, the cylinder which is squashed sideways. The recording setting of the two shapes are shown in Fig 4.16. These two baffles are made from wood by the author's handcraft. Due to the limitation of size, we only mounted 12 microphones to the rectangular one to get 5th order decomposition, and the squashed cylinder was only mounted with 10 microphones for 4th order decomposition.

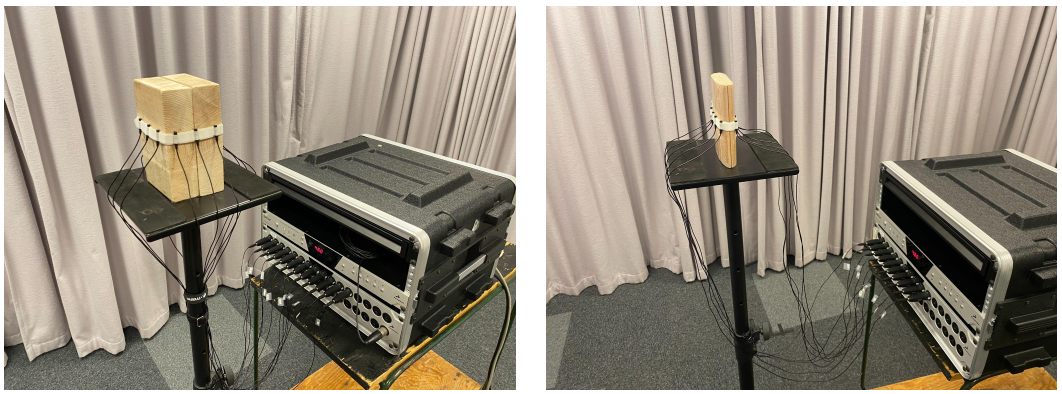


Figure 4.16: The recording setting of the square cross section baffle and the squashed cylinder

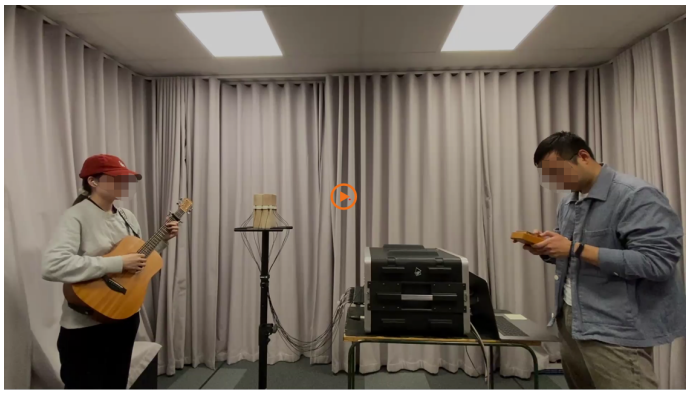


Figure 4.17: The video of recording scene.

The two recordings were connected with two videos as shown in Fig 4.17. The video shows two students playing guitar and kalimba on both sides of the baffle, and then the guitar player walking around the baffle. The XMA shows a good performance with these two non-spherical baffles that with the headsets, we can hear two different instruments' sounds coming from your left and right sides separately. And then the sound of the guitar starts to move around you. However, according to some of our listeners' feedback, the recording produced by the rectangular baffle is better than the squashed cylinder produced. We believe the difference is caused by the fact that the XMA with the squashed cylinder baffle only provides 4th order of decomposition but the rectangular one provides 5th. And the simulation results also show that the rectangular baffle can provide more spatial information in the lower frequency range.

5

Conclusion

Derived and improved from the SMA and EMA method, we investigated the XMA method and applied a set of non-spherical baffles on it to identify what spherical harmonic orders can be obtained with what accuracy.

Based on these simulations, we analyzed the normalized calibration error E() of the high order microphone arrays with non-spherical baffles and attempted to find the relationship between the accuracy and the baffle shape. The calibration error represents the error between the spherical harmonic (SH) coefficients of the known sound field and of the captured sound field that are extracted from the microphone signals. It terms out that reducing the height of baffles, squashing the baffles sideways, moving the microphone array to the edge of the baffle will introduce ambiguities and increase the magnitude of the normalized error E(ω). Those practices will make the system show the property that is very similar in an open array without any baffles.

