

# Implementation of the Radiation Characteristics of Musical Instruments in Wave Field Synthesis Applications

Dissertation zur Erlangung des Grades des Doktors der  
Philosophie an der Fakultät Geisteswissenschaften der  
Universität Hamburg im Promotionsfach Systematische  
Musikwissenschaft, vorgelegt von

Tim Ziemer  
Hamburg, im November 2014

Datum der Disputation: 20. Mai 2015

Erstprüfer: Herr Prof. Dr. Rolf Bader (Universität Hamburg)

Zweitprüfer: Herr Prof. Dr. Wolfgang Fohl (Hochschule für Angewandte Wissenschaften Hamburg)

Universität Hamburg, Fakultät für Geisteswissenschaften

**Acknowledgements**

Thanks to my advisor Prof. Rolf Bader and the examining committee: Prof. Wolfgang Fohl and Prof. Albrecht Schneider for thoroughly reviewing this dissertation and also Prof. Georg Hajdu and Prof. Clemens Wöllner for their time and expertise. I would also like to thank the team for their support, input, valuable discussions and for practicing my defense with me: Dr. Jost Leonhardt Fischer, Dr. Florian Pfeifle, Orie Takada, Claudia Stirnat, Niko Plath and Christian Koehn.

Finally, I dedicate this work to Wolli and to Gesa. From up there and down here you look after me.

# Contents

<b>1. Space in Music Perception</b>	<b>1</b>
<b>2. Psychoacoustics</b>	<b>3</b>
2.1. Thresholds and Just Noticeable Differences . . . . .	3
2.2. Critical Bands . . . . .	7
2.3. Spatial Hearing . . . . .	11
2.3.1. Research Conditions and Definition of Terms . . . . .	11
2.3.2. Localization in the Horizontal Plane . . . . .	13
2.3.3. Distance Hearing . . . . .	18
2.3.4. Localization of Various Sound Sources . . . . .	19
2.4. Auditory Scene Analysis . . . . .	21
2.4.1. Properties of Streams and their Elements . . . . .	22
2.4.2. Primitive Grouping Principles . . . . .	25
2.4.3. Schema-Based Grouping Principles . . . . .	29
2.4.4. Organization Based on Auditory Scene Analysis Principles .	30
2.5. Usability of Psychoacoustic Knowledge for Audio Systems . . . . .	33
<b>3. The Spatial Sound of Musical Instruments</b>	<b>35</b>
3.1. Wave Equation and Solutions . . . . .	35
3.1.1. Homogeneous Wave Equation . . . . .	36
3.1.2. Wave Field . . . . .	37
3.1.3. Homogeneous Helmholtz Equation . . . . .	38
3.1.4. Plane Waves . . . . .	38
3.1.5. Inhomogeneous Wave Equation . . . . .	40
3.1.6. Point Sources . . . . .	41
3.2. The Spatial Sound of Musical Instruments . . . . .	45



---

3.3. Measurement of the Radiation Characteristics of Musical Instruments	50
3.3.1. Far Field Recordings . . . . .	50
<b>4. Wave Field Synthesis</b>	<b>53</b>
4.1. Theoretical Fundamentals of Wave Field Synthesis . . . . .	53
4.1.1. Huygens' Principle . . . . .	53
4.1.2. Kirchhoff-Helmholtz Integral . . . . .	55
4.2. Practical Implementation of Wave Field Synthesis . . . . .	58
4.2.1. Constraints for Implementation . . . . .	59
4.2.2. Rayleigh-Integrals . . . . .	60
4.2.3. Spatial Border . . . . .	71
4.3. Existing Applications . . . . .	72
4.4. Wave Field Synthesis and Radiation Characteristics . . . . .	74
<b>5. Spectral Psychoacoustic Wave Field Synthesis</b>	<b>78</b>
5.1. Theoretical Framework . . . . .	79
5.1.1. Definition of the Radiation Characteristic . . . . .	80
5.1.2. Recreation of the Radiation Characteristic . . . . .	82
5.2. Practical Implementation . . . . .	83
5.2.1. Measurement . . . . .	84
5.2.2. Analysis and Storage . . . . .	86
5.2.3. Extrapolation . . . . .	88
5.2.4. Reconstruction . . . . .	90
5.3. The Loudspeaker Surrounding . . . . .	98
<b>6. Performance Tests of the Developed Audio Systems</b>	<b>102</b>
6.1. Physical Measurement of the Octupole Speaker . . . . .	102
6.2. Physical Measurement of the Surrounding Speaker System . . . . .	112
6.3. Comparison of Radiation Method and Minimum Energy Method . . . . .	116
6.3.1. Comparison of signal changes for slightly altered source positions . . . . .	117
6.3.2. Comparison of Created Wave Fields . . . . .	123
6.4. Psychoacoustic Investigations of the Surrounding Speaker System . . . . .	127
6.4.1. A comparison of Spectral Psychoacoustic Wave Field Synthesis, Stereo and Mono . . . . .	127

---

<b>7. Discussion and Prospects</b>	<b>136</b>
<b>A. Technical Data</b>	<b>154</b>
A.1. Calibration . . . . .	158
<b>B. Radiation characteristics</b>	<b>159</b>
<b>C. Wave Fields Created from Radiation- and Minimum Energy Method</b>	<b>168</b>
<b>D. Questionnaire</b>	<b>174</b>
<b>E. Summary of the Results</b>	<b>178</b>
<b>F. Kurzfassung der Ergebnisse</b>	<b>180</b>
<b>G. List of Publications Resulting from the Dissertation</b>	<b>182</b>

# List of Figures

2.1. Hearing Area . . . . .	4
2.2. JND in Sound Pressure Level. From Backus (1969), p. 86 . . . . .	4
2.3. JND in Temporal Order . . . . .	6
2.4. Cochlear Amplifier . . . . .	8
2.5. Critical Band Width. After Zwicker and Fastl (1999), p. 158 . . . . .	9
2.6. Free Field Room of the University of Göttingen. From Meyer et al. (1965), p. 340 . . . . .	12
2.7. Listening Plane. Created after Blauert (1997) . . . . .	13
2.8. Localization of Auditory Event in the Horizontal Plane. After Blauert (1997), p. 41, . . . . .	14
2.9. Localization Inversion. After Blauert (1974), p. 35 . . . . .	14
2.10. Lateralization per ILD. From Blauert (1997), p. 158 . . . . .	15
2.11. Lateralization per ITD. From Blauert (1997), p. 144 . . . . .	16
2.12. Binaural Distance Difference . . . . .	17
2.13. Distance Hearing of a Frontal Impulse. After Blauert (1997), p. 47 . . . . .	19
2.14. Demonstration of Emergence . . . . .	23
2.15. Vase-Face/Violing Group . . . . .	25
2.16. Good Continuation . . . . .	28
2.17. Closure. After Bregman (1990), p. 28 . . . . .	29
3.1. Propagating and Evanescent Wave . . . . .	40
3.2. Coordinate Systems . . . . .	43
3.3. Mauerlocheffekt . . . . .	46
3.4. ILD and IPD per listening angle and distance . . . . .	47
3.5. Frequency Range with Omnidirectional Radiation of Orchestral In- struments . . . . .	48

3.6. Microphone Array for Far Field Measurements. From Pätynen and Lokki (2010), p. 140 . . . . .	51
3.7. Polar Far Field Radiation Pattern . . . . .	52
4.1. Huygens' Principle . . . . .	54
4.2. Wave Fronts of a Breathing Sphere . . . . .	55
4.3. Acoustic Curtain. . . . .	55
4.4. Elementary Waves . . . . .	57
4.5. Three Volumes . . . . .	58
4.6. Illustration of the Incidence Angle . . . . .	68
4.7. Spatial Aliasing. . . . .	69
4.8. Truncation Effect. . . . .	71
4.9. WFS System of the Seebühne Bregenz. From Fraunhofer-Institut für Digitale Medientechnologie IDMT (2014) . . . . .	72
4.10. Fraunhofer IOSONO WFS System in the Tresor Club in Berlin. From IOSONO GmbH (2008), link: -> clubs and events. . . . .	73
4.11. WFS in the Auditorium of the TU Berlin. From Technical University Berlin (2015) . . . . .	73
4.12. Dodecahedron Speaker Array with 120 Loudspeakers. From Avizienis et al. (2006) . . . . .	75
4.13. Setup for Complex Radiation Synthesis. From Corteel (2007), p. 4 .	76
5.1. Measurement Setup for the Radiation Characteristic . . . . .	85
5.2. Radiation Patterns of Similar Frequencies . . . . .	87
5.3. Spectrum of the Bark Sound . . . . .	88
5.4. Illustration of a Case for an Interpolation . . . . .	90
5.5. Vision of the Octupole Speaker System . . . . .	91
5.6. Setup of the Octupole Speaker System . . . . .	92
5.7. Setup of the Surrounding WFS System . . . . .	93
5.8. Wavelength Deviation for an FFT over 512 Samples . . . . .	98
6.1. Measured Amplitudes of the Octupole Speaker Sound Beam Test .	109
6.2. Measured Relative Phases of the Octupole Speaker Sound Beam Test	111
6.3. Listening Room Including the Surrounding Loudspeaker Setup . . .	113

---

6.4. Setup of the Physical Performance Test of the Surrounding Loudspeaker . . . . .	114
6.5. Results of the Physical Measurement of the Surrounding Speaker System . . . . .	115
6.6. Calculated Loudspeaker Signals via Radiation- and Minimum Energy Method . . . . .	120
6.7. Comparison of Simulated Wave Field from Original Source, Radiation Method and MEM . . . . .	125
6.8. Results of the Listening Test (Mono) . . . . .	129
6.9. Results of the Listening Test (Stereo) . . . . .	130
6.10. Results of the Listening Test (WFS without Precedence Fade) . . .	131
6.11. Results of the Listening Test (WFS with Precedence Fade) . . . . .	132
A.1. Photo of the Ocupole Speaker System . . . . .	155
A.2. Photo of the Loudspeaker Surrounding . . . . .	157
B.1. Radiation Characteristic of a Shakuhachi (Amplitudes) . . . . .	161
B.2. Radiation Characteristic of a Shakuhachi (Phase) . . . . .	163
C.1. Simulated Wave Fields . . . . .	173

# List of Tables

2.1. Bark Scale . . . . .	10
6.1. Amplitudes of the Octupole Sound Beam Test . . . . .	105
6.2. Phases of the Octupole Sound Beam Test . . . . .	107
6.3. Demands on Amplitudes of Loudspeakers . . . . .	117
6.4. Loudspeaker Signals via Radiation- and Minimum Energy Method .	122
6.5. Wave Field Errors in Simulation . . . . .	126
7.1. Strengths and Weaknesses of Spectral Psychoacoustic WFS . . . . .	139

# List of symbols and abbreviations

## Symbols

$\alpha$	Angle between normal direction of wave front and secondary source
$\gamma(t - \tau)$	Time window
$\Gamma(\omega, \varphi)$	Solution for the azimuth angle of the Helmholtz equation
$\Gamma_{\vec{Q}}(\omega, \varphi)$	Horizontal radiation characteristic of the source
$\Gamma_{\vec{Y}}(\omega, \varphi)$	Horizontal radiation characteristic of the secondary source
$\delta$	Dirac-delta
$\phi$	Phase
$\varphi$	Azimuth angle
$\varphi_0$	Viewing direction
$\varphi_Q$	Azimuth angle of the source in the head-related coordinate system
$\lambda$	Wavelength
$\vartheta$	Polar angle
$\Theta$	Solution for the polar angle of the Helmholtz equation
$\rho$	Density
$\sigma$	Standard deviation
$\omega$	Angular frequency ( $\omega = 2\pi f$ )
$\Omega$	Width of the beamformer- or sound radiator lobes
$\Psi(\omega, \varphi, \vartheta)$	Spherical harmonics (solution to for azimuth and polar angle of the Helmholtz Equation)
$\hat{A}$	Amplitude

---

$A(\omega)$	Complex amplitude
$c$	Sound velocity
const	Constant
$d(\vec{Y})$	Windowing function
$D$	Manipulation factor turning source amplitude to transducer amplitude
dB	Decibels sound pressure level
dB <sub>SL</sub>	Decibels spectrum level
dB <sub>SPL</sub>	Decibels sound pressure level
dir	Direct sound
$f$	Frequency
$\vec{f}$	Steady differentiable vector function
FF	Far field
$g$	Special solution of the Green's function in time domain (impulse response)
$\tilde{g}$	General solution of the Green's function in time domain
$G$	Special solution of the Green's function in frequency domain (complex transfer function)
$\tilde{G}$	General solution of the Green's function in frequency domain
$i$	Imaginary unit $\sqrt{-1}$
$k$	Wave number ( $k = \frac{\omega}{c} = \frac{2\pi}{\lambda}$ )
$K$	Propagation matrix
$K(\omega, \varphi)$	Propagation function
$\vec{L}$	Position of the left loudspeaker
$L1$	Separation line between source area and source free area
lg	Decadic logarithm
max	Maximum
$\vec{M}_m(\vec{Q}, \varphi)$	Position of the microphones
NF	Near field
$p(t)$	Sound pressure in time domain (sound signal)
$P(\omega)$	Sound pressure in frequency domain (spectrum)
$P(t, \omega)$	Sound pressure in time-frequency domain



---

pre	Predicted
$Q$	Primary source
$\vec{Q}$	Primary source position
$Q'$	Mirror source
$\vec{Q}'$	Mirror source position
$r$	Radius
$\vec{r}$	Position vector in polar-/spherical coordinates
$\vec{r}'$	Mirror position of $\vec{r}$ in polar-/spherical coordinates
$\vec{R}$	Position of the right loudspeaker
s	Seconds
$S$	Surface of the source-free volume
$S1$	Separation plane between source volume and source-free volume (see fig. 4.5 b)
$S2$	Hemispherical separation surface between a source volume and a source-free volume (see fig. 4.5 b)
$SL$	Spectrum level
SPL	Sound pressure level
st	Static
$U$	Source volume (see fig. 4.5)
$\vec{v}$	Sonic particle velocity in time domain
$V$	Source-free volume (see fig. 4.5)
$\vec{x}$	Position vector in Cartesian coordinates
$\vec{X}$	Listening position
$\vec{Y}$	Secondary source position
$\vec{Y}'$	Mirrored secondary source position
$z$	Critical band from the Bark scale
$\nabla$	Nabla-operator
$\nabla^2$	Laplace-operator

## Abbreviations

AAC	Advanced audio coding (psychoacoustic compression algorithm)
AC-3	Adaptive transform coder no. 3 (psychoacoustic compression algorithm)
approx.	Approximately
ar.	Arithmetic
c#	C-Sharp programming language
ER	Early reflections
FFT	Fast Fourier transform
h/w/d	Height/width/depth
Hi-Fi	High-Fidelity (Quality demand on audio playback systems)
HRTF	Head related transfer function
ILD	Interaural level difference
ITD	Interaural time difference
JND	Just noticeable difference
K-H integral	Kirchhoff-Helmholtz integral
LA	Listening area
LR	Late reflections
MADI	Multichannel Audio Digital Interface
MEM	Minimum energy method
MP3	MPEG II audio layer 3 (psychoacoustic compression algorithm)
NAH	Nearfield acoustical holography
PA	Public address loudspeakers
PC	Personal computer
RMS	Root mean square (Comparative value for the power of amplifiers)
WFS	Wave field synthesis

# 1. Space in Music Perception

Space is not only important when it comes to localizing musical instruments. Even the frequency of sound is spatially coded in the cochlea<sup>1</sup>. This spatial organization is one reason of various psychoacoustic effects such as *masking*<sup>2</sup>. Although psychological mechanisms that allow for spatial orientation by acoustical signals are not fully understood, our capability of localizing sound sources has been extensively investigated and is well known today<sup>3</sup>. The way people represent the auditory outside world in a mental map for matters of orientation is called *auditory scene analysis* and is especially known from a phenomenological point of view<sup>4</sup>.

The human auditory system is not only capable of localizing sources. It is also able to distinguish original instrumental sound from loudspeaker playback by perceiving differences in their spatial sound radiation characteristics<sup>5</sup>. The sound radiation of musical instruments can be very complex due to the interplay of vibrating, radiating and reflecting parts of the instrumental body and enclosed air.

Quick advancements in computer technology and digital signal processing made it possible to physically recreating a natural sound field to a certain degree. Wave field synthesis disclosed new possibilities concerning spatial sound reproduction<sup>6</sup>. Applications typically creating virtual sources at arbitrary positions for a large listening area. Approaches exist to recreate not only the desired source location but also the desired sound radiation characteristic<sup>7</sup>.

---

<sup>1</sup>See e.g. Zwicker and Fastl (1999), p. 29.

<sup>2</sup>See e.g. Fastl (1977) and Gelfand (1990).

<sup>3</sup>See Blauert (1997).

<sup>4</sup>See Bregman (1990) and Theile (1980).

<sup>5</sup>See e.g. Warusfel et al. (1997), p. 1.

<sup>6</sup>See e.g. Berkhout (1988) and Spors et al. (2008)

<sup>7</sup>See e.g. Avizienis et al. (2006), Baalman (2008) and Corteel (2007).

In this thesis the sound radiation characteristics of musical instruments is investigated. A theoretic framework is developed to measure, store and recreate their radiation characteristics for an extended listening area based on adapted techniques of far field radiation measurements, wave field synthesis and psychoacoustic considerations. Furthermore, the *radiation method* is introduced, a method to implement the radiation characteristic of the loudspeakers to make calculations more robust and improve the precision of the reconstructed wave field. Based on this approach, further listening tests concerning the perception of instrumental sound and spaciousness are possible. In principle, the concept is transferable to conventional audio systems. It is realized in two loudspeaker systems.

The performance of the approach and the developed loudspeaker systems is investigated by physical measurements, simulations and listening tests. It has been found that the systems work especially well for stationary sounds. Recreating the radiation characteristics of a musical instrument can lead to the impression of an increased spaciousness. The perceived spaciousness of a sound does not necessarily increase with localization inaccuracy. This finding indicates that two independent mechanisms for localization and perception of spaciousness may exist. The precedence effect directs the perceived source location towards the origin of the first arriving wave front. This effect is applied systematically. By playing sound from one loudspeaker, and quickly fading in the signal of an adjacent speaker afterwards, a *fading based panning* is possible, similar to time- and amplitude based panning of conventional stereo.

Results of presented simulations, physical and psychoacoustic measurements are discussed against the background of applicability for research and development. An outlook about future steps to improve the introduced approach is given. Furthermore, potential applications areas of the methods for matters of research are named. Possibilities for combinations with other techniques, such as conventional wave field synthesis or Dolby Digital are outlined.

## 2. Psychoacoustics

The main function of the auditory system is a psychological representation of the physical world which relies on several psychoacoustic mechanisms outlined in this chapter. Perception is a result of filtering, analysis, segregation and integration. Certain absolute and relative thresholds limit the region of sound which is psychologically processed. Signals which do not surpass these thresholds are neglected for auditory processing and perception. Therefore, these thresholds are discussed next<sup>1</sup>. Many of the filtering processes are based on the spatial representation of sound within the cochlea, the critical bands, which are explained subsequently, followed by an associated psychoacoustic phenomenon, namely masking. Sound source localization and other aspects of spatial hearing are also performed separately for each critical band. Auditory scene analysis explains how those sounds that are not filtered out are grouped and mentally represented.

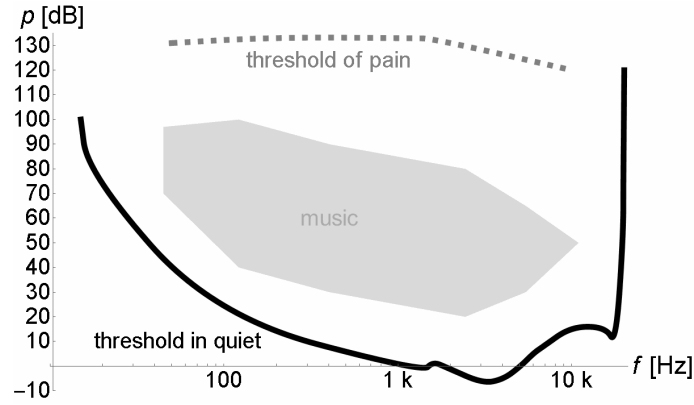
### 2.1. Thresholds and Just Noticeable Differences

The hearing area is illustrated in fig. 2.1. The solid line is the threshold in quiet the dashed curve the threshold of pain. Frequency- and sound pressure area of music does not include extreme cases like impulsive sounds in music which can cover a wider frequency range and higher sound pressure levels.

Just noticeable differences (JND) are thresholds of change of certain physical parameters. The JND in change of sound pressure roughly lies around 0.8dB, being

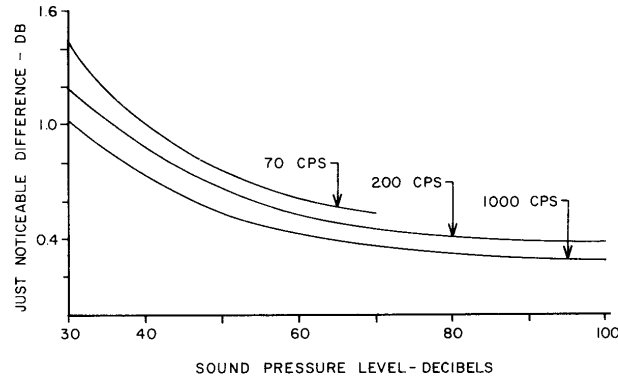
---

<sup>1</sup>Mainly based on Zwicker and Fastl (1999).



**Figure 2.1.:** Threshold of audibility and pain and area of orchestra music.

lowest around 1kHz at high sound pressure levels<sup>2</sup>. In this work, if not explicitly denoted differently, units in dB are referred to the sound pressure level, i.e. “1dB” means “1dB<sub>SPL</sub>”. The JND in SPL of successive tones is generally lower. It lies between 0.3dB 1.4dB depending on frequency and absolute SPL as can be seen in fig. 2.2.



**Figure 2.2.:** Just noticeable difference in sound pressure level for three different frequencies. “CPS” denote cycles per second. From Backus (1969), p. 86.

One temporal threshold of the auditory system is about  $T = 50\text{ms}$ , becoming less for increasing frequencies<sup>3</sup>. Successive acoustical events happening quicker than that are not discriminated but perceived as one sound or noise. This explains the lowest audible frequency of around  $\frac{1}{T} = 20\text{Hz}$ . 50ms is the time it takes for

<sup>2</sup>See Bruhn (2002b), pp. 667ff and Zwicker and Fastl (1999), pp. 175ff.

<sup>3</sup>See e.g. Bruhn (2002b), p. 669.

pitch perception to build up<sup>4</sup>. However, in an experiment the duration necessary to discriminate alternating from simultaneous complex tones has been found to lie below that threshold. This has been tested by Ziemer et al. with two complex tones with triangular wave form in the interval of a fifth in three frequency regions<sup>5</sup>. The tones are attached at their amplitudes or zero crossings to ensure that neither an impulsive sound nor silence occurs between them. Subjects were asked to judge whether the presented tones were perceived as clearly simultaneous (1), tend to be simultaneous or alternating (2 and 3) or clearly alternating (4). The arithmetic means are plotted in fig. 2.3. However, as we have an ordinal scale of measurement the median values just above and below 2.5 are considered as threshold between perceived simultaneity and alternation. With fundamental frequencies of about 1 and 1.5kHz, i.e. in the most sensitive frequency region, a duration between 7 and 11.2ms of each tone is sufficient to recognize that they are presented successively and not simultaneously. At higher frequencies a duration of 11.2 and 13.2ms is necessary. For very low frequencies even simultaneous tones were perceived as alternating but with a large deviation. This is not surprising since both fundamentals fall into the same critical band and are therefore hardly perceived as individual tones as will be discussed extensively in the subsequent section. Only at durations of 30ms and more subjects identified the alternating tones as clearly as for the other frequency regions with a small deviation. Here, considering median values, the threshold lies between 15.2 and 30.3ms. The test has been conducted with 33 musicology students, most of them trained musicians, who may have lower thresholds than an average person. The monaural temporal resolution of the auditory system is about 2ms<sup>6</sup>. It takes 2 to 5ms for a timbre to be perceived rather than a simple 'click'<sup>7</sup>. The temporal binaural resolution is even better. Interaural arrival time differences of several  $\mu$ s are sufficient for sound source localization as will be discussed below in ch. 2.3.