However, the increased amplitude of normalized calibration error E(ω) is moderate. The error of majority non-spherical baffles we simulated are lower than -40 dB in low order modes that can have a good performance in ambisonic sound field representation. Some reshaped baffles have an error that is higher than the -40 dB threshold but still lower than -30 dB, we believe they can still produce a good performance since the error is smaller than 3% of the data itself. The shape of the baffle is not very critical. A rounded surface is not even necessary since the error of the square and triangular cross section baffle with sharp corners are still low enough.

The XMA method shows a very exciting performance with the non-spherical baffles. It provides a much more convenient way in ambisonic sound field recording compared with the ordinary SMA method. And the price we need to pay for the convenience in losing accuracy seems acceptable and limited.

The frequency band limitation of which SH coefficients can be extracted reliably is possibly determined by the dimension of the baffle. The size of the baffle has a significant influence on the bandwidth. The upper frequency limit of the bandwidth is determined by the spatial aliasing which is very similar to how it is in the situation of spherical microphone array in SMA method. This spatial aliasing frequency is both determined by the highest order of SH coefficients you need to extract and the radius of the baffle. The frequency raised up higher with a smaller radius. The lower limit of the frequency range that has a very small error is determined by the aperture of the array. There is no need to have a big enough aperture in all dimensions.

According to our simulation, it is sufficient to get an accurate result if the aperture of the array is long enough only in one Cartesian dimension. However, the shortage in one of the dimensions will still introduce some error. This shortage makes the high order of modes only valid in high frequency range and at very low frequency range, only the 0th order is available.

Higher order cannot be extracted at low frequency with a small aperture of array. This is a physically limited problem since the wavelength of a low frequency sound wave can be bigger than the dimension of the microphone array and all the microphones may capture the same signal no matter what direction the sound wave comes from. Thus we can hardly deduce any spatial information from the captured signals.

However, we are still not sure whether the small baffle that the 1st order modes only valid above 1 kHz still produce satisfactory perceptual results when the captured signals are audible. And at this point, it is also unclear how small such a baffle can be while still performing well in people's auditory perception. In the previous study, for a bowling ball size array that delivers 5th and higher order in the mid frequency range, people can hardly realize the loss of spatial information at low frequency range when binaural playback is employed. The undesired missing of information is inaudible.

For those baffles who have a low simulated normalized error results and a wide mode valid frequency range for most of the orders. We are confident to say they will have a good performance in representing the ambisonic sound field. But We can not forecast people's perception of the smaller baffles array. The important aspect in future investigation is to figure out at what point people will realize the loss of the spatial information when the captured sound field is represented in the different conceivable playback formats. This will make it more reliable when evaluating the performance in capturing spatial audio information for high-order microphone array baffles .

Bibliography

[1] Thushara D Abhayapala and Darren B Ward. Theory and design of high order sound field microphones using spherical microphone array. In *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages II–1949. IEEE, 2002.

[2] J. Ahrens, H. Helmholz, D. Alon, and S. V. Amengual Garí. Spherical Harmonic Decomposition of a Sound Field Based on Observations Along the Equator of a Rigid Spherical Scatterer. *J. Acoust. Soc. Am.*, (150), 2021.

[3] Jens Ahrens. *Analytic methods of sound field synthesis*. Springer Science & Business Media, 2012.

[4] Jens Ahrens. Audio 2: Microphones. Online, `http://www.ta.chalmers.se/content/protected/courses/ata/ATA_lec12.pdf`, 2022. last accessed: 2022-11-07.

[5] Jens Ahrens. Binaural audio rendering in the spherical harmonic domain: A summary of the mathematics and its pitfalls. *arXiv preprint arXiv:2202.04393*, 2022.

[6] Jens Ahrens, Hannes Helmholz, David L. Alon, and Sebastià V. Amengual Garí. A head-mounted microphone array for binaural rendering. In *Int. 3D Audio Conference (I3DA)*, Bologna, Italy, 2021.

[7] Jens Ahrens, Hannes Helmholz, David L. Alon, and Sebastià V. Amengual Garí. Spherical harmonic decomposition of a sound field based on microphones around the circumference of a human head. In *IEEE WASPAA*, New Paltz, NY, USA, 2021.