The JND of phase is difficult to quantify. Zwicker and Fastl deal with this subject and essentially give the following quantitative and qualitative statements<sup>8</sup>: Changes

---

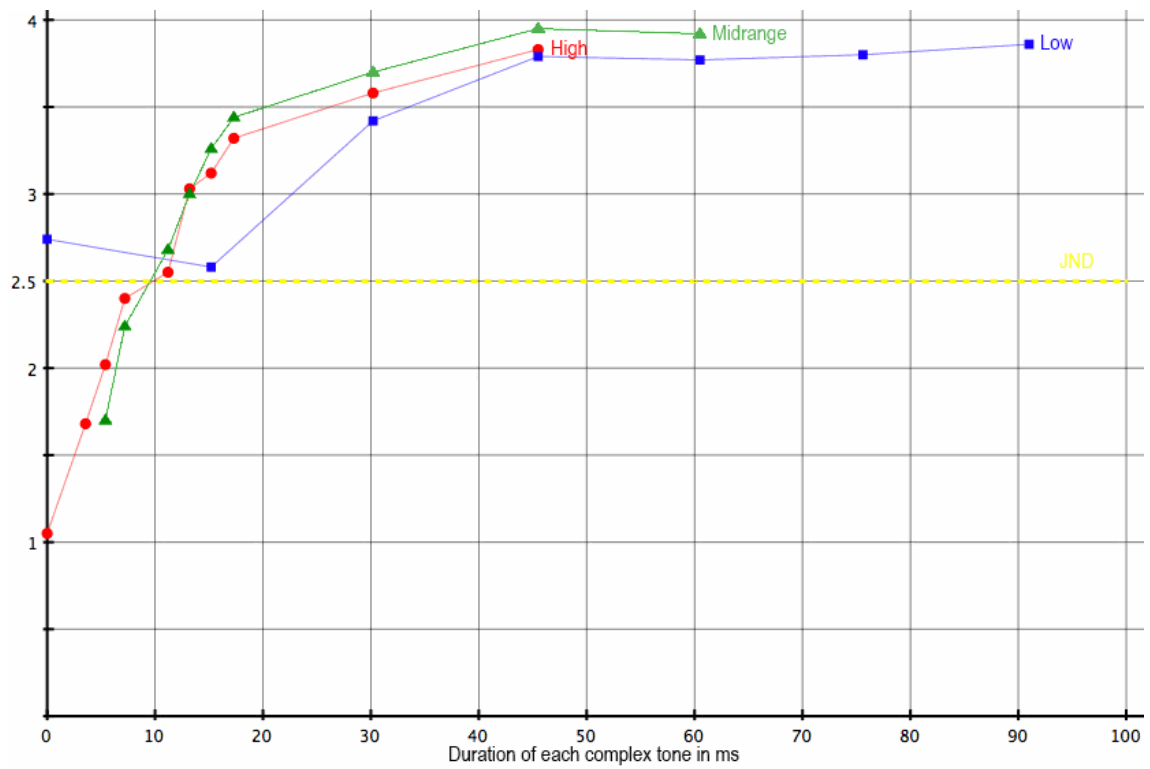
<sup>4</sup>See e.g. Bader (2013), p. 325.

<sup>5</sup>The experiment is described in detail in Ziemer et al. (2007).

<sup>6</sup>Cf. Zwicker and Fastl (1999), p. 293.

<sup>7</sup>See Bader (2013), p. 324.

<sup>8</sup>These data are given in Zwicker and Fastl (1999), pp. 188–191.



**Figure 2.3.:** Just noticeable difference in temporal order for low (33 and 49.5Hz), midrange (1056 and 1584Hz) and high (5280 and 7920Hz) sounds with triangular waveform.



of phase result in instantaneous frequency changes and a change of envelope. In experiments with three pure tones of equal amplitude the just noticeable difference in phase change of one tone relative to the others has been found to lie around  $10^\circ$  at best conditions, two times as much at worse laboratory conditions and again three times as much in everyday-environments like the living room. It is audible apparently due to the change in envelope. According to Webers, phase distortions in stationary sounds are not audible<sup>9</sup>.

Thresholds of frequency discrimination, masking and sound source position exist as well and are discussed in the following sections.

## 2.2. Critical Bands

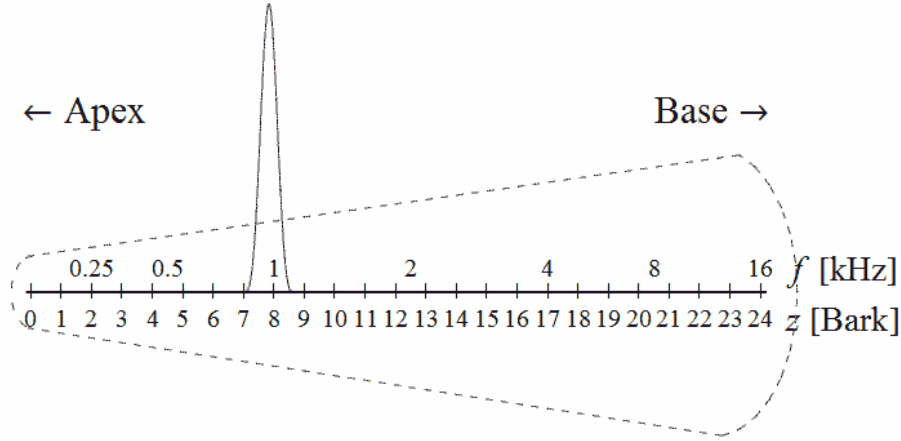
Movements of the oval window induce a traveling wave in the cochlea. This wave propagates inwards, slowly builds up and suddenly collapses after reaching the climax, its amplitude. The local elongations of the tectorial membrane relative to the basilar membrane — induced by this traveling wave — cause outer hair cells to shear which leads to a neural firing proportional to the elongation. Due to complicated biomechanical effects, the location of the amplitude is dependent on frequency. This frequency-place transformation is known as the “place principle” and means that every frequency has a corresponding area on the tectorial membrane, i.e. frequency is spatially coded within the cochlea<sup>10</sup>. This means sound pressure is processed in time-frequency domain by the auditory system. Furthermore, frequencies that do not have a corresponding resonance area in the cochlea are filtered out by that cochlear filter mechanism. This is the case for frequencies below 16Hz and above 20kHz. The minimum sound pressure necessary to activate a cochlear amplification is the absolute threshold. The higher the frequency the closer the amplitude lies to the oval window. Consequently, the traveling wave of a low frequency passes the area of higher frequencies and evokes cochlear activity. It resonates in a certain region which is narrowed by a cochlear amplifier. Behind the resonance region the traveling wave collapses. The traveling wave of a frequency

---

<sup>9</sup>See Webers (2003), p. 128.

<sup>10</sup>See e.g. Zwicker and Fastl (1999), p. 29.

has to surpass the envelope of simultaneous traveling waves to protrude and thus be audible. If not, it will be masked. Figure 2.4 schematically illustrates the envelope of a traveling wave and the cochlear amplification.



**Figure 2.4.:** Schematic diagram of a rolled-out cochlea (dashed contour) with the envelope of a traveling wave induced by a frequency of 1kHz (light gray). At its peak the neural firing is amplified (dark gray curve) by a cochlear mechanism. The abscissa illustrates the linear alignment of frequencies in Bark in contrast to the logarithmic distribution in Hertz.

The region affected by the cochlear amplifier is the so-called *critical band*. The width of a critical band is about 1.3mm and roughly includes 160 hair cells. Within one critical band approximately 25 JND in pitch can be discriminated. Frequencies that simultaneously fall into the same critical band cannot be identified individually. They create a common sound impression. Depending on their interval, it is the impression of one single note, beats or roughness. Only frequencies from different frequency bands can be heard as different tones with a certain interval. The Bark scale divides the cochlea into 25 fixed equal areas  $z$  representing critical bands as can be seen in fig 2.4. Every frequency can be transferred to its corresponding position on the cochlea:

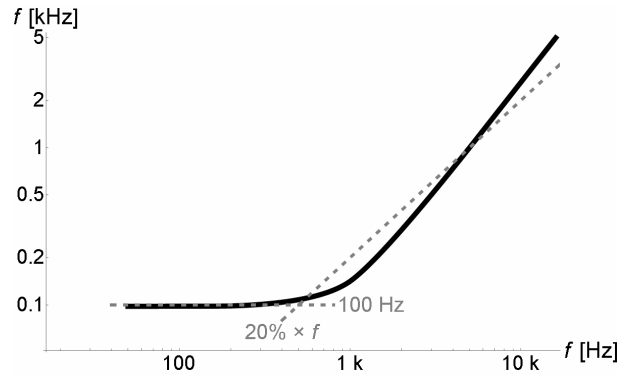
$$z[\text{Bark}] = 13 \arctan(0.76f) + 3.5 \arctan\left(\left(\frac{f}{7.5}\right)^2\right) \quad (2.1)$$

The width of one Bark, the critical band width  $\Delta f_{\text{critical}}$  can be calculated as follows:

$$\Delta f_{\text{critical}} = 25 + 75 \left(1 + 1,4(f)^2\right)^{0.69} \quad (2.2)$$

Table 2.1 lists the Bark scale and the corresponding lower and upper boundary frequency  $f_l$  and  $f_u$ , the mean frequency  $f_{\text{mean}}$  and the critical band width<sup>11</sup>. The mean frequency does not represent the arithmetic mean of lower and upper frequency but anatomically the mean position within one critical band on the basilar membrane.

Up to a center frequency of 500Hz, the critical band width is about 100Hz. From there on it is approximately 20% of the frequency. Certainly, the Bark scale does not describe the whole complex nature of critical bands. In fact, they have no fixed positions but are areas around a frequency as can be seen in fig. 2.5.



**Figure 2.5.:** Plot of the critical band width over frequency. After Zwicker and Fastl (1999), p. 158.

The auditory system does not process sound as a whole but in frequency bands<sup>12</sup>. This is an important fact in the context of spatial hearing and auditory scene analysis, which are discussed in the following.

<sup>11</sup>Cf. Zwicker and Fastl (1999), p. 159.

<sup>12</sup>See e.g. Blauert (1974), p. 173, Allen (2008), pp. 28ff and Kostek (2005), p. 9.

$z$ [Bark]	$f_l$ [Hz]	$f_u$ [Hz]	$\Delta f_{\text{critical}}$ [Hz]	$f_{\text{mean}}$ [Hz]
0	0	100	100	50
1	100	200	100	150
2	200	300	100	250
3	300	400	100	350
4	400	510	110	450
5	510	630	120	570
6	630	770	140	700
7	770	920	150	840
8	920	1080	160	1000
9	1080	1270	190	1170
10	1270	1480	210	1370
11	1480	1720	240	1600
12	1720	2000	280	1850
13	2000	2320	320	2150
14	2320	2700	380	2500
15	2700	3150	450	2900
16	3150	3700	550	3400
17	3700	4400	700	4000
18	4400	5300	900	4800
19	5300	6400	1100	5800
20	6400	7700	1300	7000
21	7700	9500	1800	8500
22	9500	12000	2500	1050
23	12000	15500	3500	13500
24	15500			

**Table 2.1.:** Bark scale and corresponding frequencies.

## 2.3. Spatial Hearing

The ability of people to localize sound sources is a well researched topic in psychoacoustics. The literature of Blauert<sup>13</sup> is considered a standard work summarizing the state of research in the seventies — with the postscript in the mid-eighties — especially derived from listening tests. Supplemented by Blauert and Braasch (2008), Dickreiter (1987), Webers (2003) and Strube (1985), the most important results concening localization in the horizontal plane are expounded in this section.

### 2.3.1. Research Conditions and Definition of Terms

The auditory system uses a variety of analytical processes of the acoustical signals impinging the eardrum for spatial hearing. These usually proceed preconscious: Sensations are assigned to a direction and distance in a bottom-up-processing<sup>14</sup>. In case of familiarity with the signal or the acoustical environment top-down-processes support the localization<sup>15</sup>. Parameters gathered for localization are *interaural time difference* (ITD) and *interaural level differences* (ILD) and an individual filtering of the signal depending on incidence angle. The *Head-Related Transfer Function* (HRTF) quantifies the changes a sound undergoes from its origin to the eardrums of a listener as resulting from sound propagation and filtering, caused by reflections in the pinna, diffractions around and acoustic shadow behind head, torso, shoulders, etc. Sound playback via headphones often leads to a localization inside the head. Hence, it is only spoken of *lateralization* in this case. Localization parameters and accuracy immensely distinguish between spatial planes, therefore they are illuminated individually.

People tend to consult visual cues additionally to hearing for the localization of sound sources, as will be discussed in detail in ch. 2.4. In the listening tests this is

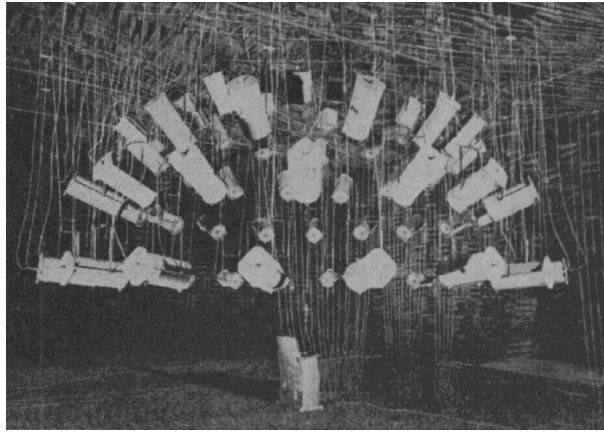
<sup>13</sup>See Blauert (1974) and Blauert (1985).

<sup>14</sup>See Blauert (1997), p. 409.

<sup>15</sup>See Bruhn (2002a), p. 444, definitions of bottom-up- and top-down-processing see e.g. Myers (2008), p. 214.

avoided by blindfolding the subjects. Also, bone conduction is neglected because its auditory threshold is more than 40dB higher than the threshold of the eardrums.

The following tests take place in free field rooms exclusively, such as the room illustrated in fig. 2.6. The probands sit in the center of a hemispherical loudspeaker array with fixed heads or in a darkened free field room and judge the assumed origin direction of a sound verbally or by pointing.

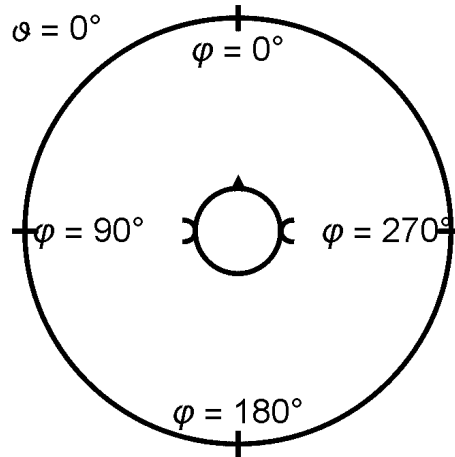


**Figure 2.6.:** Free field room of the University of Göttingen during a test arrangement with 65 loudspeakers. From Meyer et al. (1965), p. 340, with friendly Permission of Deutscher Apotheker Verlag.

Typical test stimuli are pure tones, narrow band and broad band noise and occasionally speech. Furthermore, Gaussian impulses are used. These are pure tones multiplied by a Gaussian function which gives the continuous tones the envelope of the Gaussian bell curve function resulting in a small spectral widening and imprecise note on- and offsets. If nothing else is alluded, one signal at a time is concerned. The distance of the source is so big that the wave fronts reaching the listener can be considered as being plane. Subjects are people without hearing loss.

For sound source localization it is meaningful to use a head-related spherical coordinate system with the head as the origin. The horizontal plane can be seen in fig. 2.7. In this thesis, only the horizontal plane is considered, mainly because the loudspeaker systems to be developed are controlled in this plane.

The actual direction/distance is denoted *sound event direction/distance*. *Auditory event* is the position of the source as assumed by the subject. If a subject is only



**Figure 2.7.:** The horizontal listening plane after Blauert (1997).

able to indicate an area, the range of the area is called *localization blur*. It is not exactly clear whether the localization blur denotes insecurity in localization or a perceived extent or width of the source.

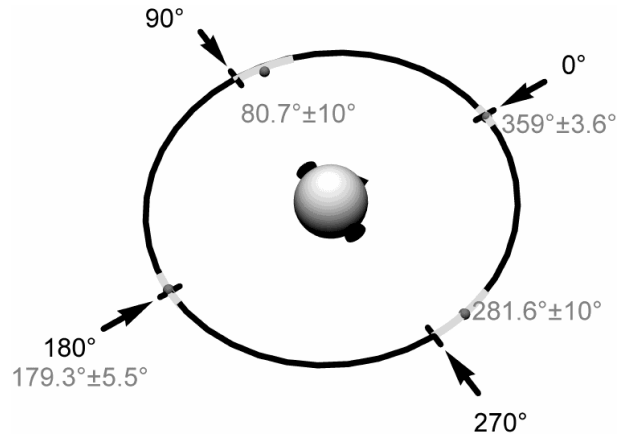
### 2.3.2. Localization in the Horizontal Plane

**Localization capability:** Mechanisms to localize sound sources are superior in the horizontal plane compared to the median plane. Horizontal location is especially based on binaural signal comparisons. The localization capability is best in the frontal area with an average accuracy of  $1^\circ$  and a localization blur of about  $\pm 3.6^\circ$  — as illustrated in fig. 2.8 — which corresponds to the JND in position in this spatial region<sup>16</sup>. However, it is dependent on the signal, being largest in a frequency range between about 1 or 1.5kHz and 3 to 4kHz.

Towards the sides the deviations and the localization blurs distinctly grow. Completely lateral signals are typically estimated too frontal. Here, the JND in position lies between  $12$  and  $18^\circ$ <sup>17</sup>. Especially for unfamiliar and narrow band signals the auditory event direction lies axial symmetric to the sound event. This effect is known as *localization inversion* and is illustrated in fig. 2.9. A typical mistake

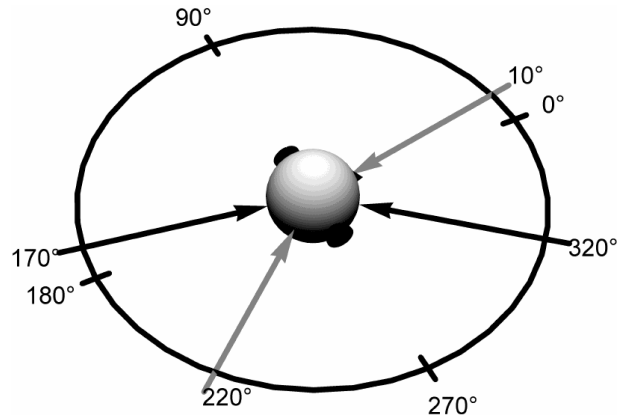
<sup>16</sup>See e.g. Blauert (1997), pp. 39f and Webers (2003), p. 120.

<sup>17</sup>See Webers (2003), p. 120.



**Figure 2.8.:** Auditory event direction and localization blur in case of a fixed sound event in the horizontal plane, determined with 600 to 900 subjects. After Blauert (1997), p. 41, data taken from Haustein and Schirmer (1970) and Preibisch-Effenberger (1966).

in localization is a “front-back reversal”, also called “front-to-rear confusion”, especially for low frequencies. At higher frequencies, the HRTF — mostly due to wave shadow behind the pinna — yield audible spectral differences between sources from the front and the rear, which inform subjects on the source position, if they are familiar with the sound<sup>18</sup>.

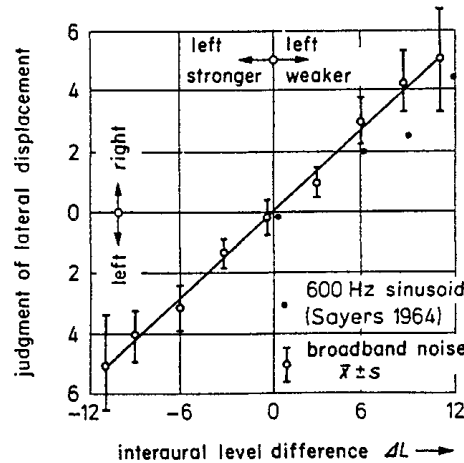


**Figure 2.9.:** Sound event direction (black) and auditory event direction (gray) in cases of localization inversions in the horizontal plane. After Blauert (1974), p. 35.

<sup>18</sup>See Kling and Riggs (1971), p. 351 and Blauert (1997), p. 360.



The coherence between ILD and lateralization of sounds through dichotic headphones is almost linear but with quite quite a localization blur especially off the center as can be seen in fig. 2.10.

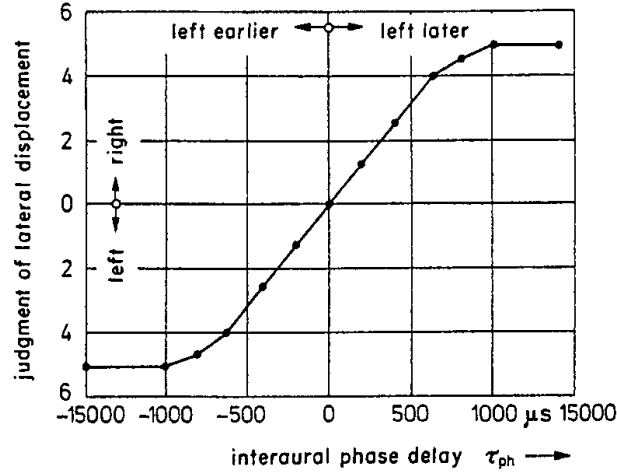


**Figure 2.10.:** Lateralization and lateralization blur per interaural level difference (in dB) linearly scaled from 0 (= center) to 5 (=at the auricular entrance). Reprint from Blauert (1997), p. 158, with friendly permission of The MIT Press.

However, the auditory event angle per ILD is dependent on frequency. For broadband signals a level difference above 11dB leads to a total lateral impression, narrow banded signals need bigger differences. The coherence between interaural phase difference and lateralization is also relatively linear in the range from 0-80% of a completely lateral angle until approximately  $640\mu\text{s}$  arrival time difference. This can be seen in fig. 2.11.

ITD above  $640\mu\text{s}$  lead to hardly any increase of the auditory event angle, probably because a source actually placed at  $90^\circ$  leads to an ITD of roughly  $640\mu\text{s}$ , assuming a head-radius of 8.5cm. It has to be taken into account that completely lateral signals are localized  $10^\circ$  too frontal, as discussed above and shown in fig. 2.8. According to Myers, the JND in ITD lies at about  $27\mu\text{s}$  whereas Zwicker and Fastl consider it to lie at about  $50\mu\text{s}$ , mentioning that individual values between 30 and  $200\mu\text{s}$  have been measured. Kling and Riggs even state that values up to  $300\mu\text{s}$  are possible<sup>19</sup>. Kling and Riggs quantified the relationship between incidence angle

<sup>19</sup>See Myers (2008), p. 240, Zwicker and Fastl (1999), pp. 293f and Kling and Riggs (1971), p. 355.



**Figure 2.11.:** Lateralization per ITD. Reprint from Blauert (1997), p. 144, with friendly permission of The MIT Press.

of a source and the length of the path of its propagating wave to the two ears for a head in the near field and in the far field of the source, considering diffraction of a wave around a sphere representing the listener's head<sup>20</sup>. This formulation is illustrated in fig. 2.12. Dividing this path difference by sound velocity yields the relationship between source angle and ITD<sup>21</sup>:

$$\begin{aligned} ITD_{\text{NF}} &= \frac{2r\varphi}{c} \\ ITD_{\text{FF}} &= \frac{r(\sin\varphi + \varphi)}{c} \end{aligned} \quad (2.3)$$

Here, the subscripts NF and FF denote the near field and the far field,  $r$  is the radius of the sphere,  $c$  is the sound velocity and  $\varphi$  is the azimuth angle of the source in a head-related coordinate system where  $-\frac{\pi}{2} \leq \varphi \leq \frac{\pi}{2}$ . For a source in the far field, this formulation can be extended to cover sources beyond the horizontal plane<sup>22</sup>:

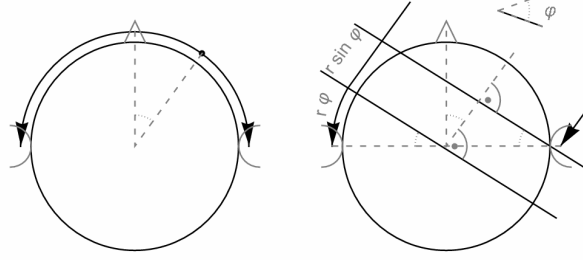
<sup>20</sup>See Kling and Riggs (1971), p. 351.

<sup>21</sup>See Kling and Riggs (1971), p. 352.

<sup>22</sup>See Larcher and Jot (1999).

$$ITD_{FF} = \frac{r (\arcsin (\cos \varphi \sin \vartheta) + \sin \varphi \cos \vartheta)}{c} \quad (2.4)$$

$-\frac{\pi}{2} \leq \vartheta \leq \frac{\pi}{2}$  is the elevation angle.



**Figure 2.12.:** Binaural distance difference for a near (left) and a far source (right).

Contradictory interaural attributes might compensate each other via *trading* or the signal is perceived as two signals from different directions where low frequencies seem to arrive from the direction suggested by phase difference, whereas high frequencies seem to arrive from the direction derived from ILD.

**Demands on the signal:** Very low frequencies reveal barely no evaluable level differences due to a negligible wave shadow behind the head. Likewise, the ITD yields no detectable phase differences. It is these minor cues that make it difficult for the auditory system to localize low frequencies. Here, the ITD — especially of onsets, transient sounds, short signals and the envelope of sound — play a central role<sup>23</sup>. Low to medium frequencies may show an evaluable phase difference due to ITD which becomes a dominant localization cue in that frequency region. In the range from about 1.5kHz to 3kHz the localization capability is poor despite the high level-sensitivity in this region. Here, on the one hand, the frequencies are too high for unambiguous phase relations. Furthermore, the auditory neurons are not capable of firing rapidly enough to display the phase difference at higher frequencies<sup>24</sup>. On the other hand, the wave lengths are too large to create noticeable level differences by acoustic wave shadow. For higher frequencies, filtering by head, hair, pinna and shoulders cause ILD and make it the dominant aspect

<sup>23</sup>See e.g. Kling and Riggs (1971), pp. 350ff and Morikawa and Hirashara (2010), p. 419.

<sup>24</sup>See Hall (2008), p. 343 and Davis (2007), p. 750.

which leads to a proper localization even for stationary sounds. Furthermore, the human auditory system is capable of detecting ITD of the envelope of high frequencies but it is unknown whether these envelope delays deliver reliable localization cues<sup>25</sup>. The spectrum resulting from the individual HRTF shows prominent peaks and notches between 3 and 14kHz which support localization. Still, localization of high pass noise above 16kHz is imprecise because ILD are the only evaluable cues<sup>26</sup>. Thus, a good localization demands a large bandwidth, transients and distinct sound envelopes. However, the bandwidth of many musical instruments is so large that both aspects, ITD and ILD, occur in combination<sup>27</sup>. For front/back localization the direction-dependent filtering, HRTF, delivers the only evaluable cues. The HRTF is very individual, depending on size and shape of head, trunk, pinna and torso.

### 2.3.3. Distance Hearing

**Localization capability:** Since during the attempts about distance hearing some sources are placed very close to the subject, the wave fronts reaching the ears cannot be considered as being plane. For distance estimation the familiarity with the signal is of great importance. For example, the distance of normal speech in the front can be estimated rather precisely. On the other hand, the distance of whispering tends to be underestimated and does not exceed 3m whereas calling out loudly sounds further away than it actually is<sup>28</sup>.

For close unfamiliar signals the auditory event lies too close up to a localization inside the head or the impression of a source directly behind the head. The perceived spectrum from familiar sounds helps to localize distances between 0.25 and 15m since the pressure level decays evenly whereas perceived loudness does not because contours of equal loudness change over the overall amplitude<sup>29</sup>. In a natural

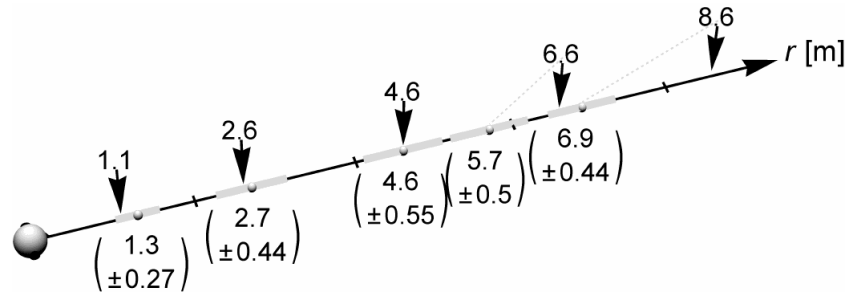
<sup>25</sup>See e.g. Middlebrooks and Green (1991), pp. 142f.

<sup>26</sup>See Morikawa and Hirashara (2010), p. 419.

<sup>27</sup>The so-called “duplex theory”, see e.g. Bruhn and Michel (2002), p. 651.

<sup>28</sup>A good illustration of the actual loudspeaker distance and the auditory event distance for normal speech, whispering and calling out loudly is given in Blauert (1997), p. 46.

<sup>29</sup>Detailed information on the *inverse distance law* which describes the sound pressure decay is given in ch. 3.1.6. Details on contours of equal loudness can be found e.g. in Zwicker and Fastl (1999), pp. 203ff.



**Figure 2.13.:** Sound source and auditory event distance for impulses of approximately 70Phon. The gray lines connect the related sound event and auditory event. After Blauert (1997), p. 47, with data from Haustein (1969).

environment the pressure level relation between direct sound and first reflections denote the distance. For distances further than about 15m high frequencies are damped stronger than lower ones, because the distance between areas of excess and reduced pressure are smaller in cases of small wavelengths. Thus, more acoustic energy transfers to heat exchange. Therefore, signals from afar sound more dull.

**Demands on the Signal:** When a listener is familiar with the signal, loudness and spectrum are important distance clues. The distance of a completely unknown sound can't be localized under laboratory conditions. Even for more familiar sounds distance hearing is imprecise. Screaming is localized too far away, whispering too close by. Best localization accuracy can be observed for distances between 1 and 5m. The localization ability notably improves when the sound is heard in a natural, known environment with its specific spatial acoustics. Then, arrival time- and loudness difference between direct sound and early reflections give applicable distance cues.

### 2.3.4. Localization of Various Sound Sources

Sound from monopole sources in the front and the back, i.e. at  $0^\circ$  and  $180^\circ$ , sound will reach both ears simultaneously. For any other location, the sound will reach the ears at different points in time. The auditory system uses these interaural time differences to localize sounds, which is only possible because it correctly interprets

the difference between the signals arriving at the two ears as a result of different traveling times of one sound coming from one source. However, this interpretation by the auditory system is prone to mistakes, since even the sounds from various sources can be perceived as coming from one source. In this case several sources are interpreted as one single source with one concrete position. According to the idea of “summing localization” this perceived position is exactly that position at which a sound source would have to be located to create similar ITD and ILD<sup>30</sup>. This position does not necessarily have to coincide with one of the actual source positions. If such a position is found, the localization is distinct. If not, a wide or diffuse sound source is perceived. Theile criticized the theory of summing localization as a result of simple comparison between localization cues of a superimposed sound field with that of a single source. In his “association model” he expands it by auditory mechanisms which later became part of the auditory scene analysis principles as formulated and extensively discussed by Bregman to explain the general psychological treatment of sound<sup>31</sup>. Auditory scene analysis will be addressed in the following section, ch. 2.4. However, in the case of several sources present, another localization effect can occur. The “precedence effect”, “Haas-Effekt” or “Law of the first wavefront”<sup>32</sup> indicates that sound events are localized solely in the direction of the first wave front arriving at the ears, even if later arriving sounds are much louder. Even if a sound arriving with a delay of 5 to 30ms is 10dB louder than the first arriving wave front it won’t affect the localization<sup>33</sup>. The effect occurs especially but not exclusively with transient signals, particularly at onsets. A frontal sound is localized correctly, even if lateral reflections reach the ears, since the first wavefront was already crucial for the localization. Premise is that the first and second arriving signal fuse, i.e. that they are integrated into one auditory stream as will be discussed in the upcoming section. From a time delay of about 50ms on, auditory event and echo are perceived individually as it is a typical threshold of the auditory system<sup>34</sup>. The precedence effect can last for seconds and more<sup>35</sup>. Zurek

---

<sup>30</sup>See e.g. Strube (1985), p. 69.

<sup>31</sup>See Theile (1980) and Bregman (1990).

<sup>32</sup>See David jr. (1988), p. 159, Friedrich (2008), p. 39, Hall (2008), p. 469 and Blauert (1997), p. 411.

<sup>33</sup>See Dickreiter (1978), p. 77.

<sup>34</sup>See e.g. Blauert (1974), p. 180 or Strube (1985), p. 68.

<sup>35</sup>According to Blauert (1974), p. 224.

and Saberi found evidence that the precedence effect does not fully suppress other localization cues that follow the onset. Rather an interaural cross correlation after the onset can stabilize or adjust the auditory event position<sup>36</sup>.

## 2.4. Auditory Scene Analysis

Bregman published an extensive elaboration regarding the perceptual organization of sound, called "auditory scene analysis". He demonstrates it on the basis of laboratory experiments with artificial sounds, such as sinusoidal tones or noise, as well as from listening experience and experiments with music. This section summarizes the essence of his work<sup>37</sup>.

Whatever we hear is our perceived part of the acoustic outside world: the auditory scene. Single units in the auditory scene are called *auditory streams*, a pendant to the *visual object*. In natural listening situations sounds from different acoustic sources overlap in time and spectrum and the sound pressures reaching the ears are always the sum of all propagated sounds and their reflections. The task of the auditory system is to analyze these complex sounds to be able to identify what parts belong together (*integration*) as well as to discriminate between different streams (*segregation*). This *grouping* is the attempt of the auditory system to create a mental representation of the world in which every stream is derived from the same environmental entity or *physical happening*. Such a categorical perception is crucial for a proper understanding of and orientation in the outside world. Auditory scene analysis is not an explanation of how exactly this is accomplished by the auditory system by means of biological, biochemical, physiological or neurological functionality. Rather it describes organization patterns which can be observed in the perception of the acoustic environment, most of which are primitive, innate, pre-attentive bottom-up grouping processes, whereas higher levels of grouping are schema-based, attention-directing top-down processes, according to our knowledge of attributes and behavior of familiar sounds<sup>38</sup>.

---

<sup>36</sup>See Zurek and Saberi (2003).

<sup>37</sup>Particularly based on Bregman (1990).

<sup>38</sup>See Bregman (1990), pp. 38f, 137, 204, 395, and 451.

If components have arisen from the same physical event they naturally have many more things in common than could be explained by chance, e.g. timing, frequency, and differential effect on our two ears. There exists not one exclusive parameter which determines auditory scene since there is no law of nature from which we could derive a concept for an adequate auditory scene analysis. Rather a complex system of certain principles is used for this task, many of which are known from Gestalt psychology. The necessity for this redundancy can be easily explained. One could think that the location of a sound source is a proper parameter to distinguish different physical happenings since only one thing can be on one position at a time. But firstly, our localization capability is quite weak e.g. in the lateral region and especially in the vertical dimension, as discussed previously in this chapter. Secondly, it is crucial to understand an echo as reflection of a direct sound and assign it to the same physical happening, even if locations are different. Thirdly, a correct localization from a mixture of sounds from various sources already requires a correct grouping. And finally, we would not be able to distinguish several sounds from a monophonic presentation if location was the only parameter considered. This deficiency is found for every single parameter, such as similarity in pitch, timbre or temporal change or proximity in onset or spectrum. Among others, organization principles base on the named parameters and are described more extensively in the following, subdivided into three categories, namely:

1. Properties of streams and their elements
2. Primitive grouping principles
3. Schema-based grouping principles

### **2.4.1. Properties of Streams and their Elements**

Several principles concern the properties of auditory streams and their elements. These are predominant and can be considered as framework for auditory scene analysis. Given examples refer to grouping principles which are explained in detail later in this subsection.



**Emergence:** Integration into streams takes a certain time which can vary. But still there is no arising or fading in of streams. Grouping takes place spontaneously, even when controlled by the listener's attention. This principle is visualized in fig. 2.14. The person in the lower right corner does not take form bit by bit but emerges as a whole although hardly any body part is drawn completely.



**Figure 2.14.:** Demonstration of emergence: The person standing on the right emerges as a whole, not body part by body part, although his legs are hardly visible.

**Simultaneous and sequential grouping:** All sounds created from a source last for a certain time and undergo some changes. Therefore, it is necessary to group sounds that arrive simultaneously at the listener (*simultaneous/vertical grouping*) and sounds arriving at different times (*sequential/horizontal grouping*). While simultaneous grouping is necessary to discriminate different auditory streams from the sum of arriving sound, sequential grouping is needed to keep track of streams and to trace back continuous or successive sounds to the same physical happening. Simultaneous and sequential grouping are not independent of one another. Principles can affect both and auditory streams typically consist of both.

**Units:** Auditory streams are units which can be embedded in larger streams and form a “higher order unit”<sup>39</sup>. Being perceived as one object does not mean we

---

<sup>39</sup>See Bregman (1990), p. 72.

cannot differentiate between single parts of the object. We do not only see the dog in fig. 2.14 but also his head, tail and so on. Imagine the dog running through the living room. Although we know the dog has a head, feet, etc. the single parts of the dog are not considered separately, since the dog cannot run through the living room leaving his head in the kitchen. The same counts for auditory streams. One note, played on the piano, will certainly be integrated into one stream. But still, at least with some training, it is possible to hear out some single frequencies from its spectrum.

**The old-plus-new-heuristic:** If a part of current sound can be interpreted as continuation of sound before, it will be integrated and then the remaining part is analyzed for grouping. It is also referred to as “wrap up all your garbage in the same bundle” heuristic<sup>40</sup>. Unattended elements can still be grouped within a stream. It even makes it easier to reject them as a group.

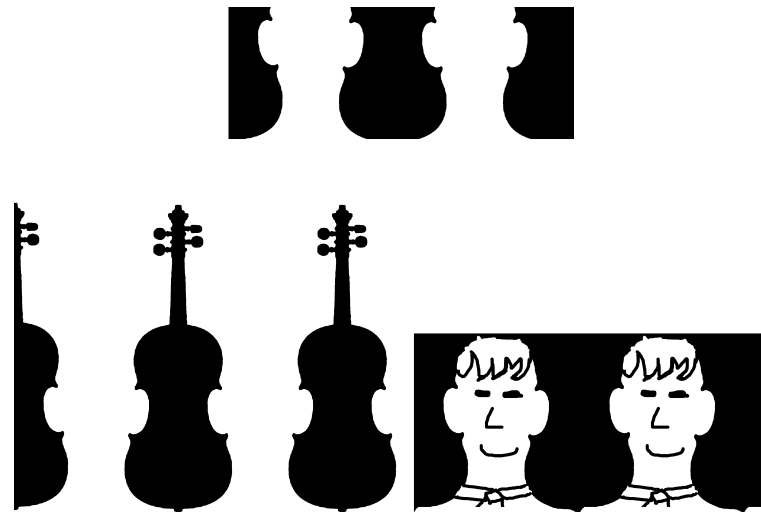
**The principle of belongingness:** The principle of belongingness forces *exclusive allocation* of sound parameters. That means every aspect of sound is always exclusively part of one stream at a time. It takes up to four seconds to establish a stream and this stream lasts until there is evidence for a more authentic new grouping for several seconds. “This conservatism prevents the system from oscillating widely among perceptions of various numbers of streams in a complex environment.”<sup>41</sup> Sudden happenings, distractions or change of attention or concentration can reset the scene analysis. In vision, the principle of belongingness can easily be seen in the example of the vase-face as illustrated in fig. 2.15. Here, the contour can either be seen as forming a vase or two faces; it cannot be considered as belonging to both objects at the same time. Additional cues can force one specific grouping and impair another.

Although a part of sound belongs to a stream it does not have to be considered as totally different from another stream. “There are levels of perceptual belongingness

---

<sup>40</sup>Bregman (1990), p. 450.

<sup>41</sup>Bregman (1990), p. 130.



**Figure 2.15.:** Illustration of the principle of belongingness. In this picture either a couple of faces or a group of violins can be seen at a time (top). Additional cues can force a specific grouping (bottom).

intermediate between 'the same thing' and 'unrelated things'."<sup>42</sup> As units can form higher order units, they may reveal relationships.

**Retroactive effects:** In the auditory stream segregation process retroactive effects can occur. Two tones starting at the same time may fuse, which means they are integrated into the same stream. But when one of the tones stops earlier, the two get reconsidered as different tones ex post.

### 2.4.2. Primitive Grouping Principles

Primitive grouping principles do not premise attention, knowledge or experience regarding sources of sound. They are typically suggesting grouping patterns based on proximity or similarity of temporal or spectral aspects of sound. The following grouping principles subsume occurring phenomena:

1. Harmonicity
2. Timbre

---

<sup>42</sup>See Bregman (1990), p. 204.

3. Proximity
4. Common Fate
5. Synchrony
6. Continuity
7. Trajectories
8. Closure
9. Spatial Location
10. Comparison with other Senses

**Harmonicity:** If simultaneous tones are harmonics of a common fundamental, they fuse, which means they are likely to integrate into one stream. In an inharmonic series the auditory system makes a “best fit”<sup>43</sup> guess, which results in a less strong integration. Yet, very loud harmonic and inharmonic spectral components can protrude and segregate from a stream.

**Timbre:** A common harmonicity will also group successive sounds into the same stream. Similar timbre, spectral balance or brightness and simplicity of the behavior of the harmonics and the attack support this integration, even when frequency relations of tones in successive sounds change.

**Proximity of a succession of frequencies:** Resembling sounds are grouped. Especially in fast sequences short movements in frequency are preferred for a grouping<sup>44</sup>. Bregman shows this in a listening test in which three high notes 4, 5 and 6 are interlocking with three deep notes 1, 2 and 3. Although the actual sequence is 1-4-2-5-3-6, the “apparent motion”<sup>45</sup> is one 1-2-3- and one 4-5-6-sequence. The actual sequence, which was jumping between high and low notes is segregated into two streams, one with high and one with low notes. Faster sequences and bigger frequency distances between the high and low notes increase the grouping strength.

---

<sup>43</sup>Bregman (1990), p. 236.

<sup>44</sup>Referred to as “Körte’s law”, see Bregman (1990), p. 22.

<sup>45</sup>Bregman (1990), p. 21.

Accurate judgments about the order of notes is only possible for notes within one stream. This experiment even works with missing fundamentals and when every note is presented randomly to one ear only, as long as the other ear is simultaneously stimulated e. g. by noise. Spectral edges, spectral balance and frequency proximity play a central role for this grouping<sup>46</sup>. Also can pure tones integrate with complex sounds if it is similar to one of the partials.

**Common fate:** If different parts of sound change in the same way, with a common ratio, they are integrated in the same stream, especially concerning frequency- and amplitude-modulations, even micromodulations and periodicity of beats. On the other hand, changing the frequency, amplitude or phase of one partial only, will segregate it from the stream. Also, echo suppression can be explained by common fate. Echoes are a slightly changed repetition of the direct sound and therefore integrate into the same stream, as long as their temporal distance is not too big.

**Synchrony:** Synchrony of tones, especially synchronous onset, leads to an integration into one stream, particularly if attack and decay of higher harmonics and the corresponding degree of spectral fluctuations coincide. This fusion happens with harmonic and inharmonic sounds. Synchronous changes in time of frequency, amplitude or spatial direction impose an integration.

**Continuity:** A continuous, smooth change leads to a better grouping than sudden changes. A tone changing pitch over time (gliding) is likely to be integrated into one stream, whereas a sudden change of pitch will be perceived as two successive tones, which can support a segregation. In vision this is referred to as good continuation which is demonstrated in fig. 2.16. In version 1, the irregular figure is perceived as unitary form due to the smooth continuation, despite the fact that it is separated by a line which seems to belong to the white rectangle. In version 2, the shape seems discontinuous through a shift of the part on the right hand side of the vertical line. In version 3, in addition to a vertical shift, the slope of the right part of the shape is differed. This further dissociated the two halves.

---

<sup>46</sup>See Bregman (1990), p. 76, 90, and 93.



**Figure 2.16.:** Illustration of the principle of good continuation by three slightly changed versions of the same figure.

**Trajectories:** Crossing lines in vision are not perceived as equal angles being tangent to each other. The contrary holds in auditory perception. Sequences crossing in pitch are perceived as sequences which converge, meet, and diverge again. “When a falling sequence of tones crossed a rising sequence, the listeners tended not to be able to follow either sequence across the crossing point.”<sup>47</sup>

**Closure:** Masked sounds seem to continue, even if they are physically not present. A repeated note with short silence between the notes is heard as a sequence. But when the gaps between the notes are filled with a masking sound, the repeated note is perceived as one continuous sound, interrupted by the masking signal. We continue to follow a stream even when it is masked or over. Fig. 2.17 illustrates this principle in vision and hearing. The lines with periodic incline and decline, interrupted by gaps, are perceived as distinct angles, though one can think of it as being connected. Filling up the gaps with stripes leads to the impression of a continuous zigzag line, being partly covered by the stripes. The same counts for sound. Tones with periodically ascending and descending pitch glides, interrupted by silence, are perceived as single, though related, events. Filling up the silent parts with masking noise leads to the perception of one continuous tone, occasionally masked by the noise.

**Spatial location:** Sounds that are perceived to emerge from the same spatial origin tend to be grouped to one auditory stream. Each frequency band is treated separately, therefore sound from each band must be localized according to the localization mechanisms explained in ch. 2.3.

<sup>47</sup>See Bregman (1990), p. 447.



**Figure 2.17.:** Illustration of the principle of closure in vision and hearing. A tone, systematically gliding in pitch, interrupted by silence, is represented by an interrupted zigzag line. When the silence is filled up with noise (bars), the pitch-gliding tone seems to be continuous, as seems the zigzag line. After Bregman (1990), p. 28.

**Other senses:** Comparison with other senses, such as vision, can influence grouping. We believe a sound is coming from the source suggested by the visual perception, e.g. by similar temporal patterns, especially change in intensity of motion or a corresponding alteration of vertical position and pitch. This principle is known as “ventriloquism effect”<sup>48</sup>. Typically, vision is even slightly dominant over hearing. Sense of touch can also influence grouping. Wind from behind, together with a tumbling sound of a wooden wall indicates that there was a physical happening in the rear.

### 2.4.3. Schema-Based Grouping Principles

Already infants between 1.5 and 3.5 months show evidence of auditory stream segregation<sup>49</sup>. But a six-month old child can locate sound sources only with an accuracy of approximately 15°, although the physiological development of spectral resolution is completed. Furthermore, they need larger cues<sup>50</sup>. From that one can assume that further improvements in localization are based on experience which may also be the case for other organization cues.

<sup>48</sup>See e.g. Bregman (1990), p. 183 or Schmidhuber et al. (2011).

<sup>49</sup>See e.g. Bregman (1990), p. 405.

<sup>50</sup>See Werner (2012), pp. 4ff.

Learned patterns, like diatonic sequences can lead to auditory stream segregation. E.g. a non-diatonic note in a diatonic sequence “pops out”, in other words segregates, since it does not fit into a learned pattern<sup>51</sup>. Intention of a listener can prefer a desired way of grouping within certain limits. For example hearing a sequence of tones as one or as two separate streams can be chosen by will, as long as the tempo is not too high and intervals are not too small for a segregation or too big for an integration. A similar phenomenon is the ability to concentrate on certain aspects of sound, like hearing out a particular tone or instrument, and therefore reorganize the auditory scene where necessary. It is easier to segregate a part of a sequence against the grouping, forced by the primitive principles, than to integrate something that would be segregated by primitive grouping.

#### **2.4.4. Organization Based on Auditory Scene Analysis Principles**

The more principles suggest the same way of grouping the stronger the grouping gets. In some cases the different principles will lead to a particular scene analysis with distinct integrations into single auditory streams and a clear segregation between them. According to the psychologist Garner, “[. . .] pattern goodness is correlated with pattern redundancy.”<sup>52</sup> However, in many cases the grouping resulting from one or more principles will conflict with the grouping gathered from others. In these cases principles can predominate others, forcing their particular organization preference, as already illustrated in the vase-face figure 2.15. Sometimes this leads to insecurity about the grouping. In worst case conflicting principles may even lead to a total confusion. Bregman speaks of a “competition among alternative organizations”<sup>53</sup>.

Although grouping principles are based on certain parameters of the sound, an auditory stream can have one group-value for a particular parameter which differs from the values of the single components of the group. E.g. when sounds from

---

<sup>51</sup>See Bregman (1990), p. 136.

<sup>52</sup>See Garner (1974), p. 10.

<sup>53</sup>See Bregman (1990), p. 165.



different locations are integrated into one stream due to dominance of other principles — like harmonicity, timbre, common fate, synchrony etc. — they are likely to obtain one common group location. “The auditory system seems to want to hear all the parts of one sound as coming from the same location, and so when other cues favor the fusion of components, discrepant locations for these components may be ignored. It is as if the auditory system wanted to tell a nice, consistent story about the sound.”<sup>54</sup>

An auditory stream can obtain qualities which the single elements within the stream do not have. According to von Ehrenfels, the whole can be more than the sum<sup>55</sup>. Furthermore, relations are clearer between elements within one stream than between those of two different streams, e.g. intervals between notes of a chord from one instrument are easier to identify than between instruments with different timbres and locations. Also, dissonance between frequencies of the same stream are perceived much stronger than between frequencies of different streams. It is harder to tap or count along with a metronome if the clicks fall into different streams, e.g. the location of the click sound alters or the single clicks strongly differ in spectrum. Often a perception of temporal overlap between elements of different streams arises for up to 50ms of silence between them. This shows the uncertainty concerning a comparison between elements from different streams. It is hard to hear out a melody from a musical piece when the single notes are elements of different streams or when they are integrated into one stream together with distracting other notes. Although all notes of the melody are physically heard, it is almost impossible to recognize it in this case. Bregman calls this phenomenon “sequential blanking” or “recognition masking”<sup>56</sup>.