[8] Jens Ahrens and Ziyi Hu. Evaluation of non-spherical scattering bodies for ambisonic microphone arrays. In *Audio Engineering Society Conference: AES 2022 International Audio for Virtual and Augmented Reality Conference*. Audio Engineering Society, 2022.

[9] Sennheiser ambeo vr mic. Online, `https://en-us.sennheiser.com/microphone-3d-audio-ambeo-vr-mic`, 2022. last accessed: 2022-11-07.

[10] Head-related transfer function (hrtf). Online, `https://dictionary.apa.org/head-related-transfer-function`, 2022. last accessed: 2022-11-07.

[11] Ilya Balmages and Boaz Rafaely. Open-sphere designs for spherical microphone arrays. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(2):727–732, 2007.

[12] svein berge. acoustically hard 2d arrays for 3d hoa. *journal of the audio engineering society*, march 2019.

[13] Benjamin Bernschütz. *Microphone arrays and sound field decomposition for dynamic binaural recording*. Technische Universitaet Berlin (Germany), 2016.

[14] Terence Betlehem and Mark Poletti. Measuring the spherical-harmonic representation of a sound field using a cylindrical array. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 955–959. IEEE, 2019.

[15] Type 8606 spherical beamforming software. Online, `https://mhacoustics.com/products`, 2022. last accessed: 2022-11-07.

[16] miguel blanco galindo, philip coleman, and philip j. b. jackson. microphone array geometries for horizontal spatial audio object capture with beamforming. *journal of the audio engineering society*, 68(5):324–337, may 2020.

[17] Jens Blauert. *Spatial hearing: the psychophysics of human sound localization*. MIT press, 1997.

[18] Common techniques for stereo miking. Online, `https://www.shure.com/en-US/performance-production/louder/common-techniques-for-stereo-miking`, 2022. last accessed: 2022-11-7.

[19] Core sound tetramic™ 1st-order ambisonic microphone. Online, `https://www.core-sound.com/products/tetramic`, 2022. last accessed: 2022-11-07.

[20] Peter Graham Craven and Michael Anthony Gerzon. Coincident microphone simulation covering three dimensional space and yielding various directional outputs, August 16 1977. US Patent 4,042,779.

[21] Jérôme Daniel. Evolving views on hoa: From technological to pragmatic concerns. In *1st Ambisonics Symposium,(Graz)*, 2009.

[22] Stereo recording techniques and setups. Online, `https://www.dpamicrophones.com/mic-university/stereo-recording-techniques-and-setups`, 2022. last accessed: 2022-11-7.

[23] Eigenmike mh acoustics' patented em32 eigenmike microphone array. Online, `https://mhacoustics.com/products`, 2022. last accessed: 2022-11-07.

[24] Peter B Fellgett. Ambisonic reproduction of directionality in surround-sound systems. *Nature*, 252(5484):534–538, 1974.

[25] Michael A Gerzon. The design of precisely coincident microphone arrays for stereo and surround sound. In *Audio Engineering Society Convention 50*. Audio Engineering Society, 1975.

[26] Gopro max-6k 360 camera -gopro official site. Online, `https://gopro.com/en/us/shop/cameras/max/CHDHZ-202-master`, 2022. last accessed: 2022-11-7.

[27] Nail A Gumerov and Ramani Duraiswami. *Fast multipole methods for the Helmholtz equation in three dimensions*. Elsevier, 2005.

[28] Helmut Haas. The influence of a single echo on the audibility of speech. *Journal of the Audio Engineering Society*, 20(2):146–159, 1972.

[29] Head-related transfer function. Online, `https://en.wikipedia.org/wiki/Head-related_transfer_function`, 2022. last accessed: 2022-11-07.

[30] Joo Young Hong, Jianjun He, Bhan Lam, Rishabh Gupta, and Woon-Seng Gan. Spatial audio for soundscape design: Recording and reproduction. *Applied sciences*, 7(6):627, 2017.

[31] Kemar – the manikin for hearing aid testing and rd. Online, `"https://www.grasacoustics.com/industries/audiology/kemar"`, 2022. last accessed: 2022-11-07.

[32] Irving Langmuir, VJ Schaefer, CV Ferguson, and EF Hennelly. A study of binaural perception of the direction of a sound source. *General Electric Research Laboratory Rep*, 1944.

[33] Live planet vr, the first live streaming 360 stereo vr camera. Online, `https://liveplanetvr.com/camera`, 2022. last accessed: 2022-11-7.