On the other hand, sounds integrated in a stream can “[...] lose some of their original salience when they are incorporated into a larger unit.”<sup>57</sup> For example it is not easy to distinguish all partials of a complex sound or all notes of a 4- or 5-note chord. In a fast sequence of four notes, integrated into one stream, subjects were not able to tell the order of the second and third note. “[...] [T]he phrasing

---

<sup>54</sup>See Bregman (1990), pp. 305f.

<sup>55</sup>This gestalt quality is known as “Übersummenhaftigkeit”, see von Ehrenfels (1890), pp. 249ff.

<sup>56</sup>See Bregman (1990), p. 172.

<sup>57</sup>See Bregman (1990), p. 140.

of a sequence of notes makes some of their pitches less distinct.”<sup>58</sup> The localization of a source is based on grouping those components of complex, interfering sounds, which can be associated to the same auditory event<sup>59</sup>. Consequently, a subjectively secure localization of a source can suppress the perception of timbre, which explains the inaudibility of the comb filter effect in stereo playback<sup>60</sup>. Despite this loss for details, a formation of separate streams allows for a comparison of global properties.

As grouping principles can conflict, there are also cases in which principles concerning the properties of auditory streams are violated. E.g. when prime and fifth of a triad are presented to one ear, while the third is presented to the other, many listeners perceive a full chord in one ear and the single note in the other<sup>61</sup>. A paradox in grouping is that timbre is an important parameter for sequential grouping, even though timbre is already a result of spectral grouping.

If a part of sound cannot be integrated into a stream it is likely to be neglected for further perception; it will be overheard. This natural selection is necessary to reduce the masses of information from the environment to an amount we can handle, to avoid a sensory overload. This reduction of information may be misinterpreted as a deficiency. But it is auditory scene analysis which provides us with reliable information about the acoustical outside world. It is the basis of our understanding of the auditory world as interpreted from all superimposed acoustical signals that confront us. Computational auditory scene analysis has arisen from the idea to compute this impressive capability by imitating mechanisms of the auditory system. This approach might have the potential to assign parts of sound to their physical happening, thus to identify musical instruments from an orchestral recording or recognize speech in a noisy environment<sup>62</sup>.

---

<sup>58</sup>See Bregman (1990), p. 475.

<sup>59</sup>See Theile (1980), especially p. 24.

<sup>60</sup>See Ono et al. (2002), p. 2 and Theile (1980), p. 12.

<sup>61</sup>See Bregman (1990), p. 693.

<sup>62</sup>Concepts, algorithms and the state of research are extensively illuminated in Rosenthal and Okuno (1998) and Wang and Brown (2006).

## 2.5. Usability of Psychoacoustic Knowledge for Audio Systems

The physical nature of the sound in our environment is highly complex. The auditory system supplies numerous mechanisms to adequately transfer the physical signals into psychologically interpretable representations. Not all aspects of physical sound stimuli actually contribute to this mental representation. Some magnitudes lie below thresholds, some changes below just noticeable differences. Thus, absolute thresholds and masking thresholds determine the audible frequency- and level region to be reproduced by an audio system. Just noticeable differences as well as integration times tell about the necessary spatial and temporal precision for the reproduction of amplitude, amplitude change, phase change, source direction etc.

Hence, psychoacoustic knowledge can be used to reduce the amount of data to be recorded, processed, and reproduced without audible effects. This is commonly done in application. For example microphones for musical recordings only record the audible frequency range. All digital systems make use of the temporal and dynamic resolution capacity of the ears by sampling continuous sound into time-discrete values with a certain accuracy in relative sound pressure. The audio-cd reduces continuous sound pressure changes to 44100 discrete sound pressure states per second and codes the dynamic range with 16 bit which allows  $2^{16}$  possible values. Bader discusses the approach to efficiently code sounds by using gammatones, imitating the nerve cell output of the auditory system<sup>63</sup>. Masking can be considered as a threshold increase caused by a loud sound. Experiments on simultaneous masking led to quite accurate and valid results, concerning the relationship between masker frequency and amplitude and the resulting masking pattern. Temporal masking shows bigger variance and less reliability. Furthermore, interaction between maskers in different frequency regions as well as between masker and maskee has been observed but not fully understood. The existence of binaural masking is an evidence that another masking mechanism exists besides the masking originated in cochlear processes. Of course, pure tones, critical band wide noise, white noise and Gaussian sound bursts are not the kind of sound typically faced in a

---

<sup>63</sup>See Bader (2013), pp. 632ff.

natural listening situation, in communication or musical performance. The same applies for pure monotic or dichotic performance. The masking effects occurring in natural listening contexts may be some sort of mixture of both monaural and binaural masking. Still, investigations of masking led to an understanding of the phenomenon — temporally and spectrally — which gave rise to psychoacoustic audio compression methods quantified in technical applications such as AAC, AC-3 and MP3<sup>64</sup>. It is also qualified for an implementation in an analysis and synthesis system for musical instruments as will be discussed in ch. 5.3 and 6.

Conventional audio systems are mainly based on psychoacoustic methods to recreate a natural auditory impression rather than aiming at recreating all physical quantities. Spatial localization of sound sources can be accomplished by the auditory system with a high precision especially concerning the horizontal angle in the frontal directions. Due to this fact, early stereophonic audio systems concentrated on sound playback in this region. Further developments added sounds from the rear directions but rarely involved distance or the third dimension since distance hearing and localization capability at the sides and in the median plane are weak, anyway. Auditory streams obtain a group value, therefore it can be sufficient to reproduce only some of the acoustical properties to maintain the original auditory scene. The theory of summing localization is used in stereo systems to create the impression of one sound source at any position between two loudspeakers by playing systematically manipulated signals via two loudspeakers.

The theory of wave field synthesis is discussed in ch. 4. It aims at physically recreating all sound properties, as discussed in ch. 3, in a large listening area. At first glance, wave field synthesis seems to make applications of psychoacoustic methods superfluous. But when it comes to actual implementation, psychoacoustic considerations are essential as will become clear in ch. 4.2. Many researchers predict that the future of audio systems lies in psychoacoustics<sup>65</sup>.

---

<sup>64</sup>See e.g. Lerch (2008), pp. 872ff. Extensive discussion about MP3 can be found in Ruckert (2005).

<sup>65</sup>See e.g. Blauert (2008) and Fastl (2010).

## 3. The Spatial Sound of Musical Instruments

To reach a listener, the sound of musical instruments has to travel, typically through air. Thus, the next section deals with the basic physical principles of sound propagation. This leads to a better understanding of spatial attributes of sound, such as propagation and directivity patterns of musical instruments, which are discussed subsequently. These spatial attributes strongly contribute to the individual sound character of musical instruments. Therefore, many methods have been developed to investigate the radiation characteristics of musical instruments and to represent it in ways that allow for qualitative and quantitative statements. A discussion of these methods completes this chapter.

### 3.1. Wave Equation and Solutions

In this section the physical fundamentals of sound in air are illuminated<sup>1</sup>. They are the basis of acoustics in the free field and describe sound propagation e.g. of musical instruments and loudspeakers.

---

<sup>1</sup>As described in Ziemer (2011), mainly based on Pierce (2007), Williams (1999), Morse and Ingard (1986), Rabenstein et al. (2006) and Ahrens (2012).

### 3.1.1. Homogeneous Wave Equation

Euler's equation of motion

$$\rho_0 \frac{\partial \vec{v}(\vec{x}, t)}{\partial t} = -\nabla p(\vec{x}, t) \quad (3.1)$$

is the first base equation of the wave field. It explains the flow of frictionless fluids by means of time  $t$ , direction vector  $\vec{x}$ , particle velocity vector  $\vec{v}$ , pressure  $p$ , ambient density  $\rho_0$  and nabla operator  $\nabla$ . In Cartesian coordinates the following is valid:

$$\begin{aligned} \vec{x} &= \begin{bmatrix} x \\ y \\ z \end{bmatrix} \\ \vec{v} &= \begin{bmatrix} u(x) \\ v(y) \\ w(z) \end{bmatrix} \\ \nabla &\equiv \frac{\partial}{\partial \vec{x}} = \frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z} \end{aligned} \quad (3.2)$$

The second base equation of the wave field is the continuity equation (conservation of mass)

$$c^2 \rho_0 \nabla \vec{v}(\vec{x}, t) + \frac{\partial p(\vec{x}, t)}{\partial t} = 0. \quad (3.3)$$

with propagation velocity  $c$ . Differentiating eq. 3.3 with respect to time and replacing the velocity term by the right side of the equation of motion, eq. 3.1, yields the homogeneous wave equation for pressure

$$\nabla^2 p(\vec{x}, t) - \frac{1}{c^2} \frac{\partial^2 p(\vec{x}, t)}{\partial t^2} = 0. \quad (3.4)$$

By differentiating the continuity equation with respect to  $\vec{x}$  and the equation of motion with respect to  $t$  yields the homogeneous wave equation for velocity

$$\nabla^2 \vec{v}(\vec{x}, t) - \frac{1}{c^2} \frac{\partial^2 \vec{v}(\vec{x}, t)}{\partial t^2} = 0. \quad (3.5)$$

### 3.1.2. Wave Field

The sound field magnitudes sound pressure  $p$  and sound velocity  $\vec{v}$  are perturbations of the state of equilibrium which propagate as waves.  $c$  is the sound propagation velocity and  $\nabla^2$  is the Laplace operator

$$\nabla^2 \equiv \frac{\partial^2}{\partial^2 \vec{x}} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \quad (3.6)$$

Solutions of the wave equation, eq. 3.4 and eq. 3.5, are called *sound field* or *wave field*. Note, that although these two equations look very similar, the dependent variables  $p$  and  $\vec{v}$  are not equal. Their relationship is described by the equation of motion, eq. 3.1. The equations assume the following conditions:<sup>2</sup>

1. The propagation medium is homogeneous
2. The medium is quiescent and vortex free
3. State changes are adiabatic, i.e. no heat interchange between areas of low pressure and areas of high pressure due to the rapid movement of the particles
4. Pressure and density perturbations are small compared to static pressure and density
5. Relationships in the medium are subject to linear differential equations
6. The medium exhibits no viscosity
7. The medium is source-free

---

<sup>2</sup>See Mechel (2008), pp. 5f, Teutsch (2007), Wöhe (1984), Pierce (2007), p. 36 and Baalman (2008), p. 23.

### 3.1.3. Homogeneous Helmholtz Equation

The pressure term is transformable via Fourier transform

$$P(\vec{x}, \omega) = \int_{t=-\infty}^{\infty} p(\vec{x}, t) e^{i\omega t} dt \quad (3.7)$$

from the time domain to the frequency domain and back via inverse Fourier transform

$$p(\vec{x}, t) = \frac{1}{2\pi} \int_{\omega=-\infty}^{\infty} P(\vec{x}, \omega) e^{-i\omega t} d\omega. \quad (3.8)$$

$e$  is Euler's number ( $e \approx 2.718\dots$ ),  $i = \sqrt{-1}$  is the imaginary unit,  $\omega = 2\pi f$  is the angular frequency and  $f$  the frequency. The wave equation in the frequency domain reads

$$\nabla^2 P(\vec{x}, \omega) + k^2 P(\vec{x}, \omega) = 0 \quad (3.9)$$

with *wave number* or *spatial frequency*  $k = \frac{\omega}{c} = \frac{2\pi}{\lambda}$  and wave length  $\lambda$  and is called *Helmholtz equation*. Since the Fourier transform is an integral over time, the Helmholtz equation is only valid for stationary signals, i.e. periodic vibrations, and not for transients<sup>3</sup>.

### 3.1.4. Plane Waves

A general solution of the wave equation is d'Alembert's solution:

$$p(\vec{x}, t) = f(\vec{x} - ct) + \tilde{f}(\vec{x} + ct) \quad (3.10)$$

The first term describes the propagation of a pressure state in  $\vec{x}$  direction, the second a propagation in the opposite direction. For waves the principle of superposition applies, i.e. they interfere without affecting each other. Assuming the second term to be 0, only one wave in  $\vec{x}$  direction remains. Other directions can simply be

---

<sup>3</sup>See e.g. Meyer et al. (2001), p. 2.



added. One possible solution function  $f(\vec{x} - ct)$  is the function of a plane wave:

$$\begin{aligned} p(\vec{x}, t) &= A(\omega) e^{-i(k\vec{x} - \omega t)}, \text{ or, respectively} \\ P(\vec{x}, \omega) &= A(\omega) e^{ik\vec{x}} \end{aligned} \quad (3.11)$$

Here  $A(\omega)$  is an arbitrary complex amplitude in the form  $\hat{A}e^{i\phi}$  whose absolute value is the amplitude  $\hat{A}$  and whose argument is the phase  $\phi$  of a frequency contained in the signal.  $k^2 = k_x^2 + k_y^2 + k_z^2$  is the squared wave number in direction  $\vec{x}$ ,  $\lambda^2 = \lambda_x^2 + \lambda_y^2 + \lambda_z^2$ , the wave length in  $\vec{x}$ -direction, respectively. A plane wave propagates in direction  $\vec{x}$  whereat phase changes with respect to location.  $k_x$ ,  $k_y$  and  $k_z$  are called “trace wavenumbers”<sup>4</sup>,  $\lambda_x$ ,  $\lambda_y$  and  $\lambda_z$  are *trace wavelengths*. They are projections to the spatial axes. The “wavefronts”<sup>5</sup> are infinite planes of equal pressure perpendicular to vector  $\vec{x}$ .

For a wave with non-negative  $k$ , two formulations for  $k_y$  point out two different sorts of wave<sup>6</sup>:

$$k_y = \begin{cases} \pm \sqrt{k^2 - k_x^2 - k_z^2}, & k^2 \geq k_x^2 + k_z^2 \\ \pm i \sqrt{-k^2 + k_x^2 + k_z^2}, & k^2 \leq k_x^2 + k_z^2 \end{cases} \quad (3.12)$$

In the first case all components are real, indicating a propagating plane wave. In the second case  $k_y$  is imaginary, leading to an *evanescent wave*. Inserting the second case in eq. 3.11 yields:

$$P(\vec{x}, \omega) = A(\omega) e^{\pm \sqrt{-k^2 + k_x^2 + k_z^2} y} e^{i(k_x x + k_z z)} \quad (3.13)$$

In this case the first exponential term is real, indicating an exponential decay in  $y$ -direction<sup>7</sup>. Both types of waves are illustrated in fig. 3.1. Note that in this example the propagation direction of the propagating wave and the evanescent wave are the same.

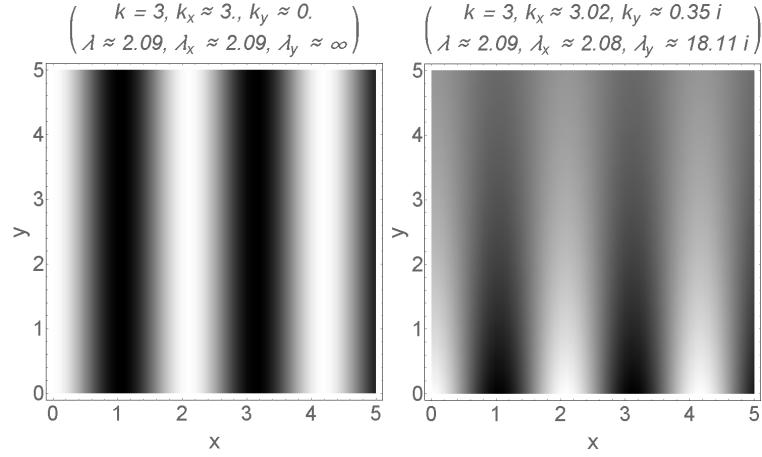
---

<sup>4</sup>See e.g. Williams (1999), p. 21.

<sup>5</sup>See Williams (1999), p. 22.

<sup>6</sup>See Ahrens (2012), p. 23.

<sup>7</sup>Or an exponential increase which is ignored since it is non-physical, see Ahrens (2012), p. 23.



**Figure 3.1.:** Two dimensional visualization of a propagating plane wave (left) and an evanescent wave (right).

For periodic functions the motion equation (eq. 3.1) yields:

$$\nabla p(\vec{x}, t) = -i\omega\rho_0\vec{v}(\vec{x}, t) \quad (3.14)$$

and in the frequency domain

$$\nabla P(\vec{x}, \omega) = -ik\rho_0 c\vec{V}(\vec{x}, \omega), \quad (3.15)$$

where  $\vec{V}$  is the sound velocity in frequency domain.

### 3.1.5. Inhomogeneous Wave Equation

The homogeneous wave equation assumes a source free medium. But every sound field has at least one source which adds acoustic energy to the medium, propagating as a wave pursuant to the wave equation. To account for this, the eighth condition listed in ch. 3.1.2 is dropped and a source term is added to the homogeneous wave equation. Then a solution  $p(\vec{x}, t)$  is sought describing the temporal and spatial behavior of the source signal in the system.

$$\nabla^2 p(\vec{x}, t) - \frac{1}{c^2} \frac{\partial^2 p(\vec{x}, t)}{\partial t^2} = -4\pi\delta(\vec{x} - \vec{x}_0, t - t_0) \quad (3.16)$$

This wave equation is called *inhomogeneous wave equation*.  $\delta(\vec{x}, t)$  is the Dirac delta function, an impulse. It is defined as being  $\infty$  at point  $\vec{x}_0$  at time  $t_0$ , otherwise it is 0. A transformation of the Dirac delta function into the spectral domain

$$\delta(\omega) = \int_{t=-\infty}^{\infty} \delta(t - t_0) e^{i\omega t} dt = 1 \quad (3.17)$$

shows that its amplitude for every frequency is 1, i.e. all frequencies have an equal amplitude and are in phase. That means every arbitrary function  $p(t)$  can be expressed by weighted and delayed Dirac delta functions  $\delta(\vec{x}, t)$ . Amplitude and phase of spectral components  $P(\omega)$  of sound signals may be arbitrary so they can be expressed as multiplication of the spectra of the Dirac delta function by frequency-dependent complex amplitudes  $A(\omega)$ . That conforms a convolution of a sound signal with the Dirac delta function in the time domain.

### 3.1.6. Point Sources

One solution for the inhomogeneous wave equation is the point source. A point source is a sound source with no volume. In the easiest case, its radiation is equal in each direction. This is referred to as a *monopole source* or *zero order radiator*. Amplitude and phase are dependent on frequency and distance but independent of direction. Therefore, a formulation in spherical coordinates is meaningful. For spherical coordinates the following holds:

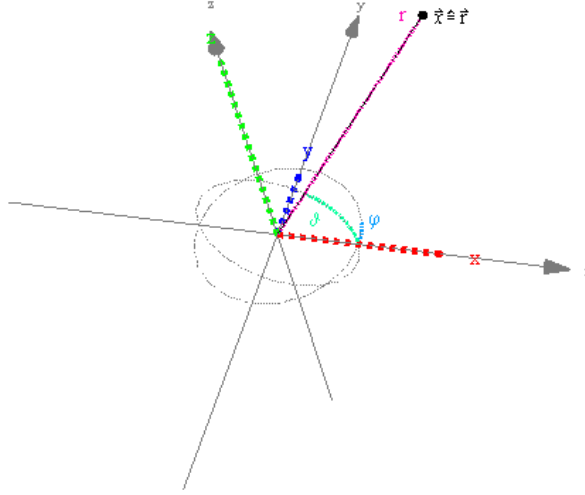
$$\begin{aligned}
\vec{r} &= \begin{bmatrix} r \\ \varphi \\ \vartheta \end{bmatrix} \\
r &= \sqrt{x^2 + y^2 + z^2} \\
\varphi &= \arctan\left(\frac{y}{x}\right) \\
\vartheta &= \arccos\left(\frac{z}{r}\right) \\
\nabla_{\text{spherical}} &\equiv \frac{\partial}{\partial r} + \frac{1}{r} \frac{\partial}{\partial \vartheta} + \frac{1}{r \sin \vartheta} \frac{\partial}{\partial \varphi}
\end{aligned} \tag{3.18}$$

$$\begin{aligned}
\vec{x} &= \begin{bmatrix} x \\ y \\ z \end{bmatrix} \\
x &= r \cos \varphi \cos \vartheta \\
y &= r \sin \varphi \cos \vartheta \\
z &= r \sin \vartheta \\
\nabla_{\text{Cartesian}} &\equiv \frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z}
\end{aligned}$$

With radius  $r$ , azimuth angle  $\varphi$  and polar angle  $\vartheta$ . Respectively, the position vector  $\vec{x}$  is redefined to  $\vec{r}$ . In principle, this spherical coordinate system is consistent with the head related spherical coordinates used in ch. 2.3 for describing directional hearing. But in this case the coordinate origin is not the listener's head but the source position. Fig. 3.2 illustrates the relations of Cartesian and spherical coordinate systems.

Thus, the inhomogeneous wave equation (3.16) takes the form

$$\frac{1}{r} \frac{\partial \left( r^2 \frac{\partial p}{\partial r} \right)}{\partial r} + \frac{1}{r^2 \sin \vartheta} \frac{\partial \left( \sin \vartheta \frac{\partial p}{\partial \vartheta} \right)}{\partial \vartheta} + \frac{1}{r^2 \sin^2 \vartheta} \frac{\partial^2 p}{\partial \varphi^2} - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = -4\pi \delta(\vec{x} - \vec{x}_0, t - t_0). \tag{3.19}$$



**Figure 3.2.:** Representation of the position vector  $\vec{x}$  or, respectively  $\vec{r}$  via Cartesian coordinates and spherical coordinates.

Since radiation of a monopole is independent of  $\varphi$  and  $\vartheta$  the wave equation simplifies to

$$\frac{\partial^2 p(r, t)}{\partial r^2} + \frac{2}{r} \frac{\partial p(r, t)}{\partial r} - \frac{1}{c^2} \frac{\partial^2 p(r, t)}{\partial t^2} = -4\pi \delta(r - r_0, t - t_0), \quad (3.20)$$

and the Helmholtz equation appropriately to

$$\frac{\partial^2 P(r, t)}{\partial r^2} + \frac{2}{r} \frac{\partial P(r, t)}{\partial r} - k^2 P(r, t) = -4\pi \delta(r - r_0, \omega). \quad (3.21)$$

The point source solution for this case is

$$\begin{aligned} p(r, t) &= g(r, t) + \tilde{g}(r, t) \\ &= A(t) \frac{e^{-i(kr - \omega t)}}{r} + \tilde{g}(r, t), \text{ or, respectively} \\ P(r, \omega) &= G(r, \omega) + \tilde{G}(r, \omega) \\ &= A(\omega) \frac{e^{-ikr}}{r} + \tilde{G}(r, \omega). \end{aligned} \quad (3.22)$$

It is a *Green's function* comprised of a linear combination of a special solution —  $g(r, t)$ , or  $G(r, \omega)$ , respectively — and a general solution —  $\tilde{g}(r, t)$ , or  $\tilde{G}(r, \omega)$  —

which are arbitrary solutions of the homogeneous wave equation (eq. 3.4) and Helmholtz equation (eq. 3.9). It is also called “impulse response” in the time domain and “complex transfer function” in the frequency domain<sup>8</sup>. Since the first term of the impulse response is already a complete solution of the inhomogeneous Helmholtz equation, the second term can be assumed to be zero. This case is called *free field Green’s function* and describes the radiation of a monopole sound source. The exponential term describes the phase shift per distance of the propagating wave from the source. The fraction represents the amplitude decay per distance, the so-called *inverse distance law* or *1/r distance law*<sup>9</sup>, which is owed to the fact that the surface of the wave front increases with an increasing sphere radius, so the pressure distributes on a growing area.

The surface of a sphere  $S$  is given as

$$S = 4\pi r^2 \quad (3.23)$$

so the sound intensity  $I_0$  in the origin of the point source at  $r = 0$  spreads out to the surface with  $I(r) = I_0 \frac{1}{4\pi r^2}$  and is thus directly proportional to  $\frac{1}{r^2}$ . Since  $I$  is proportional to  $p^2$ ,  $p(r)$  it is directly proportional to  $\frac{1}{r}$ <sup>10</sup>:

$$\begin{aligned} I(r) &\propto \frac{1}{r^2} \\ p(r) &\propto \frac{1}{r} \end{aligned} \quad (3.24)$$

The wave front of a propagating plane wave, in contrast, is assumed to be infinite and thus does not decay. In the far field — i.e. ignoring near field effects which show a complicated behavior close to the source — any stationary sound source can be simplified by considering it as point source<sup>11</sup>. These point sources, however,

---

<sup>8</sup>See e.g. Müller (2008), p. 65.

<sup>9</sup>See e.g. Vorländer (2008).

<sup>10</sup>See e.g. Roederer (2008) pp. 89f.

<sup>11</sup>See Ahrens (2012), p. 42.

do not necessarily have to be monopoles. A dependence on direction  $\Psi(\omega, \varphi, \vartheta)$  can be introduced ex post by reconsidering  $A(\omega)$  as  $A(\varphi, \vartheta, t)$  or, respectively,  $\Psi(\omega, \varphi, \vartheta) A(\omega)$  for the far field:

$$\begin{aligned}
 p(\varphi, \vartheta, r, t) &= g(\varphi, \vartheta, r, t) + \tilde{g}(\varphi, \vartheta, r, t) \\
 &= \Psi(\omega, \varphi, \vartheta) A(\omega) \frac{e^{-i(kr - \omega t)}}{r} + \tilde{g}(r, t), \text{ or, respectively} \\
 P(\varphi, \vartheta, r, \omega) &= G(\varphi, \vartheta, r, \omega) + \tilde{G}(\varphi, \vartheta, r, \omega) \\
 &= \Psi(\omega, \varphi, \vartheta) A(\omega) \frac{e^{-ikr}}{r} + \tilde{G}(r, \omega).
 \end{aligned} \tag{3.25}$$

Due to the complex factor  $\Psi(\omega, \varphi, \vartheta)$ , the amplitude  $A(\omega)$  is modified for any direction. Note, that the Green's function with a direction-dependent radiation factor is not a solution to the inhomogeneous Helmholtz function as such<sup>12</sup>. It rather comprises the spherical harmonics, which are a solution to the angular dependencies of the Helmholtz equation in spherical coordinates over a sphere rather than a point. The radiation characteristic of point sources can be any arbitrary function of angles  $\varphi$  and  $\vartheta$ , which can be composed by a linear combination of mono- and multipoles. In the literature, point sources with a direction-dependent radiation factor are called “multipole point sources”, “higher mode radiators” or “point multipoles”, the directivity is called “far-field signature function”<sup>13</sup>

## 3.2. The Spatial Sound of Musical Instruments

The sound of musical instruments contains a lot of spectral, temporal and spatial features;

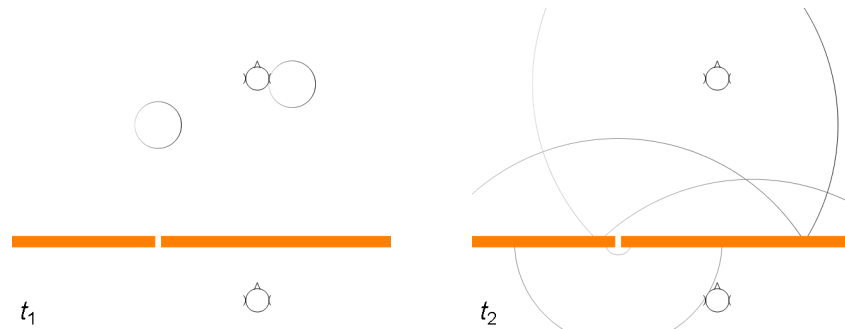
”[...] the main acoustic features of musical instruments include:

<sup>12</sup>Cf. e.g. Ahrens (2012), p. 66.

<sup>13</sup>See e.g. Mechel (2013), p. 2, Magalhães and Tenenbaum (2004), p. 204, Ahrens (2012), p. 42.

- musical scale,
- dynamics,
- timbre of sound,
- time envelope of the sound,
- sound radiation characteristics.”<sup>14</sup>

The first four features are easily recordable via microphone and can be played back in a good approximation by any High-Fidelity (Hi-Fi) loudspeaker. Still, a listener is often able to distinguish simply recorded and played-back sound from the original instrumental sound. “Composers and musicians often complain about the way loudspeakers sound when aiming at reproducing or amplifying signals from musical instruments.”<sup>15</sup> The reason for this is the so-called “Mauerlocheffekt”<sup>16</sup>. It effects that a monophonic playback sounds like a single-slit diffraction, i.e. like hearing a concert through a keyhole<sup>17</sup>. Independent of directivity or incidence angle of sounds the further sound radiation will be the same for all wavelengths larger than the slit: From the slit on, as for monophonic sound, the feature “sound radiation characteristics” is missing.



**Figure 3.3.:** Demonstration of the Mauerlocheffekt. Inside the room, source position and radiation patterns can be heard (left). Behind the slit all sources sound like a monopole located at the slit (right).

<sup>14</sup>From Kostek (2005), p. 24.

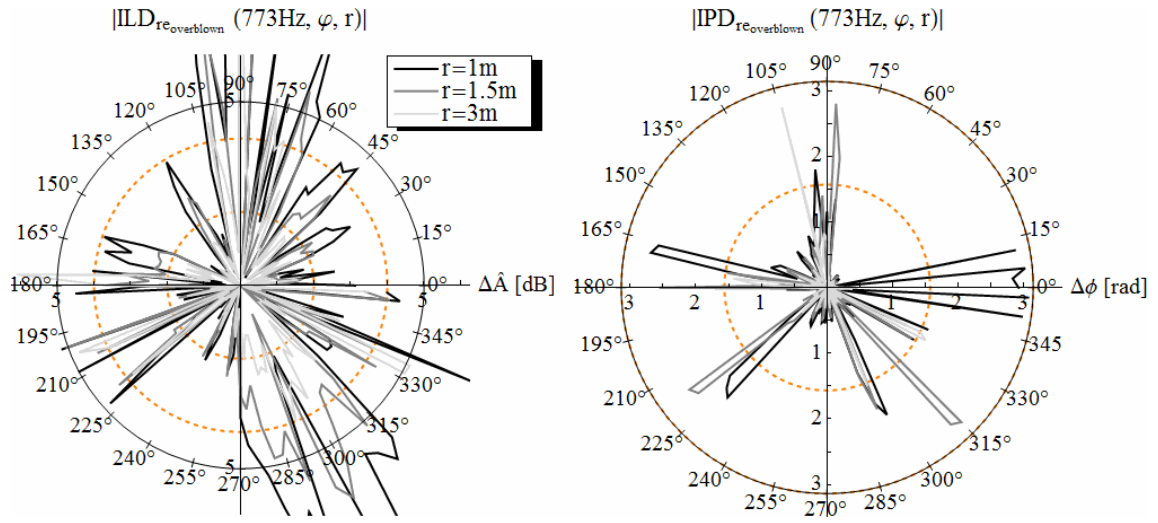
<sup>15</sup>From Warusfel et al. (1997), p. 1.

<sup>16</sup>See Schanz (1966), p. 2.

<sup>17</sup>Cf. Rossing (1990), p. 48.



The complicated radiation characteristics of musical instruments, especially for higher frequencies, create slightly different arrival times of wavefronts and relative amplitude and phase differences between each direction which lead to the impression of a natural, broad, spatial source for a listener. It is an important characteristic of instrumental sound, sometimes referred to as “directional tonal color”<sup>18</sup>. Even small instruments with simple geometries — such as the shakuhachi — create interaural sound differences for listeners which decrease as distance increases. ILD and IPD of a shakuhachi partial are illustrated in fig. 3.4 for different listening positions.<sup>19</sup>



**Figure 3.4.:** Interaural level differences (left) and interaural phase differences (right) of one shakuhachu partial for listeners at different listening angles and distances.

Without recreating this spatial aspect listeners will be able to distinguish between original instrumental sound and loudspeaker playback. “But only few systems incorporate the directivity characteristic of natural sources.”<sup>20</sup> How this radiation characteristic occurs and how it is perceived by a listener is described in the following.

A body radiating with its whole surface in phase, as a so-called *breathing sphere*, radiates as a monopole. This approximately accounts for all wave lengths larger

<sup>18</sup>See Fletcher and Rossing (2008), p. 308.

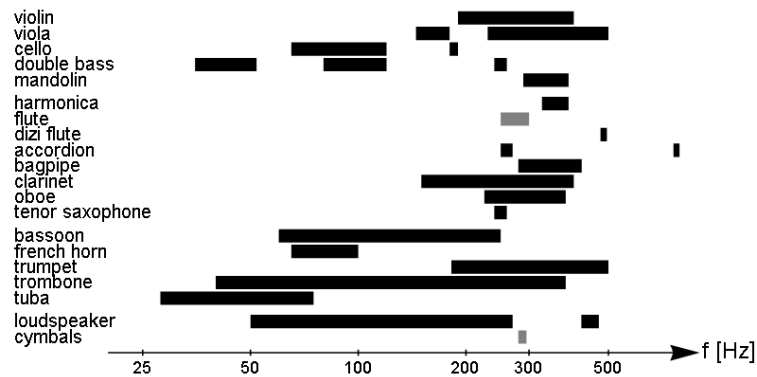
<sup>19</sup>An examination of the relationship between features of direct sound and perceived source extent can be found e.g. in Ziemer (2014).

<sup>20</sup>Albrecht et al. (2005), p. 1.

than the radiating body. For musical instruments this is often roughly the case for frequencies below 500Hz. Small acoustic sources compared to the radiated wave lengths can be considered as *point sources* whose wave front is a sphere increasing concentric around the source.

The amplitude on this wave front may be dependent on direction. This direction-dependency varies with frequency and is caused by interfering sound radiation from different areas on the body (modes), issues from different apertures (i.e. finger holes or f-holes)<sup>21</sup>, directive radiation e.g. from the bell of a trumpet or from diffraction and acoustic shadow behind instrument and instrumentalist<sup>22</sup>. The directional characteristic of a frequency is typically independent of the dynamic but sometimes dependent on the played keynote, especially in the family of string- and woodwind instruments<sup>23</sup>.

An overview of the sound radiation of musical instruments is given by Jürgen Meyer and Pätynen and Lokki<sup>24</sup>.



**Figure 3.5.:** Frequency regions with approximately monopole- (black) and dipole-shaped (gray) sound radiation of musical instruments. Data from Meyer (2009), p. 130 and from my own measurements.

Figure 3.5 illustrates frequency regions in which musical instruments show an omnidirectional radiation characteristic or a dipole-like radiation. Other frequency

<sup>21</sup>Referred to as “structure- and air-borne sound”, see e.g. Blauert and Xiang (2009), p. 177.

<sup>22</sup>See Hall (2008), pp. 290–294.

<sup>23</sup>See Meyer (2008), p. 156, Warusfel et al. (1997), p. 4, Pätynen and Lokki (2010) and Otondo and Rindel (2005).

<sup>24</sup>In Meyer (2009), pp. 129–177 and Meyer (2008), pp. 123–180 and in Pätynen and Lokki (2010).

regions provide different spectra per direction. This means that the listening impression is dependent on the position of listeners and on movements of instrumentalists. This is especially the case in close proximity to the source. Thus, not only the audience but particularly the instrumentalist experiences spectral changes when moving relative to the instrument. A pianist for example is exposed to complicated interferences which may strongly vary with head movements. This is a natural experience which is typically not reproduced in electric pianos. This lack of spatial interference can make them sound static and boring.

The instrumental sound can be divided into phases with different spectral and temporal behavior, and possibly different sound radiation characteristics<sup>25</sup>:

The *transient phase* offers a dense, broad spectrum generated by the main mechanisms of sound production. For example the impulse of a hammer on a piano string, the irregular sticking and sliding of a bow and cello string or the wind burst of a trumpeter excite an amount of frequencies. The transient phase is additionally characterized by side noise like the grasping sound of a guitar, the clicking of trumpet valves or the quick inhaling of singers. For classical musical instruments the transients last for about 30ms<sup>26</sup>. The duration depends on the instrument, pitch and playing technique. The transient sound plays an important role for the perception of timbre and the recognition of musical instruments.

The *quasi-stationary phase* is almost periodic. It contains the eigen frequencies of the instrument which established while other frequencies lost their energy e.g. by radiation, destructive interference of standing waves or energy transmission between modes. The long lasting steady sound of an organ or a viola are examples of a quasi-stationary phase which can also be damped as in case of a piano string.

---

<sup>25</sup>See Meyer (2009), p. 24, Hammond and White (2008), pp. 4–7 and Hall (2008), pp. 124–125.

<sup>26</sup>See Bruhn (2002c), p. 452.

### 3.3. Measurement of the Radiation Characteristics of Musical Instruments

As described previously in this chapter, the sound of musical instruments can radiate from their surface or containing air. The whole body can vibrate like a breathing sphere, or only parts of it, which may lead to complicated interferences and near field effects as well as complex radiation patterns in the far field. It is difficult to measure the vibrations of body and air without affecting the observed system. Therefore, measurements are done with a microphone array in the far field of the instruments in a free field room.

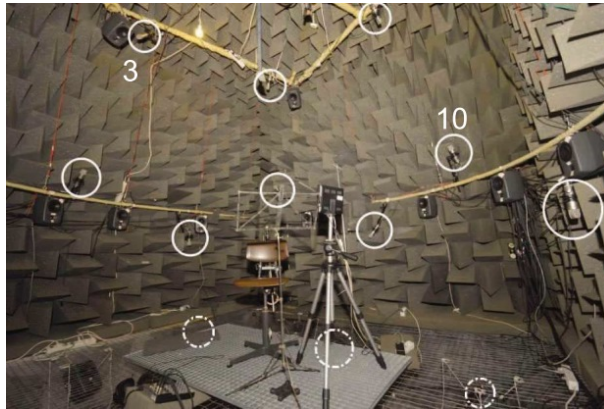
#### 3.3.1. Far Field Recordings

The radiation characteristics of a musical instrument can be measured by simultaneous far field recordings. The far field is typically defined as  $kr \gg 1$  or, respectively  $r \gg \frac{\lambda}{2\pi}$ .

From a distance greater than the dimensions of the musical instrument it is valid to consider the instrument a complex point source or, respectively, a spherical source with infinitesimal volume. The measured wave field is assumed to originate in solely this point. This simplification only holds for big wave lengths compared to the dimensions of the source and is an oversimplification for small wavelengths. Furthermore, it is only valid in the far field of the source and does not inform about near field effects. Thus, the radiation characteristics of the point source can be calculated back from far field recordings. Choosing a meaningful position of the virtual point source in, on or very close to the actual body of the instrument is crucial for a reliable description. For circularly- or spherically-shaped instruments the position of choice might be the center, obviously. However, there is typically no single position which can be considered the “acoustical center” of the radiating sound<sup>27</sup>. There will hardly be a plausible argumentation to pick e.g. geometric

<sup>27</sup>See e.g. Pätynen and Lokki (2010), p. 139.

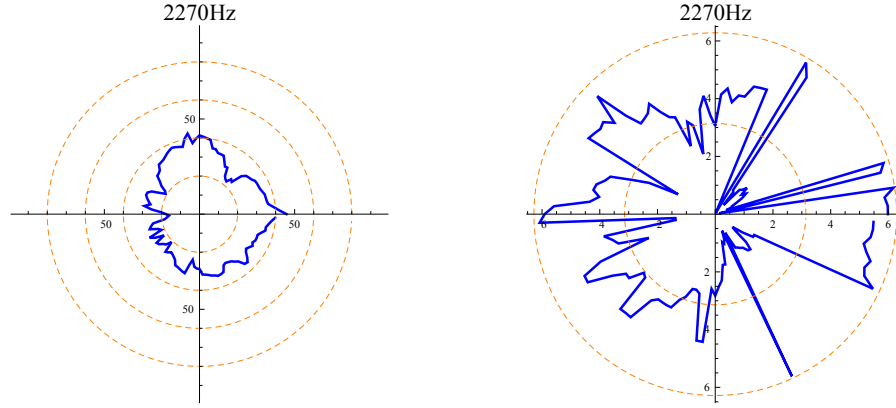
centroid or center of gravity of the instrument's mass as acoustic center. One has to find a center position that fits the specific situation or intention. Then, microphones are arranged equidistantly around this center position, i.e. circularly or spherically. Fig. 3.6 is a photo of a microphone array in a free field room for measuring the radiation characteristics of musical instruments. It is a spherical arrangement consisting of four groups of five circularly arranged microphones plus two additional microphones in front and above the investigated instrument.



**Figure 3.6.:** Photo of a microphone array for far field recordings of musical instruments. From Pätynen and Lokki (2010), p. 140, with friendly permission of Deutscher Apotheker Verlag.

Assuming the source to be a point rather than an area or a volume, the measured relative complex pressure at a microphone position represents not only the pressure at that very position but it can be regarded the pressure factor for that angle. Pressure and phase per direction of one frequency of a shakuhachi tone are illustrated in fig. 3.7. These may or may not be interpolated to approximate factors for the angles in between the measurement angles. Except a Fourier transform (eq. 3.7), no calculation needs to be done.

The accuracy of this simple method can be increased by increasing the number of microphones. Complex factors for angles in between the measurement angles do not have to be approximated by interpolation but may as well be derived from spherical harmonic decomposition, as will be subsequently discussed. Often, the measured directional factors are not taken from single frequencies but are mean



**Figure 3.7.:** Polar far field radiation pattern of amplitude (left) and phase (right) of one shakuhachi frequency measured at a distance of 1m with 128 microphones, linearly interpolated. Note, that the phase is periodic, i.e.  $\phi(2\pi) = \phi(0)$ .

values of several frequencies within octave- or third octave bands<sup>28</sup>. In this thesis, the far field signature function is used to implement natural radiation patterns in a wave field synthesis system.

<sup>28</sup>See e.g. Otondo and Rindel (2004) p. 1179 or Otondo and Rindel (2005), p. 903, Pelzer et al. (2012), Pätynen and Lokki (2010) and Zotter et al. (2007).

## 4. Wave Field Synthesis

The idea of wave field synthesis is not to create a sound which is only psychoacoustically equivalent to a natural sound at one specific listening position but to create a sound which is physically equal to a natural sound or any desired wave field. Its theoretical core is derived from several mathematical theorems and physical considerations. The derivation is explained step by step in this chapter<sup>1</sup>. Several constraints make it applicable but lead to synthesis errors, which are diminishable by propositions from the literature. These are addressed in ch. 4.2.

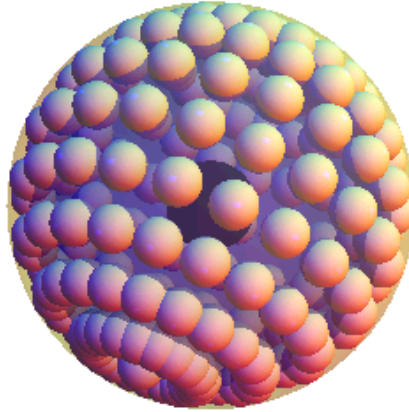
### 4.1. Theoretical Fundamentals of Wave Field Synthesis

#### 4.1.1. Huygens' Principle

Every arbitrary radiation from a sound source can be described as integral of point sources on its surface. In addition, each point on a wave front can be considered as origin of an elementary wave. Together they form the advanced wave front via superposition. This finding is called *Huygens' principle* and is the fundament on which wave field synthesis is based on.

---

<sup>1</sup>Mainly based on Pierce (2007), Williams (1999), Morse and Ingard (1986) and Rabenstein et al. (2006).



**Figure 4.1.:** Illustration of the Huygens' principle.

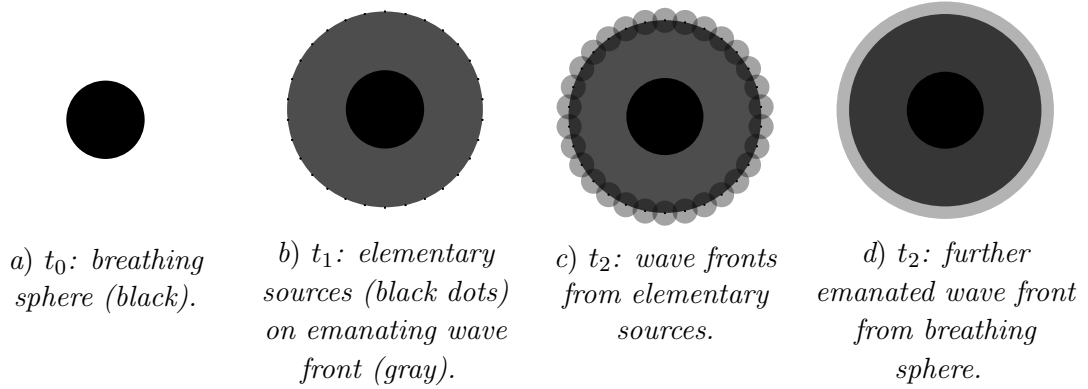
Figure 4.1 illustrates the Huygens' principle. Figure 4.2 clarifies this illustration by reducing it to two dimensions and splitting it into states at different points in time. The black disk in fig. 4.2 a) represents the source at  $t_0$  which creates a wavefront that spreads out concentrically. This wavefront is illustrated in dark gray in fig. 4.2 b) with some points on it. Each point on this wave front can be considered the origin of an elementary source, which again create a wave front, represented by the gray disks in fig. 4.2 c). Together, these wave fronts form the propagated wave front of the original source at a later point in time illustrated in fig. 4.2 d). The distance between those elementary waves has to be infinitesimally small. A monopole-shaped radiation of these elementary waves would create a second wave front at time  $t_2$ . This second wave front would be inside the earlier wave front, closer to the original breathing sphere again. This can clearly be seen in both fig. 4.1 and 4.2 c): One half of the elementary waves are located inside the dark gray wave front. This is physically untrue; the elementary waves must have a radiation characteristic which is 0 geared towards the source. This radiation characteristic is described by the Kirchhoff-Helmholtz integral (K-H integral), discussed in the subsequent subsection 4.1.2.

This principle constitutes the idea of an acoustic curtain<sup>2</sup>: If an array of microphones records elementary waves, the playback of these recordings via loudspeakers which are arranged in the exact same way as the microphones should recreate the

---

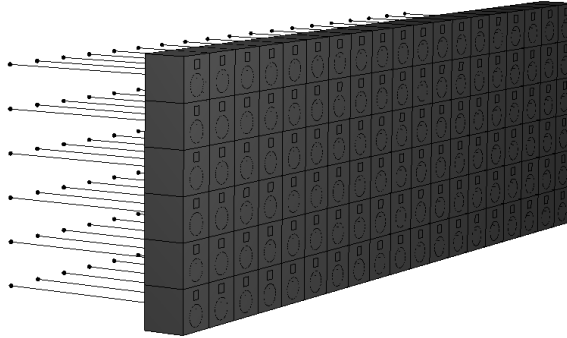
<sup>2</sup>See e.g. Ahrens (2012), pp. 8ff.





**Figure 4.2.:** Wave fronts of a breathing sphere at three points in time in 2D. The breathing sphere at  $t_0$  (a) creates a wave front at  $t_1$  (b). Points on this wave front can be considered as elementary sources which also create wave fronts at  $t_2$  (c). By superposition these wave fronts equal the further emanated wave front of the breathing sphere (d)

original wave field to a certain degree. This idea is a basic concept of wave field synthesis. It is illustrated in fig. 4.3.



**Figure 4.3.:** Scheme of an acoustic curtain: Loudspeakers are arranged in the same way as a plane microphone array (dots).

#### 4.1.2. Kirchhoff-Helmholtz Integral

The Gauss' theorem<sup>3</sup> states that spatial area integrals of a function over a volume  $V$  are equal to surface integrals of the normal components of a function over the volume's surface  $S$

<sup>3</sup>Also called "divergence theorem", see e.g. Pierce (2007), p. 58.

$$\int_V \nabla \vec{f} \, dV = \int_S \vec{f} \vec{n} \, dS \quad (4.1)$$

if it has a piecewise smooth boundary and the function  $\vec{f}$  is a steady, differentiable vector function<sup>4</sup>. A special case of the Gauss' theorem is described by Green's second theorem<sup>5</sup>:

$$\int_V \vec{f} \nabla^2 \vec{g} - \vec{g} \nabla^2 \vec{f} \, dV = \int_S \vec{f} \nabla \vec{g} \vec{n} - \vec{g} \nabla \vec{f} \vec{n} \, dS \quad (4.2)$$

From Green's second theorem and the wave equations, eq. 3.4 and 3.16, the Kirchhoff-Helmholtz integral can be derived, which links the wave field of a source-free volume  $V$  with sources  $\vec{Y}$  on its surface  $S$ :

$$-\frac{1}{4\pi} \oint_S \left( G(\omega, \Delta \vec{r}) \frac{\partial P(\omega, \vec{Y})}{\partial \vec{n}} - P(\omega, \vec{Y}) \frac{\partial G(\omega, \Delta \vec{r})}{\partial \vec{n}} \right) dS = \begin{cases} P(\omega, \vec{X}), & \vec{r} \in V \\ \frac{1}{2} P(\omega, \vec{X}), & \vec{r} \in S \\ 0, & \vec{r} \notin V \end{cases} \quad (4.3)$$

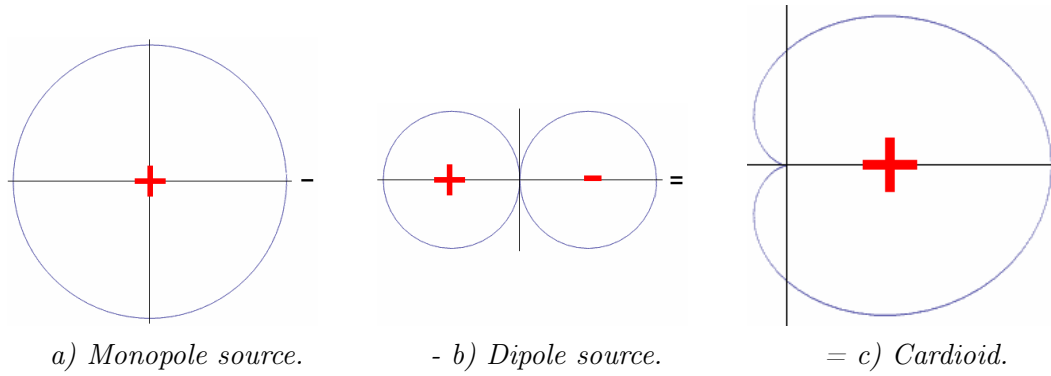
The K-H integral is a nonlinear differential equation which states that the spectrum  $P(\omega, \vec{X})$  at each point  $\vec{X}$  in a source-free volume  $V$  is the integral of the spectra  $P(\omega, \vec{Y})$  at every point  $\vec{Y}$  on the bounding surface  $S$  and their propagation function  $G(\omega, \Delta \vec{r})$  in the direction of the normal vector  $\vec{n}$  pointing inwards.  $G(\omega, \Delta \vec{r})$  is a Green's function, a solution of the inhomogeneous Helmholtz equation eq. 3.22, and  $P(\omega, \vec{Y})$  is a spectrum, a solution for the homogeneous Helmholtz equation eq. 3.9.  $\Delta \vec{r}$  is the Euclidean distance  $\|\vec{Y} - \vec{X}\|_2$ . The sources  $\vec{Y}$  on the boundary surface are secondary sources, excited by primary sources  $\vec{Q}$  which lie in the source volume  $U$ . The first term of the closed double contour integral describes a wave

<sup>4</sup>See Merziger and Wirth (2006), p. 551.

<sup>5</sup>See Merziger and Wirth (2006), p. 555.

which propagates as monopole since the propagation term  $G(\omega, \Delta\vec{r}) = \frac{e^{-ik\Delta\vec{r}}}{\Delta\vec{r}}$  is a monopole. From the periodic motion equation eq. 3.14 it emerges that  $\frac{\partial P}{\partial n}$  is proportional to sound particle velocity in normal direction  $\vec{V}_n$ . The second term of the integral is a wave which radiates as dipole, since  $\frac{\partial G(\omega, \Delta\vec{r})}{\partial \vec{n}} = \frac{1+ik\Delta\vec{r}}{\Delta\vec{r}^2} \cos(\varphi) e^{-ik\Delta\vec{r}}$  is a dipole term. Sound field quantities  $P$  and  $V$  are convertible into each other after Euler's equation of motion eq. 3.1 so the K-H integral is over determined and several approaches to a solution exist.

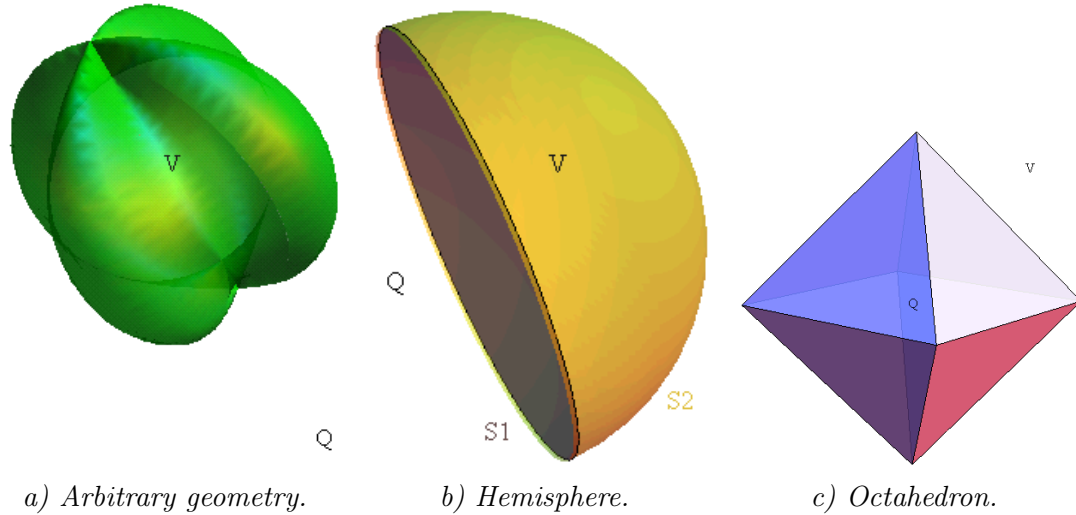
As already stated, the secondary sources on the surface of the source-free medium are monopole- and dipole-sources. Inwards they radiate in phase and outwards inversely phased. So the radiation doubles inwardly by constructive interference and outwardly becomes 0 by destructive interference. Combined, they create a *cardioid*, also referred to as *kidney* or *heart*. It is illustrated in fig. 4.4.



**Figure 4.4.:** Two dimensional illustration of superposition. Monopol- and dipole-source form a cardioid-shaped radiation.

The boundary surface could be the wave front around a source and the source free volume could be the room beyond this wave front. Then, the K-H integral is a quantified formulation of the Huygens' principle. But the volume could also be any arbitrary other geometry and the surface a physically existing or non-existing boundary. This boundary is the separation surface between a source volume, which contains one or more sources, and a source-free volume, which contains the listening area. Any arbitrary closed boundary is conceivable as long as the premises of the Gauss' theorem are observed. Figure 4.5 illustrates three examples for a volume boundary, which will be regarded in later chapters. Two equivalent setups exist:

Surrounding the listener with secondary sources — as in fig. 4.5 a) and c) — or surrounding the primary source(s), as illustrated in fig. 4.5 b)<sup>6</sup>.



**Figure 4.5.:** Three volumes  $V$  with possible source positions  $Q$ .

The Kirchhoff-Helmholtz integral describes analytically how spectrum and radiation on a volume surface are related to any arbitrary wave field inside a source-free volume. Therefore, this integral is the core of wave field synthesis<sup>7</sup>.

## 4.2. Practical Implementation of Wave Field Synthesis

The Kirchhoff-Helmholtz integral is a theoretical construct which cannot simply be put into practice by technical means. It demands a continuous distribution of an infinite number of secondary sources with infinitesimal distance, surrounding a volume entirely. That means sound pressure and velocity need to be controllable everywhere on the volume surface, which is hardly possible. However, what we can control is the sound pressure of loudspeakers. But an infinite number of infinitesimally distanced loudspeakers would be required, completely separating a listening

<sup>6</sup>Cf. Daniel et al. (2003), p. 3.

<sup>7</sup>See Berkhout et al. (1993), p. 2769.

area from a source volume and radiating inwards the listening area but not outwards. This is still hardly implementable. For a practical realization the reduction of secondary sources to a real number of loudspeakers with discrete distances radiating approximately as monopoles or dipoles is feasible<sup>8</sup>. These have to be fed with the correct “driving function”<sup>9</sup>. Surrounding an entire room with speakers is impracticable — as already mentioned in chapter 2.3.1 and illustrated in fig. 2.6 — and requires enormous technical challenges, computational power, acquisition- and operating-costs. Therefore, concepts with plane arrays<sup>10</sup> and line arrays<sup>11</sup> of the speakers are proposed.

For implementing such WFS systems the K-H integral has to be adjusted to the restrictive circumstances, which leads to errors in the synthesis. Adjustment steps, resulting synthesis mistakes and their effects as well as compensation methods are discussed in the following.

#### 4.2.1. Constraints for Implementation

A number of constraints simplify the K-H integral in a way which allows for a technical implementation of the theory by means of loudspeaker arrays<sup>12</sup>:

1. Reduction of the boundary surface to a separation plane between source-free volume and source volume
2. Restriction to one type of radiator (monopole or dipole)
3. Reduction of three-dimensional synthesis to two dimensions

---

<sup>8</sup>See e.g. Spors et al. (2008).

<sup>9</sup>See e.g. Spors et al. (2008).

<sup>10</sup>See e.g. Oellers (2010).

<sup>11</sup>One line, see Gauthier and Berry (2007), Baalman (2008), Kolundzija et al. (2009a), Cho et al. (2010), Reisinger (2003), Reisinger (2002) and Spors (2007), circular array, see Spors (2007), Rabenstein et al. (2006), Reisinger (2003), Reisinger (2002) and Rabenstein and Spors (2008), and three to four lines surrounding the listening area, see Spors et al. (2003), Reisinger (2003), Reisinger (2002), Rabenstein et al. (2006).

<sup>12</sup>These or similar simplifications are also proposed by Rabenstein et al. (2006), p. 529.

4. Discretization of the surface
5. Introduction of a spatial border

The particular steps will be successively accomplished in the following subsections.

### 4.2.2. Rayleigh-Integrals

Imagine a volume  $V$  consisting of a circular plane  $S1$  closing a hemisphere  $S2$ , as illustrated in fig. 4.5 b), whose radius converges to  $\infty$ . The influence of the radiation from the source on  $S2$  becomes 0 for the area in front of  $S1$ . This coherence satisfies the so-called *Sommerfeld condition*. It remains a separating plane between source free volume and source volume. The K-H integral then consists of an integral over the plane  $S1$  and thus fulfills the first simplification criterion from ch. 4.2.1:

$$-\frac{1}{4\pi} \iint_{S1} \left( G(\omega, \Delta\vec{r}) \frac{\partial P(\omega, \vec{Y})}{\partial \vec{n}} - P(\omega, \vec{Y}) \frac{G(\omega, \Delta\vec{r})}{\partial \vec{n}} \right) dS = \begin{cases} P(\omega, \vec{X}), & \vec{X} \in V \\ 0, & \vec{X} \notin V \end{cases} \quad (4.4)$$

This step reduces the area of secondary sources from a three-dimensional surrounding of a source-free volume to a separation plane.

Since the Green's function eq. 3.22 is a linear combination of a special solution and a general solution, one term of the integral can be eliminated by adding a deftly chosen general solution to the free-field Green's function. So the radiation can be restricted to one type of radiator. If the Green's function is chosen to be

$$G_D(\omega, \Delta\vec{r}) = \frac{e^{-ik\Delta\vec{r}}}{\Delta\vec{r}} + \frac{e^{-ik\Delta\vec{r}'}}{\Delta\vec{r}'}, \quad (4.5)$$

$G_D(\omega, \Delta\vec{r})$  is 0 on the surface  $S$  — which satisfies the so-called *homogeneous Dirichlet boundary condition*<sup>13</sup> — and the second term vanishes if  $\Delta\vec{r}'$  is the mirrored

---

<sup>13</sup>See e.g. Burns (1992).

position of  $\vec{X}$ , mirrored at the tangent of point  $\vec{Y}$  on  $S$ . This implicitly models the boundary as a rigid surface<sup>14</sup>, leading to the *Rayleigh I integral* for secondary monopole sources:

$$P(\omega, \vec{X}) = -\frac{1}{2\pi} \iint_{S_1} \left( G_D(\omega, \Delta\vec{r}) \frac{\partial P(\omega, \vec{Y})}{\partial \vec{n}} \right) dS. \quad (4.6)$$

Now, considering  $\frac{\partial P(\omega, \vec{Y})}{\partial \vec{n}}$  the desired source signal, an explicit solution can be found e.g. by means of wave field expansion. This approach is called “simple source approach” and is the basis of some sound field reconstruction methods such as higher-order ambisonics<sup>15</sup>.

Since the distance  $|\Delta\vec{r}|$  between secondary source position  $\vec{Y}$  and considered position in the source-free volume  $\vec{X}$  equals the distance between the secondary source position and the mirror position  $|\Delta\vec{r}'|$ ,  $G_D(\omega, \Delta\vec{r})$  is nothing but a doubling of the free-field Green’s function  $G(\omega, \Delta\vec{r})$ :

$$G_D(\omega, \Delta\vec{r}) = 2G(\omega, \Delta\vec{r}) \quad (4.7)$$

Assuming  $\frac{G_N(\omega, \Delta\vec{r})}{\partial n}$  to be 0 satisfy the *homogeneous Neumann boundary condition*<sup>16</sup> and the first term of eq. 4.4 vanishes. This is accomplished by choosing

$$G_N(\omega, \Delta\vec{r}) = \frac{e^{-ik\Delta\vec{r}}}{\Delta\vec{r}} - \frac{e^{-ik\Delta\vec{r}'}}{\Delta\vec{r}'}, \quad (4.8)$$

yielding the *Rayleigh II integral* for secondary dipole sources:

<sup>14</sup>See Spors et al. (2008), p. 4 and Baalman (2008), p. 27.

<sup>15</sup>Further information on higher-order ambisonics can be found e.g. in Spors et al. (2008), pp. 3f, Ahrens and Spors (2008), Williams (1999), pp. 267ff, Spors and Ahrens (2008), Daniel et al. (2003), Menzies and Al-Akaidi (2007), Daniel (2003) and Elen (2001).

<sup>16</sup>See e.g. Burns (1992).

$$P(\omega, \vec{X}) = -\frac{1}{2\pi} \iint_{S_1} \left( P(\omega, \vec{Y}) \frac{\partial G(\omega, \Delta\vec{r})}{\partial \vec{n}} \right) dS. \quad (4.9)$$

In both cases the second simplification criterion from chapter 4.2.1 is satisfied. But since the destructive interference outside the source-free volume is missing,  $P(\omega, \vec{X})$  for  $\vec{X} \notin V$  is not 0. A mirrored sound field in the source volume is the consequence. In case of monopoles the sound field created by the secondary sources is identical with the one inside the source-free volume. In case of dipole sources the phase in the source volume is the inverse of the phase inside the source-free volume. Additionally, the sound pressure or, respectively the particle velocity, duplicate by adding the general solution of the Green's function.

Both formulations do not apply for arbitrary volume surfaces but for separation planes only<sup>17</sup>. To ensure that any position around the listening area can be a source position, the listening area has to be surrounded by several separation planes. If eq. 4.6 and eq. 4.9 are applied to other geometries, they still deliver approximate results<sup>18</sup>. In any case, the source-free volume has to be convex so that no mirrored sound field lies inside the source-free volume, i.e. volume a) in fig. 4.5 is inappropriate<sup>19</sup>. Since  $S_1$  is implicitly modeled as a rigid surface, several reflections occur when a listening area is surrounded by several separation planes. These artifacts can be reduced by spatial “windowing”<sup>20</sup> technique applied to the Rayleigh I integral:

$$P(\omega, \vec{X}) = d(\vec{Y}) \frac{P(\omega, \vec{Y})}{\partial \vec{n}} 2G(\omega, \vec{Y}) \quad (4.10)$$

$$d(\vec{Y}) = \begin{cases} 1, & \text{if } \langle \vec{Y} - \vec{Q}, \vec{n}(\vec{Y}) \rangle > 0 \\ 0, & \text{otherwise} \end{cases}$$

<sup>17</sup>See Spors et al. (2008), p. 5.

<sup>18</sup>See Spors et al. (2008), p. 5.

<sup>19</sup>See Spors and Ahrens (2008), pp. 4f.

<sup>20</sup>See de Vries et al. (1994), Spors et al. (2008), p. 5 and Gauthier and Berry (2007), p. 3.



$d(\vec{Y})$  is the windowing function for spherical waves which is 1 if the local propagation direction of the sound of the virtual source at the position of the secondary source has a positive component in normal direction of the secondary source. If the deviation is  $\frac{\pi}{2}$  or more,  $d(\vec{Y})$  becomes 0 and the speaker is muted. That means only those loudspeakers whose normal component resembles the tangent of the wave front of the virtual source are active.  $G(\omega, \Delta\vec{r})$  describes the directivity function of the secondary source, i.e. of each loudspeaker. The other terms are the sought-after driving functions  $D$  of the loudspeakers<sup>21</sup>:

$$D(\omega, \vec{Y}) = 2d(\vec{Y}) \frac{P(\omega, \vec{Y})}{\partial \vec{n}} \quad (4.11)$$

Although considered as source- and obstacle-free field, it is to a certain extent possible to recreate the wave field of a virtual source within the source-free volume. This is achieved by assuming an inverse propagation and calculating a concave wave front at the surface which focuses at the position of the virtual source and creates a convex wave front from then on. These sources are called “focused sources”.<sup>22</sup>

## Two Dimensions

For applications in which the audience is organized more or less in plane, it is sufficient to recreate the wave field correctly for that listening plane only, rather than in the whole listening volume. Luckily, in many situations of listening to music, the listeners are organized roughly in plane, e.g. in many concert halls, opera halls, cinemas, theaters, in the car, on the couch etc. Furthermore, one or several one-dimensional distributions of loudspeakers are easier implementable than covering a complete room surface with loudspeakers. Reducing the three-dimensional wave field synthesis to two dimensions reduces the separation plane  $S1$  to a separation

<sup>21</sup>See Spors et al. (2008), p. 5.

<sup>22</sup>The derivation of the secondary source signals and further information on these sources can be found e.g. in Kim et al. (2009), Geier et al. (2010), Ahrens and Spors (2009b).

line  $L1$ . In theory, one could simply reduce the surface integral to a simple integral and the Rayleigh integrals would take the forms

$$P(\omega, \vec{X}) = \frac{1}{2\pi} \int_{L1} \left( G(\omega, \Delta\vec{r}) \frac{\partial P(\omega, \vec{Y})}{\partial \vec{n}} \right) dS1 \quad (4.12)$$

and

$$P(\omega, \vec{X}) = \frac{1}{2\pi} \int_{L1} \left( P(\omega, \vec{Y}) \frac{\partial G(\omega, \Delta\vec{r})}{\partial \vec{n}} \right) dS1. \quad (4.13)$$

In these cases  $\vec{X}$  is two-dimensional

$$\vec{X} = \begin{bmatrix} x \\ y \end{bmatrix}. \quad (4.14)$$

This solution was satisfying if no third dimension existed, e.g. if wave fronts of the secondary sources had no spherical but a circular or cylindrical propagation<sup>23</sup>. Then, the propagation function  $G(\omega, \Delta\vec{r})$  was different, having an amplitude decay of  $\frac{1}{\sqrt{r}}$  instead of  $\frac{1}{r}$ . This is owed to the fact that the surface  $S$  of a circle or cylinder doubles with a doubled circle radius  $r_{\text{circle}}$

$$S = 2\pi r_{\text{circle}} \quad (4.15)$$

in contrast to the spherical case in which it squares with the doubled radius as already indicated in eq. 3.24 in ch. 3.1.6. In this case

$$I \propto \frac{1}{r} \quad (4.16)$$

and thus

---

<sup>23</sup>See e.g. Spors et al. (2008) pp. 8f, Rabenstein et al. (2006), pp. 521ff.

$$p \propto \frac{1}{\sqrt{r}} \quad (4.17)$$

So the practical benefit of 4.12 and 4.13 is minor since transducers with a cylindrical radiation in the far field are hardly available<sup>24</sup>. An approximately cylindrical radiation could be achieved with line arrays of loudspeakers<sup>25</sup>. But replacing each individual loudspeaker by a line array of speakers contradicts our goal to reduce the number of loudspeakers. Simply replacing cylindrically radiating speakers by conventional loudspeakers which have a spherical radiation function leads to errors in this wave field synthesis formulation due to the deviant amplitude decay.

The Huygens' principle states that a wave front can be considered as consisting of infinitesimally distanced elementary sources. An infinite planar arrangement of elementary point sources with a spherical radiation could (re-)construct a plane wave, since the amplitude decay which is owed to the  $1/r$ -distance law is compensated by the contribution of the other sources. Imagining secondary line sources with a cylindrical radiation, linear arrangement of sources would be sufficient to create a planar wave front. In a linear arrangement of elementary point sources, the contribution of the sources from the second dimension is missing, resulting in an amplitude decay. Therefore, a “2.5D-operator” including a “far field approximation” which modifies the free-field Green's function to approximate a cylindrical propagation is used<sup>26</sup>. This changes the driving function to

$$D_{2.5D}(\omega, \vec{Y}) = \sqrt{\frac{2\pi |\vec{Y} - \vec{X}_{\text{ref}}|}{ik}} D(\omega, \vec{Y}) \quad (4.18)$$

with  $\vec{X}_{\text{ref}}$  being a reference point in the source-free volume. This yields the “2.5-Dimensional” Rayleigh integral<sup>27</sup>:

<sup>24</sup>Cf. Spors and Ahrens (2008), p. 6 and Goertz (2008), p. 444.

<sup>25</sup>As often applied in PA systems for concerts, see e.g. Friedrich (2008), pp. 316ff.

<sup>26</sup>See e.g. Spors et al. (2008), pp. 9f or Wittek (2007), p. 58.

<sup>27</sup>See Spors et al. (2008), p. 11, Baalman (2008), pp. 28–46 and Verheijen (1997), pp. 37–49 and pp. 153–156. The derivation of the 2.5D-operator is given in Ahrens (2012), pp. 288f.

$$P(\omega, \vec{X}) = - \int_{-\infty}^{\infty} D_{2.5D}(\omega, \vec{Y}) G(\omega, \Delta\vec{r}) \quad (4.19)$$

Taking reference points  $\vec{X}_{\text{ref}}$  parallel to the loudspeaker array, the wave field can be synthesized correctly along a reference line. Between the speakers and the reference line, the sound pressures are too high, behind it they are too low.

Until now, free-field conditions are assumed. However, if not installed in the free field, reflections may occur and superimpose with the intended wave field created by the loudspeaker system. Under the term “listening room compensation” a variety of methods are proposed to reduce the influence of reflections. The simplest form is passive listening room compensation which means that the room is heavily damped. This is an approved method, applied e.g. in cinemas. However, for some listening rooms, for example living rooms, damping is impractical. Therefore, active solutions are proposed, like adding a filtering function which eliminates the first reflections of the room to the calculated loudspeaker signals<sup>28</sup>. “Adaptive wave field synthesis”<sup>29</sup> uses error sensors which measure errors occurring during WFS of a test stimulus emerging e.g. from reflections. Then any WFS solution is modified by a regularization factor which minimizes the squared error. This is of course a vicious circle since compensation signals corrupt the synthesized wave field and are reflected, too, adding further errors. Due to an exponentially increasing reflection density it is hardly possible to account for all higher order reflections. Thus, the approach is limited to first order reflections

---

<sup>28</sup>See Horbach et al. (1999), Corteel and Nicol (2003), Spors et al. (2003), Spors et al. (2007), Spors et al. (2004), pp. 333–337.

<sup>29</sup>See Gauthier and Berry (2007).

## Discretization

A discretization of the Rayleigh integrals adopts the continuous formulation to discrete secondary source positions:

$$P(\omega, \vec{X}) = \frac{1}{2\pi} \sum_{\vec{r}_Y=-\infty}^{\infty} \left( G(\omega, \Delta\vec{r}) \frac{\partial P(\omega, \vec{Y})}{\partial \vec{n}} \right) \Delta\vec{r}_Y \quad (4.20)$$

and

$$P(\omega, \vec{X}) = \frac{1}{2\pi} \sum_{\vec{r}_Y=-\infty}^{\infty} \left( P(\omega, \vec{Y}) \frac{\partial G(\omega, \Delta\vec{r})}{\partial \vec{n}} \right) \Delta\vec{r}_Y \quad (4.21)$$

Thereby the Nyquist-Shannon sampling theorem has to be regarded: The sampling frequency has to be at least twice the highest frequency of the signal to be presented for no aliasing to occur. The highest frequency to be represented error-free is the *critical frequency* or *aliasing frequency*. In this case the sampling frequency is spatial; the speaker distance  $\Delta\vec{Y}$  has to be maximally half the distance of the largest presentable wavelength

$$f_{\max} = \frac{c}{2\Delta\vec{Y}} \quad (4.22)$$

between the speakers. The spatial sampling of the secondary source distribution is a process of sampling and interpolation; the interpolator is given by the radiation characteristics of the loudspeakers<sup>30</sup>. For the trace wavelength between the speakers

$$\lambda_{\Delta\vec{Y}} = \lambda |\sin \alpha| \quad (4.23)$$

is valid, where  $\alpha$  is the angle between the normal direction of a loudspeaker and the wave when striking this loudspeaker. »Respectively, it can be considered as

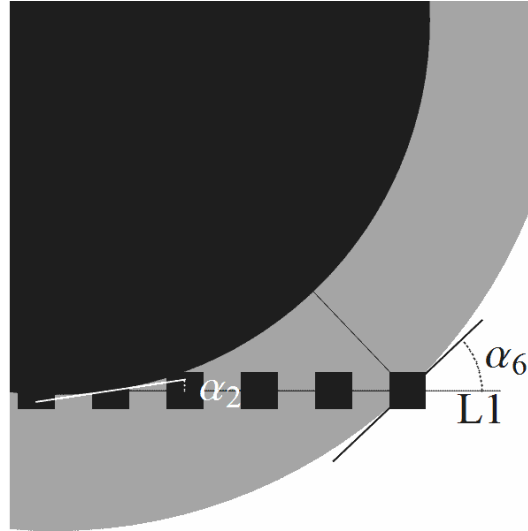
---

<sup>30</sup>See Spors (2008), p. 1. An adaption of WFS to the radiation characteristic of the loudspeakers is derived in de Vries (1996).

angle between separation line  $L1$  and the tangent of the wave front when striking the speaker position. This leads to an adjustment of eq. 4.22 to

$$f_{\max} = \frac{c}{2\Delta\vec{Y} \sin \alpha}. \quad (4.24)$$

The angle  $\alpha$  may vary dependent on position and radiation of the source in a range between  $\frac{\pi}{2}$  and  $\frac{3\pi}{2}$ . Two examples for  $\alpha$  are illustrated in fig. 4.6. to clarify the coherency. The black disk represents the source, the dark and light gray disks the wave front at two different points in time, just as in fig. 4.1 and fig. 4.2 in chapter 4.1.1.



**Figure 4.6.:** Several incidence angles for one source position.

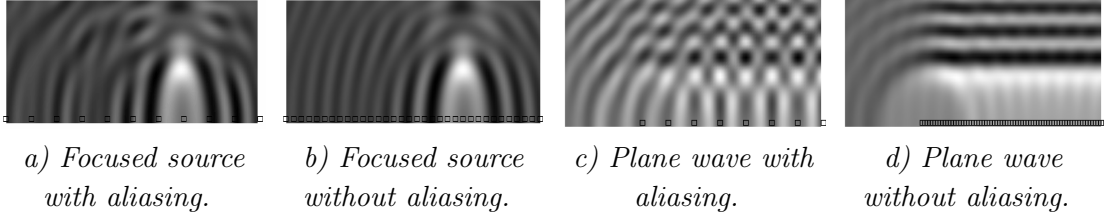
Undersampling creates erroneous wavefronts above  $f_{\max}$ . These erroneous wavefronts, *artifacts*, contain the frequencies above the critical frequency and cause perceivable changes in sound color and disturb the localization of the auditory event<sup>31</sup>. They can be heard as echoes or, respectively, as pre-echoes in case of focused sources and superimpose with the desired sound field as can be seen in fig. 4.7. The artifacts left in the aliasing-free plots result from the spatial borders which are treated in the next subsection.

<sup>31</sup>See Spors et al. (2008), p. 14 and Daniel et al. (2003), p. 15.

As long as the condition

$$|\sin \alpha(\omega)| < \frac{c}{2\Delta\vec{Y} f_{\max}} = \frac{\pi c}{\Delta\vec{Y} \omega_{\max}} \quad (4.25)$$

is satisfied no aliasing wavefronts will occur.



**Figure 4.7.:** Spatial aliasing of a focused source source a) and plane wave c). The aliasing-echoes vanish with small speaker distances, illustrated in b) and d).

One suggestion to reduce artifacts is to process frequencies above the critical frequency not by means of WFS but by conventional stereophonic sound between two to three loudspeakers. This method is called OPSI (“optimized phantom source imaging”<sup>32</sup>) and combines WFS with conventional intensity panning. Thus, OPSI reintroduces psychoacoustic considerations to WFS. In that manner no aliasing echoes as such occur but the common disadvantages of stereophonic sound become effective: A comb filter effect arises, the display of depth becomes worse and high frequencies are only located correctly on the sweet spot. At other positions two to three wave fronts arrive slightly shifted in time. Also the radiation characteristic of the instrument cannot be displayed. Lopez et al. suggest a related approach, called “sub-band approach”, playing frequencies above the aliasing frequency through the one loudspeaker with the most similar direction to the virtual source only<sup>33</sup>. This approach does not bring along the disadvantages of stereo but still a more or less correct localization is only possible in a small part of the listening area. Furthermore, the presentation of the complicated radiation characteristic of high frequencies is not possible. By randomizing the phase of the high frequencies of the loudspeakers artifacts are smeared<sup>34</sup>. This minimizes the sound coloration but localization accuracy is reduced<sup>35</sup>, since amplitude and phase have to be correct for a proper

<sup>32</sup>See e.g. Spors et al. (2008), p. 15 and Wittek (2007), pp. 96–105.

<sup>33</sup>See Lopez et al. (2005).

<sup>34</sup>See Spors et al. (2008), p. 17.

<sup>35</sup>See Wittek (2007), p. 88.

localization via WFS. Also, the resulting wave field does not correspond to the desired one. A reconstruction of the radiation characteristic of musical instruments is impossible with these methods.

In all three cases the signal is divided by the critical frequency into two frequency regions. For the lower frequency region the theory of WFS is applied. Frequencies in the higher region are not processed to recreate an original wave field but to remain the natural temporal and spectral properties as well as an approximately correct source position. The methods are based on the same psychoacoustic considerations: Partials of a sound tend to fuse, higher frequencies tend to be masked by lower frequencies and altogether the audible portion of sound will be integrated into one auditory stream with one group source position. Then, the lower frequency region — which offers very precise localization cues due to the correct reconstruction of the wave field — is crucial for a distinct and correct localization and the wrong localization cues of higher frequencies are neglected by the auditory system. All these psychoacoustic phenomena have been illuminated already in ch. 2.

Of course, these methods work best if the chosen distance between adjacent speakers is so small that the aliasing frequency is as high as possible. Then it can even be speculated that the influence of the frequencies above the critical frequency is weak concerning sound coloration and localization. Spors et al. confirm this assumption:

“However, the human auditory system seems to be not too sensible to spatial aliasing if the loudspeaker spacing is chosen in the range  $\Delta x = 10 \dots 30\text{cm}$ .”<sup>36</sup>

Quite a different method is to recreate the wave field not for the discrete loudspeaker positions but for discrete listening positions sampling the listening area. One approach is called “sound field reconstruction”<sup>37</sup>. Sampling positions are chosen under the assumption that if a wave field is reproduced correctly on a grid satisfying the Nyquist-Shannon sampling theorem, the wave field is correct everywhere inside the grid.

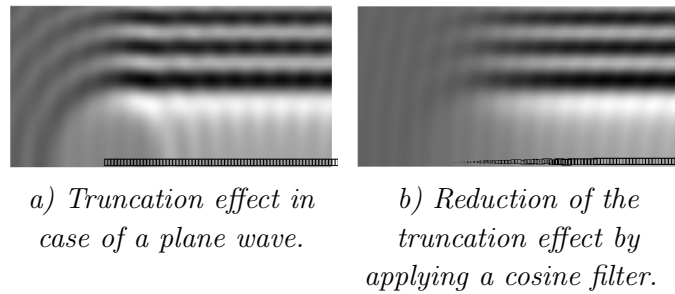
<sup>36</sup>Spors et al. (2008), p. 17. Note that Spors et al. name the speaker positions “ $x$ ”, in this thesis they are called  $\vec{Y}$ .

<sup>37</sup>Cf. Kolundzija et al. (2009b), p. 4.



### 4.2.3. Spatial Border

A constraint of the discrete Rayleigh integrals eq. 4.20 and 4.21 to a finite number of speaker positions is the 5th simplification of the list in ch. 4.2.1. This creates two borders from which the created wave front curvatures fade to the wave front of the speaker itself. This effect is called “truncation”<sup>38</sup>. It appears like diffraction through a gap and has the effect that the wave field cannot be synthesized in the area beyond the border. Furthermore, a spherical wave propagates from the border originated in the last speaker<sup>39</sup>, since the compensatory effect of adjacent speakers is missing. The truncation effect can be compensated by reducing the amplitudes of the outer speakers but this limits the listening area. Fig. 4.8 shows this artifact and its correction by applying a cosine filter on the outer speakers.



**Figure 4.8.:** Illustration of the truncation effect on synthesizing a plane wave (a) and its compensation by applying a cosine filter at the cost of a wide listening area (b). The size of the squares represents its amplitude. The spherical truncation wave emanating from the outer speaker is eliminated.

Compensation sources as used in listening room compensation minimize truncation by using speakers with antiphased signals. When two finite speaker arrays hit each other — as described in the beginning of chapter 4.2 for surrounding the listening area by three to four speaker line arrays — the border effect is inferior<sup>40</sup>.

<sup>38</sup>See Verheijen (1997), pp. 50ff and Baalman (2008), pp.37 ff.

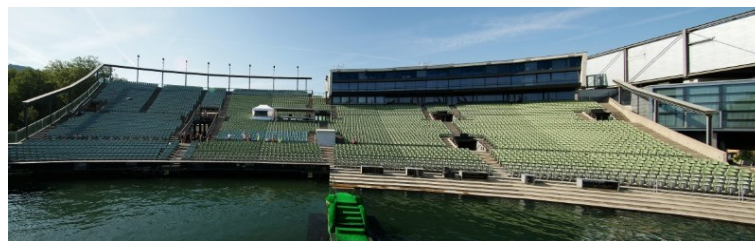
<sup>39</sup>See Spors et al. (2008), p. 14.

<sup>40</sup>See Verheijen (1997), p. 53.

### 4.3. Existing Applications

Although wave field synthesis is still at a stage of research and development, several systems are already in use, e.g. in cinemas, theaters, clubs, themed environments, schools and universities<sup>41</sup>.

One WFS system containing 832 loudspeakers fills the Seebühne Bregenz with sound, as illustrated in fig. 4.9. Another WFS system has been installed in the



**Figure 4.9.:** Photo of the WFS loudspeaker system at the Seebühne Bregenz. The speakers are arranged beside and behind the audience. From Fraunhofer-Institut für Digitale Medientechnologie IDMT (2014). ©Bregenzer Festspiele / Karl Forster.

Tresor club in Berlin as can be seen in fig. 4.10). A third system for research and public events can be found in the auditorium of the Technische Universität Berlin, illustrated in fig. 4.11.

These systems use hundreds of channels to synthesize a natural wave field in a listening area of umpteen to hundreds of square meters. The result is a natural sound with a good locatability in the whole listening area<sup>42</sup>.

---

<sup>41</sup>Most of all the WFS-system of the Technical University Berlin in cooperation with Deutsche Telekom Laboratories or IOSONO systems, the WFS system of the Fraunhofer Institute. Further information, see e.g. Baalman (2008), pp. 47ff and IOSONO GmbH (2008).

<sup>42</sup>See e.g. Pulkki (2008), p. 754.



**Figure 4.10.:** Photo of the IOSONO WFS loudspeaker system in the Tresor club Berlin. The speakers used to be arranged above the audience surrounding the dance floor. From IOSONO GmbH (2008).



**Figure 4.11.:** Photo of the WFS loudspeaker system in an auditorium of the Technische Universität Berlin with 832 channels and more than 2700 loudspeakers. ©Pressestelle der TU Berlin, taken from Technical University Berlin (2015).

## 4.4. Wave Field Synthesis and Radiation Characteristics

Radiation synthesis of musical sound is a research topic which receives growing interest<sup>43</sup>. Synthesizing the wave field as propagating from a source with a complex radiation pattern is a difficult task. One approach to recreate the natural sound radiation of musical instruments is to use an array of densely spaced loudspeakers. This approach is called “sound radiation synthesis”<sup>44</sup>. The idea has been derived and applied by several researchers<sup>45</sup>. They use platonic solids, other regular polyhedrons or a spherical loudspeaker arrangement, as illustrated in fig. 4.12, to create a high number of spherical harmonics. In the literature it is sometimes referred to as “spatial additive synthesis”<sup>46</sup>. The more loudspeakers are used the more complex radiation patterns can be synthesized. Assuming each loudspeaker to radiate as a monopole, the application is straightforward. However, implementing the actual radiation pattern of each loudspeaker and considering diffraction is challenging. An overview of methods is given in Zotter<sup>47</sup>. From auralizations in virtual physical models and from impulse response measurements in actual rooms it has been found that the directivity of musical instruments significantly affects the room response and leads to changes in the perceived naturalness and loudness<sup>48</sup>.

The massive advancement of wave field synthesis led to market-ready loudspeaker systems which are able to create impressively realistic sounds with a distinct location of the source. But typically, virtual monopole sources or plane waves are created, which have small perceived dimensions<sup>49</sup>. There have been many attempts already to recreate the sound radiation characteristics of musical instruments via wave field synthesis. Menzel et al. proposed a WFS method to create binaural signals for a single listening position<sup>50</sup>. Baalman uses several monopole sources

---

<sup>43</sup>See Ahrens (2012), p. 13.

<sup>44</sup>See e.g. Ahrens (2012), p. 13.

<sup>45</sup>See e.g. Avizienis et al. (2006), Pollow and Behler (2009) and Kassakian and Wessel (2004).

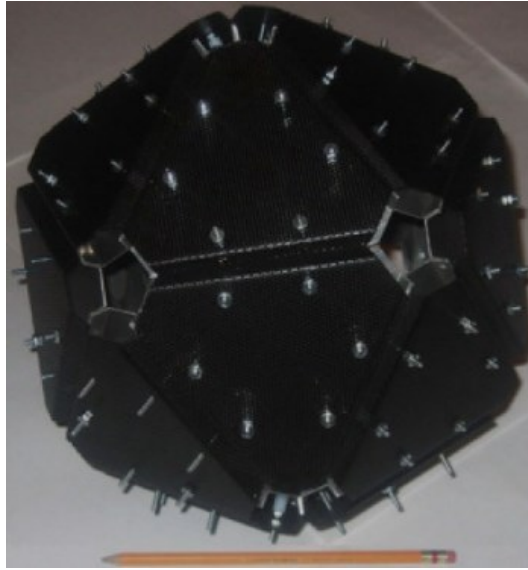
<sup>46</sup>See e.g. Warusfel and Misdariis (2004), p. 3.

<sup>47</sup>See Zotter (2009), pp. 111–152.

<sup>48</sup>See e.g. Martín et al. (2007), p. 395 and Otondo and Rindel (2004), p. 1183.

<sup>49</sup>See e.g. Ahrens (2012), p. 198ff.

<sup>50</sup>See Menzel et al. (2006).



**Figure 4.12.:** 120 loudspeakers mounted on the surface of a dodecahedron for matters of sound radiation synthesis. From Avizienis et al. (2006), with friendly permission of the Audio Engineering Society.

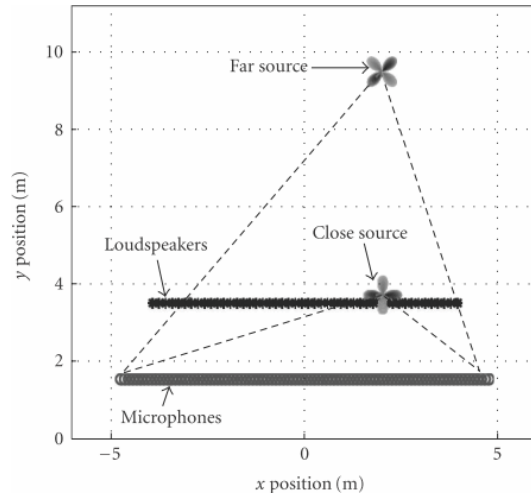
on the body of the virtual sound source to recreate its radiation patterns<sup>51</sup>. This approach is promising but the application is a compromise: A small number of monopole sources does not meet the complexity of many sound sources. A high number of monopole sources on the other hand may lead to an optimal recreation of the radiation characteristic but the computational costs are enormous. However, in more than 70% of the cases subjects of listening tests reported a higher “naturalness” for sources with complex radiation patterns compared to virtual monopoles<sup>52</sup>. Corteel uses a combination of spherical harmonics to calculate driving functions for WFS<sup>53</sup>. He created single and combined spherical harmonics with functions of degree  $-2$  to  $2$  using closed cabinet loudspeakers and multi-actuator panels (MAPs). A setup for his simulation and actual implementation is illustrated in fig. 4.13. Although artifacts appear, increasing with decreasing source distance and increasing radiation complexity, he found that such complex source radiations created natural variations while wandering through the listening area as well as an increased perceived “presence” compared to monopole sources<sup>54</sup>.

<sup>51</sup>See Baalman (2008), p. 97ff.

<sup>52</sup>See Baalman (2008), p. 19.

<sup>53</sup>See Corteel (2007).

<sup>54</sup>See Corteel (2007), p. 15.



**Figure 4.13.:** Setup for simulation and actual implementation of synthesizing a complex radiation pattern using wave field synthesis. From Corteel (2007), p. 4, under <http://creativecommons.org/licenses/by/4.0/legalcode>.

Ahrens gives two formulations to create virtual sound sources with complex radiation patterns via wave field synthesis: A finite line sources which is divided into sections vibrating with alternating algebraic sign and spheres vibrating in higher modes. However, he states that these approaches are intermediate steps to a solution to the problem of unnatural spatial radiation characteristics of virtual sources. Firstly, because the computational costs are enormous. And secondly, to him proper knowledge about the parameters that cause the perception of a certain source extent is needed to enable us to create a sound field that creates the desired spatial impression psychoacoustically, even if the physical wave field is different from a natural wave field emitted by a musical instrument.<sup>55</sup>

Blauert also proposes the use of psychoacoustics to overcome physical limitations of wave field synthesis applications. Fastl sees the future of room acoustics and auralization in psychoacoustics as well.<sup>56</sup> In the following chapter, a theoretical framework is developed to define, measure and store the radiation characteristics of musical instruments as well as to calculate and reconstruct the wave field as emitted from a natural source via psychoacoustic sound field synthesis. Subse-

<sup>55</sup>See Ahrens (2012), pp. 198ff.

<sup>56</sup>See Blauert (2008) and Fastl (2010), p. 9.

---

quently, the theory is is put into practice and its performance is tested by physical measurements, simulations and listening tests.

## 5. Spectral Psychoacoustic Wave Field Synthesis

One subject of this thesis is to investigate the radiation characteristics (RC) of musical instruments. The focus lies on its influence on the perception of instrumental sound in musical performances. The RC's main influence on perception appears to be of spatial nature. The radiation characteristic is a feature of instrumental sound that cannot be measured easily, compared e.g. to dynamics or musical scale, as listed in ch. 3.2. An adequate measurement method must be found to quantitatively describe it, make it comparable and eventually utilizable for psychoacoustic investigations. Not only that the radiation characteristic is difficult to measure, it also cannot be varied easily. As already exposed in ch. 3.2, it is only dependent on playing technique for some instruments, like flutes. Typically, it rather it depends on material properties and geometry of the source and the complicated interaction of its components. Typically, instruments with different RC tend to differ in many more parameters. Therefore, listening tests in which subjects evaluate a large number of sounds from many different instruments to explore the effect of RC by deducing it from factor analysis or multidimensional scaling is unpromising. It will suffer under multicollinearity. A more promising approach is to isolate the RC and vary it while keeping all other parameters constant. This idea seems likely but its application is not straightforward. It demands a reasonable definition and determination of the RC as physical parameter, and the possibility to measure, manipulate and recreate it. Methods like far field recordings, beamforming and nearfield acoustical holography can serve to determine the RC and isolate it from all other acoustic features. Provided that RC is adequately determined, wave field synthesis can serve to recreate sound fields that would have emanated from such



sources with arbitrary RC with high precision. Such an auralization is a highly challenging aim because it demands the use of up-to-date technology on a plausible theoretical background. Therefore, the derivation of a theoretical basis as well as the development of technical measurement- and recreation systems are an important milestone. Such a system needs to be applicable. Therefore, a method is developed using a small number of loudspeakers but being compatible to common loudspeaker systems — like stereo and Dolby Digital setups — as well as to WFS-applications with a high number of loudspeakers. For a practical realization with a small number of loudspeakers psychoacoustic considerations are necessary to provide a stable listening experience for listeners in an expanded listening area. Therefore, two loudspeaker systems are being developed and tested which make use of the auditive perception apparatus. They apply *spectral psychoacoustic wave field synthesis* to recreate the radiation characteristics of musical instruments, which will be introduced in this chapter.

The theory is described next, divided into the area of RC definition and acquisition and RC recreation. It is followed by the actual implementation which includes necessary adaption for an applicable measurement- and playback system. The first loudspeaker system, the *octupole speaker*, is able to recreate the RC of musical instruments to a certain degree. The second loudspeaker system, the *loudspeaker surrounding*, is capable of performing the same task but for various source positions. To accomplish that, an extended theory is necessary, which will be given in the last part of this chapter.

## 5.1. Theoretical Framework

This section is subdivided into two parts: The definition of the radiation characteristic and theory to reconstruct it.

### 5.1.1. Definition of the Radiation Characteristic

As already explained in ch. 3.2, it is feasible to consider a musical instrument as point source if its volume is small compared to the radiated wave length. But since musical instruments typically radiate sound from different parts there exists not one exclusive source point. Therefore, the head of the musician is taken as source point<sup>1</sup>. The instrument's spectrum  $P_{\vec{Q}}(\omega)$  is the solution to the homogeneous Helmholtz equation. It radiates from the source point according to its complex transfer function, the free-field Green's function  $G(\omega, r)$ . In case of a monopole it is independent of the propagation direction. Higher-order radiators have complex angular factors  $\Psi(\omega, \varphi, \vartheta)$  — which comprise of an azimuth and a polar factor  $\Gamma(\omega, \varphi)$  and  $\Theta(\omega, \vartheta)$  — that describes an amplitude- and phase shift per direction. Reducing the considered wave field to two dimensions for the measurement and reconstruction, as commonly done in wave field synthesis applications, eliminates the term  $\Theta(\omega, \vartheta)$ . Then,  $\Gamma(\omega, \varphi)$  is the only angle-dependent term which modifies pressure and phase on the wave front. It can be any arbitrary complex function of azimuth angle. Theoretically, even a Dirac impulse, which corresponds to a ray, could be composed of an infinite series of circular harmonics. Since  $\Gamma(\omega, \varphi)$  is the only term which is dependent on azimuth angle, it is sufficient to determine the RC of a point source in the far field. Eq. 5.1 quantifies the relationship between frequency spectrum, complex angular factor, spatial propagation and a measured signal at any point in space for a free field with one complex point source:

$$P_{\vec{X}}(\omega) = P_{\vec{Q}}(\omega) \Gamma_{\vec{Q}}(\omega, \varphi) G(\omega, \Delta\vec{r}) \quad (5.1)$$

Albrecht et al. suggest a similar way of formulating the sound source directivity as function of angle<sup>2</sup>. Recording the signal at any point in space  $\vec{X}$  and assuming the source to originate in exactly one point  $\vec{Q}$ , the measured signal  $P_{\vec{X}}(\omega)$  equals the source signal  $P_{\vec{Q}}(\omega)$ , modified by the amplification towards the direction  $\Gamma(\omega, \varphi)$  and distance  $G(\omega, \Delta\vec{r})$  between source point and receiver points for any frequency

<sup>1</sup>As commonly done, e.g. in Pätynen and Lokki (2010).

<sup>2</sup>See Albrecht et al. (2005).

$\omega$ . Together, the second and third term on the right hand side of eq. 5.1 constitute the transfer function of the sound. Propagating the sound this way yields plausible wave fields which exhibit a natural behavior: interaural level and phase differences decrease with increasing distance from the source<sup>3</sup>.

From that equation the radiation characteristic  $\Gamma(\omega, \varphi)$  cannot be calculated as long as the original spectrum  $P_{\vec{Q}}(\omega)$  remains unknown. But since  $G(\omega, \Delta\vec{r})$  is equal for all directions, what we can calculate is a relative  $\Gamma(\omega, \varphi)$ ,  $\Gamma' = \frac{\Gamma_{\vec{Q}}(\omega, \varphi, \vartheta)}{P_{\vec{Q}}(\omega)}$ , which can be considered an adequate measure to describe and compare the radiation characteristic of any complex point source. Standard works, like Meyer (2009) and Fletcher and Rossing (2008) use the same method to describe and compare the radiation characteristic of musical instruments, even for instruments that are much bigger than the considered wavelengths<sup>4</sup>. This simplification is physically untrue and neglects interferences between radiations from various parts of the instrument, different arrival times and amplitude decays. But besides the immense advantage of a uniformly defined measure, which makes the RC describable and comparable, it has another very important benefit: Its simplicity in application. Due to the reduction of the radiation characteristic to one function of angle only, complicated measurement techniques like beamforming and acoustical holography are dispensable. Furthermore, calculating the wave field emerging from such a complex point source is straight forward for any position and practical implementations showed that “a circular array of microphones leads to the best extrapolation results”<sup>5</sup>, which is of particular importance, as will be illuminated more specifically in ch. 5.2.4. An arbitrary source signal  $P(\omega)$  can be multiplied by the relative angular factor for the different frequency regions for the desired direction and forward-propagated according to the free-field Green’s function for the desired distance. No superposition of spherical harmonics or integral of point sources is necessary.

---

<sup>3</sup>This has been demonstrated in Ziemer (2014).

<sup>4</sup>See Fletcher and Rossing (2008), p. 395, showing the plots from Meyer (2009), pp. 163ff.

<sup>5</sup>Ajdler et al. (2008), p. 157.

### 5.1.2. Recreation of the Radiation Characteristic

The wave field which emanates from one source can be calculated for any position under free-field condition as described above. For psychoacoustic listening tests, this wave field needs to be reconstructed at least for the listener's position. An obvious way to recreate the spatial radiation characteristics musical instruments — which are considered as complex point sources — is to create a source which has a controllable RC using sound radiation synthesis as describes in ch. 4.4. Such a source with a controllable RC needs to be small enough to actually act as complex point source. Or it consists of densely spaced monopole sources, since every angular radiation pattern can be composed by a series of spherical harmonics, which themselves can be considered as densely spaced inversely phased monopoles. The radiation patterns to be recreated by the loudspeaker systems are typically far field recordings from circular microphone arrays as suggested earlier in this chapter. But for its reconstruction via spatial additive synthesis, loudspeakers are assumed to radiate as monopoles. This is only approximately true for loudspeakers with a closed cabinet and does not hold true for higher frequencies due to wave shadow, membrane modes etc. Furthermore, a high number of loudspeakers is necessary to recreate radiation characteristics of high order, as those that can be found for higher frequencies of musical instruments. Especially these higher order radiations can lead to interaural pressure- and phase differences which make the RC of musical instruments vivid and unique. That is why e.g. Avizienis et al. use 120 independent loudspeakers<sup>6</sup>. For a practical application, a solution with a small number of loudspeakers is needed which even accounts for nuances of very high order, at least for the position of the listener.

Although  $P_{\vec{Q}}(\omega)$ ,  $\Gamma(\omega, \varphi)$  and  $G(\omega, r)$  are all dependent on frequency,  $\omega$  is neglected in the following for matters of simplicity. The wave field at a listening position can be calculated by simply forward-propagating the sound of the musical instrument with the propagation function. The formula to calculate the pressure at any point in space  $P_{\vec{X}}$  in the far field area of a musical instrument is already given in eq. 5.1, assuming one complex point source in a free field. Together,

---

<sup>6</sup>See Avizienis et al. (2006).

$\Gamma_{\vec{Q}}(\varphi_{\vec{Q}-\vec{x}_m})$  and  $G(\Delta\vec{r})$  constitute the transfer function  $K(\Delta\vec{r})$  from the source to any chosen receiver position. Physically synthesizing these calculated wave fields by means of a loudspeaker system yields the perfect representation of the radiation characteristic of musical instruments for psychoacoustic investigations, if the position of the listener is chosen as receiver position. The theory of WFS is already given in ch. 4 as well as some adoptions for practical implementation. These will be partly applied and partly adjusted to this specific case as will be explained in the following section. Recreating the calculated wave fields at the desired points in space is the most crucial achievement for psychoacoustic investigations.

## 5.2. Practical Implementation

A major aim of this thesis is to develop a method to manipulate a recording of a musical instrument in such manner that the natural radiation characteristic of the instrument is recreated for listeners in an extended listening area. Compatibility to existing loudspeaker systems from 5.1 to WFS-systems is beneficial. Two loudspeaker systems are developed for this purpose. They are based on the same mathematical core. Its practical implementation is discussed in this section divided into four steps:

1. Measurement
2. Analysis and storage
3. Extrapolation
4. Reconstruction

First, the microphone setup to measure the radiation characteristic of musical instruments is described, followed by a subsection that explains the analysis and storage of the recorded data. With these data the wave field can be calculated at any point in the far field of a musical instrument with any arbitrary source position and signal by means of extrapolation or sound propagation as will be demonstrated

in ch. 5.2.3. At last, it is discussed how this calculated wave field is recreated by means of loudspeakers.

### 5.2.1. Measurement

A measurement setup is installed in a free field room to keep ambient noise and room-reflections small<sup>7</sup>. To measure the radiation characteristics of musical instruments, 128 omnidirectional microphones  $\vec{M}_1$  to  $\vec{M}_{128}$  are arranged concentrically to form a circular microphone array. The radius of this array is 1m, so the microphones have a distance of about 0.05m, providing one microphone every  $2.8^\circ$ . The instrument is driven naturally by the instrumentalist to create the original familiar instrumental sound, unlike the procedure of other studies where the instrument is artificially driven<sup>8</sup>. Note that this way the measurement contains the wave shadow behind the instrumentalist.

Instrumentalists are positioned in the center  $\vec{Q}$  with the instrument in plane with the array, facing microphone  $\vec{M}_1$  at angle  $\varphi_0$ . Fig. 5.1 is a photo of the actual setup. A polar coordinate system is assumed, with the instrument as origin. It corresponds to the polar coordinate system in ch. 3.1.6, eq. 3.18, but neglects the third dimension, the polar angle.

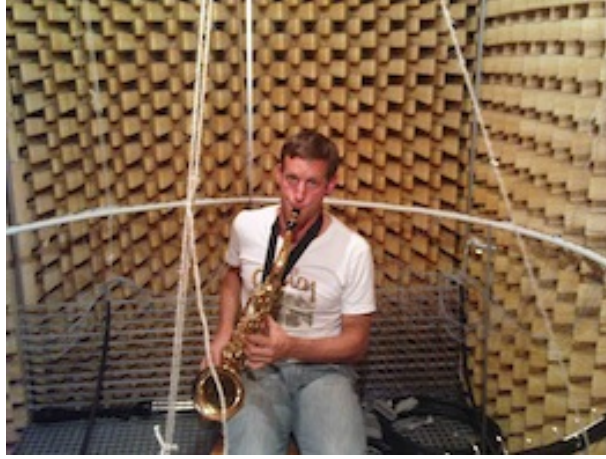
$$\vec{Q} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{5.2}$$

$$\vec{M}_m = \begin{bmatrix} 1 \\ \frac{2\pi(m-1)}{128} \end{bmatrix}, \quad m = 1, 2, 3, \dots, 128$$

---

<sup>7</sup>Technical data of the measurement equipment is given in appendix A.

<sup>8</sup>Cf. e.g. Warusfel et al. (1997).



**Figure 5.1.:** Photo of the measurement setup recording the radiation characteristic of a tenor saxophone. The microphones stick out of the circular rim as can be seen in the lower-right corner.

First, the microphone array is calibrated with a monopole source in the center to assure equal sensitivity of all microphones<sup>9</sup>. Then the instrumentalist plays a note on the instrument. All microphones simultaneously record two seconds of the sound with a sample rate of 48000Hz and a sample depth of 24bit. The first second contains silence caused by a short reaction time of the instrumentalist. It also contains the transient phase of the sound which is not taken into consideration since spectrum and radiation may change rapidly during the transient state. Several cycles are run with different notes covering the whole frequency range. A Fourier transform of the 128 signals yields one complex amplitude per frequency and direction. All the same the term “complex” will be left out since all amplitudes are meant to be complex. The measured amplitudes  $\Gamma_{\vec{Q}}(\varphi)$  describe the radiation characteristic of the instrument. Several instruments were recorded in that way.

The same measurement is to be done for the loudspeakers used in the WFS setup, too, always facing the first microphone. Its amplitudes per direction are named  $\Gamma_{\vec{Y}}(\varphi)$ . This is an important step as will be discussed extensively in ch. 5.2.4.

---

<sup>9</sup>See appendix A.1 for a detailed description of the calibration.

### 5.2.2. Analysis and Storage

From the recorded sounds, the transients are removed. One second of quasi stationary sound is transformed to spectral domain, using a discrete Fourier Transform as described in ch. 3.1.3. Thus, up to 24000 radiation patterns could theoretically be determined; one per detectable frequency. However, it is meaningful to limit it to a manageable number of patterns. Otondo and Rindel use octave bands and average the directivities of all partials within that octave<sup>10</sup>. Pätynen and Lokki use octave and third octave bands and interpolate between the measurement positions<sup>11</sup>. Other researchers decompose the recorded amplitudes at the discrete microphone positions to spherical harmonics which are continuous functions and therefore need no interpolation<sup>12</sup>.

As described in chapter 3.2, frequencies within the same range often tend to radiate in a similar way from the same instrument. This coincides with measurements as can be seen in fig. 5.2. The term “frequency range” is quite vague, therefore, I argue for the critical band width as an adequate range to cover the whole frequency range while keeping a manageable amount of data. The loudest frequency tends to mask the others within that critical band as already discussed in ch. 2.2. When this is the case, one radiation pattern per critical band is sufficient to psychoacoustically describe the whole radiation characteristic of that region even if several radiation patterns were present. No more than the loudest frequency could be heard by the listener anyway. 25 fixed bands are chosen according to the Bark scale as presented in ch. 2.2. Besides, the amount of data to be stored and processed is a good compromise between scrupulously accuracy and a separation into octave bands — the typical parameter in publications and applications concerning room acoustics.

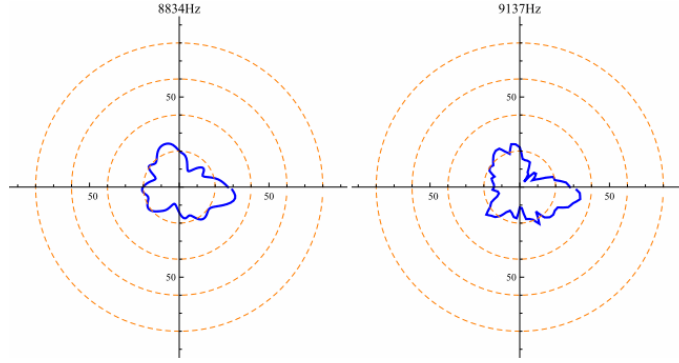
Thus, the 128 measured and calibrated complex amplitudes of one dominant frequency in each of the 25 critical bands is chosen to represent the radiation pattern of that frequency region. The radiation of all frequency bands of a shakuhachi is

<sup>10</sup>See e.g. Otondo and Rindel (2004), p. 1179 and Otondo and Rindel (2005), p. 903.

<sup>11</sup>See Pätynen and Lokki (2010), p. 139.

<sup>12</sup>See e.g. Pelzer et al. (2012) and Zotter et al. (2007).





**Figure 5.2.:** Radiation characteristic of two close violin frequencies. They have a very similar radiation pattern.

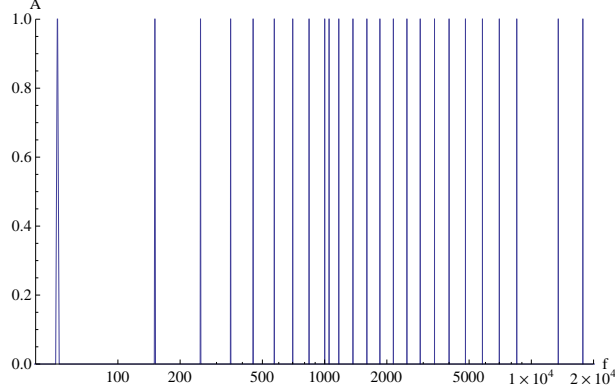
given as an example in the appendix ch. B. For matters of visualization they are presented in logarithmic dB-scale, all referred to the same sound pressure. For the signal processing, the amplitudes per frequency are not transformed into a logarithmic scale but normalized to a maximal absolute value of 1, which can be reached at any angle, by

$$\Gamma_{\vec{Q}}(\varphi) = \frac{P_{\vec{Q}}(\varphi)}{P_{\vec{Q}}(\varphi)_{\max}} \quad (5.3)$$

and stored in a database.

The measurement of the loudspeaker radiation has to be slightly different. A *Bark sound* is created, a quasi-stationary sound containing the center frequency of every critical band with equal amplitudes. The spectrum of the Bark sound is illustrated in fig. 5.3. The measured radiation must not be normalized for each frequency but as a whole. Since the input signal of the transducers has equal amplitudes, the measured output is already related to the same value. A loudspeaker is able to play exactly the same signal several times, therefore, the signal does not have to be recorded simultaneously from all directions but can be recorded successively

if no microphone array is available. One setup contains equal loudspeakers whose radiation characteristics are assumed to be the same<sup>13</sup>.



**Figure 5.3.:** Spectrum of the Bark sound.

The radiation characteristic of the loudspeakers in the octupole speaker system are measured individually since the cabinet is self-built and the speakers are mounted manually. This may cause slightly different radiation characteristics per speaker due to misplacements, asymmetry and inhomogeneous material. A detailed description of both loudspeaker systems is given subsequently in ch. 5.2.4.

### 5.2.3. Extrapolation

Once the radiation characteristic of a musical instrument is measured and stored, any source signal can be multiplied by the radiation characteristic and propagated towards a listening area according to eq. 5.1. This forward-propagation is an extrapolation of the signal from the source to the listening position.  $\Gamma_{\vec{Q}}(\varphi_{\vec{Q}-\vec{x}_m})$  and  $G(\Delta\vec{r})$  constitute the transfer function from the source to the listening position, which can be combined to one function  $K(\varphi, \Delta\vec{r})$ . Choosing discrete listening positions yields an “extrapolation matrix”<sup>14</sup>:

<sup>13</sup>Which is a premise for sound field reproduction setups, according to Ahrens and Spors (2009a), p. 2.

<sup>14</sup>See e.g. Daniel et al. (2003), p. 4.

$$K(\vec{Q}, \vec{r}) = \Gamma_{\vec{Q}}(\varphi) G(\Delta\vec{r}) \quad (5.4)$$

It describes the modification of the signal from the source to the listening positions. Summarizing the transfer function to one term  $K$ , function 5.1, which calculates the wave field radiated from a source, can be rewritten as

$$P_{\vec{X}} = P_{\vec{Q}} K(\Delta\vec{r}). \quad (5.5)$$

Calculating  $P_{\vec{X}}$  from a known  $K(\Delta\vec{r})$  is called “data-based rendering”, whereas “model-based rendering” allows to choose an arbitrary radiation characteristic<sup>15</sup>. The amplitudes  $P_{\vec{X}}$  describe the wave field at the listening positions under free field conditions which are to be reproduced by the loudspeaker system.

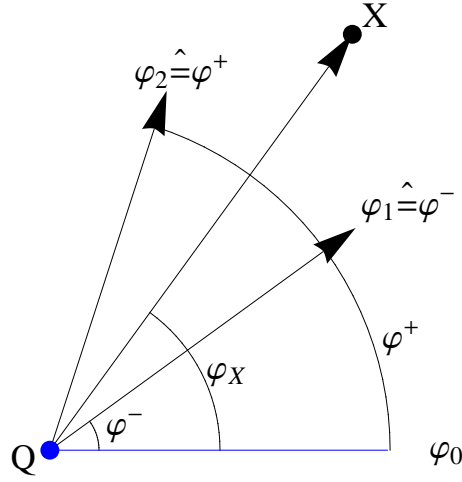
It is possible that an angle between source- and listening-position does not coincide with one of the 128 measured angles. In this case an interpolation between the directional factors of the two closest known angles offers an approximated factor. Eq. 5.6 gives the linear interpolation function. The interpolation is illustrated in fig. 5.4.

$$\Gamma_{\vec{Q}}(\varphi_{\vec{X}_m}) = \frac{\Gamma_{\vec{Q}}(\varphi^-) |\varphi_{\vec{X}_m} - \varphi^+| + \Gamma_{\vec{Q}}(\varphi^+) |\varphi_{\vec{X}_m} - \varphi^-|}{|\varphi^+ - \varphi^-|} \quad (5.6)$$

$\varphi^+$  and  $\varphi^-$  are the two measured angles closest to the actual angle between source and listening position  $\varphi_{\vec{X}_m}$ . After the wave field at the listening positions is determined by extrapolation — and interpolation where required — they need to be recreated eventually by means of a loudspeaker system.

---

<sup>15</sup>See Bleda et al. (2005), p. 2.



**Figure 5.4.:** Illustration of a case for an interpolation. The number of measured angles is reduced for a better visualization.

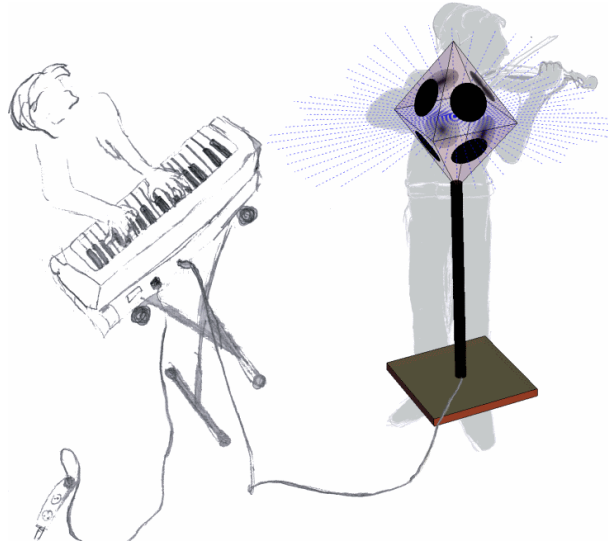
#### 5.2.4. Reconstruction

The calculated wave fields at the listening positions are modified versions of the original source signal. The modification is caused by the propagation through the room and the different radiation factors per angle. These wave fields are to be reproduced by a loudspeaker system. In other words, the speakers have to be fed by versions of the source signal which are manipulated in a way that results in exactly the calculated wave field at the listening positions. The setup of the two developed loudspeaker systems will be discussed next, followed by the calculation of the needed loudspeaker driving signals. The first system closely surrounds the virtual source position, the second system surrounds the listening area. Both systems are based on the same physical considerations and signal processing principles. However, the second system shall be able to simulate various virtual source positions which requires additional profound deliberations and even some psychoacoustic investigations, which are exhibited in an extra subsection. So, this subsection is structured as follows:

- The setup
- Calculation of the driving signals for the loudspeakers

- The loudspeaker surrounding

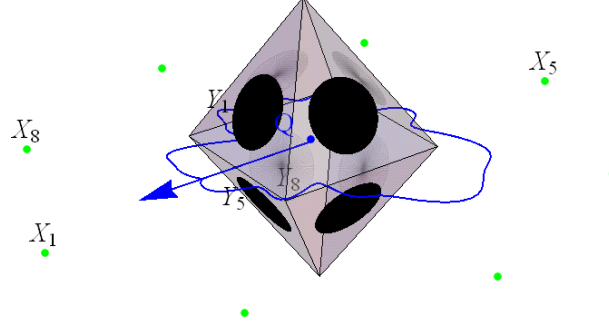
**The setup:** The first speaker system is an octahedron-shaped loudspeaker array, the *octupole speaker*, as illustrated in fig. 5.5. A photograph is given in appendix A.



**Figure 5.5.:** Vision of the octupole speaker system recreating the radiation characteristic of a violin from a keyboard input and thus emphasizing the perception of a real violin sound.

The geometry and the idea of the octupole speaker system were already introduced and shown in broad strokes in figure 4.5 in chapter 4.1.2. The whole setup is exemplary illustrated in fig 5.6. It has 8 fixed speaker positions  $\vec{Y}_n$  (black disks) which surround the virtual source position  $\vec{Q}$ . Details regarding the geometry of the octupole speaker system are given in appendix A. A virtual instrumentalist is placed at position  $\vec{Q}$  — indicated by dot in the center of the speaker — with viewing direction  $\varphi_0$  (denoted by the arrow). The contour represents the instrument's directional radiation factor  $\Gamma_{\vec{Q}}(\varphi)$ . In this case it is the radiation of a 7kHz partial of a violin. The contour is not to be confounded with the wave front which is assumed to be spherical in the far field. Rather, it can be considered as amplification factor per direction or as a curve of equal sound pressure in a propagating wave field. The triangular faces of the octahedron represent the separation surface between the

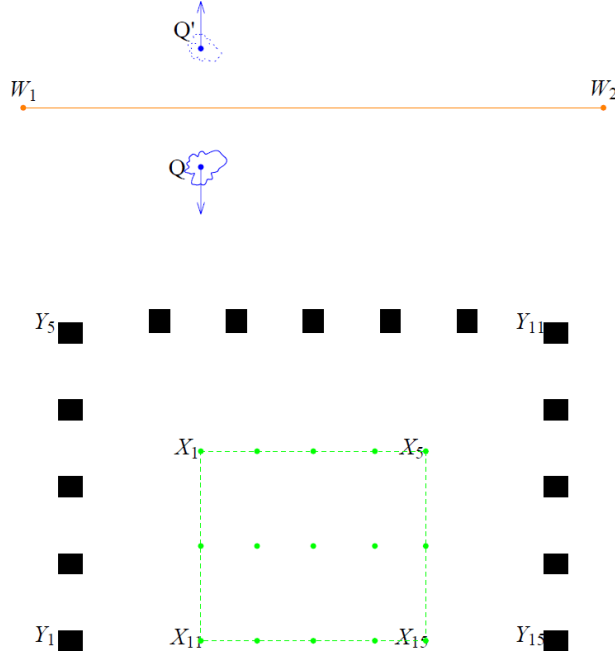
source volume  $U$  and the source free volume  $V$ . Any positions outside the source volume can be chosen as listening positions  $\vec{X}_m$ .



**Figure 5.6.:** Spatial setup of the octupole speaker system.

The second speaker array — the *loudspeaker surrounding* — contains three one-dimensional loudspeaker lines which surround the listening area in a rectangular array from three sides. Adjacent loudspeakers  $\vec{Y}_n$  have a distance of 0.65m. Listening positions  $\vec{X}_m$  are chosen, which span the listening area. Then the virtual source position  $\vec{Q}$  inside a virtual room and also the viewing direction  $\varphi_0$  are decided. The virtual room can be modeled with virtual walls ( $W$ ) via straight line functions or the definition of two points, e.g. corner points  $W_{12}$  = straight line from  $\vec{W}_1$  to  $\vec{W}_2$ . The setup is illustrated in fig. 5.7.

The black rectangles denote the loudspeakers  $\vec{Y}_1$  to  $\vec{Y}_{15}$ . They surround the listening area — denoted by the dashed rectangle — from three sides. The listening area is spanned by listening positions  $\vec{X}_{1\text{to}15}$ , denoted by the dots inside the dashed rectangle. The solid contour and the arrow are radiation factor  $\Gamma_{\vec{Q}}(\varphi)$  and viewing direction  $\varphi_0$  of the source as in the octupole setup. The dashed contour denotes radiation of mirror source  $\vec{Q}'$  which is used to model the first reflection. The amplitude drop caused by absorption is revealed by the size, the mirroring by the mirrored contour and arrow. The mirror source is mirrored at the wall which is a straight line between corner points  $W_1$  and  $W_2$ . For initial physical and psychoacoustic tests, no reflections are modeled.



**Figure 5.7.:** Spatial setup of the surrounding WFS system.

**Calculating the driving signals for the loudspeakers:** When driving the speakers, the amplitudes at each listening point  $P_{\vec{X}_m}$  are the sum of all propagated speaker signals:

$$\sum_{n=1}^N P_{\vec{Y}_n} \Gamma_{\vec{Y}_n}(\varphi_{\vec{X}_m}) G(\Delta\vec{r}) = P_{\vec{X}_m} \quad (5.7)$$

$P_{\vec{Y}_n}$  are the amplitudes of the speakers,  $\Gamma_{\vec{Y}_n}(\varphi)$  and  $G(\Delta\vec{r})$  form the transfer function  $K_{\vec{Y}_n}(\Delta\vec{r})$ ,  $\Delta\vec{r}$  indicates the distance between  $\vec{Y}_n$  and  $\vec{X}_m$ . In this equation the loudspeakers are assumed to radiate as complex point sources. In principle, this corresponds the discrete Rayleigh integral as introduced in ch. 4.2.2 but  $G$  is extended by a directional dependency  $\Gamma$ .

Eq. 5.7 in matrix notation:

$$\begin{bmatrix} K_{\vec{Y}_1, \vec{X}_1} & \cdots & K_{\vec{Y}_N, \vec{X}_1} \\ \vdots & \ddots & \vdots \\ K_{\vec{Y}_1, \vec{X}_M} & \cdots & K_{\vec{Y}_N, \vec{X}_M} \end{bmatrix} \cdot \begin{bmatrix} P_