[34] Tim Lübeck, Hannes Helmholz, Johannes M Arend, Christoph Pörschmann, and Jens Ahrens. Perceptual evaluation of mitigation approaches of impairments due to spatial undersampling in binaural rendering of spherical microphone array data. *Journal of the Audio Engineering Society*, 68(6):428–440, 2020.

[35] Jens Meyer and Gary Elko. A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield. In *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages II–1781. IEEE, 2002.

[36] Boaz Rafaely. Analysis and design of spherical microphone arrays. *IEEE Transactions on speech and audio processing*, 13(1):135–143, 2004.

[37] Boaz Rafaely. Plane-wave decomposition of the sound field on a sphere by spherical convolution. *The Journal of the Acoustical Society of America*, 116(4):2149–2157, 2004.

[38] Boaz Rafaely and Amir Avni. Interaural cross correlation in a sound field represented by spherical harmonics. *The Journal of the Acoustical Society of America*, 127(2):823–828, 2010.

[39] Boaz Rafaely, Barak Weiss, and Eitan Bachmat. Spatial aliasing in spherical microphone arrays. *IEEE Transactions on Signal Processing*, 55(3):1003–1010, 2007.

[40] Thomas D Rossing and Thomas D Rossing. *Springer handbook of acoustics*. Springer, 2014.

[41] Claude E Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949.

[42] RHY So, NM Leung, J Braasch, and KL Leung. A low cost, non-individualized surround sound system based upon head related transfer functions: An ergonomics study and prototype development. *Applied ergonomics*, 37(6):695–707, 2006.

[43] Sascha Spors, Heinz Teutsch, Achim Kuntz, and Rudolf Rabenstein. Sound field synthesis. In *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, pages 323–344. Springer, 2004.

[44] Soundfield sps200 software controlled microphone. Online, `https://www.trewaudio.com/product/soundfield-sps200-software-controlled-microphone/`, 2022. last accessed: 2022-11-07.

[45] JC Steinberg and WB Snow. Physical factors. *Bell System Technical Journal*, 13(2):245–258, 1934.

[46] Sweetwater. 7 stereo mic techniques you should try. Online, `https://www.sweetwater.com/insync/stereo-mic-techniques/`, 2016. last accessed: 2022-11-13.

[47] Sweetwater. Visual representations of the first few real spherical harmonics. Online, `https://en.wikipedia.org/wiki/Spherical_harmonics`, 2022. last accessed: 2022-11-13.

[48] Günther Theile. On the localisation in the superimposed soundfield. *Technische Universität Berlin*, 1980.

[49] Type 4101-b binaural microphone. Online, `https://www.bksv.com/ko/transducers/acoustic/binaural/binaural-microphone`, 2022. last accessed: 2022-11-07.

[50] Vuze camera: Home page. Online, `"https://vuze.camera/"`, 2022. last accessed: 2022-11-7.

[51] Markus Zaunschirm, Christian Schörkhuber, and Robert Höldrich. Binaural rendering of ambisonic signals by head-related impulse response time alignment and a diffuseness constraint. *The Journal of the Acoustical Society of America*, 143(6):3616–3627, 2018.

[52] Harald Ziegelwanger, Wolfgang Kreuzer, and Piotr Majdak. Mesh2hrtf: Open-source software package for the numerical calculation of head-related transfer functions. In *22nd International Congress on Sound and Vibration*, 2015.

[53] Harald Ziegelwanger, Piotr Majdak, and Wolfgang Kreuzer. Numerical calculation of listener-specific head-related transfer functions and sound localization: Microphone model and mesh discretization. *The Journal of the Acoustical Society of America*, 138(1):208–222, 2015.

[54] Dmitry N Zotkin, Ramani Duraiswami, and Nail A Gumerov. Plane-wave decomposition of acoustical scenes via spherical and cylindrical microphone arrays. *IEEE transactions on audio, speech, and language processing*, 18(1):2–16, 2009.

[55] Frank Zotter, Matthias Frank, Matthias Kronlachner, and Jung-Woo Choi. *Efficient phantom source widening and diffuseness in ambisonics*. 2014.

[56] Franz Zotter and Matthias Frank. *Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality*. Springer Nature, 2019.

Department of Architecture and Civil Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY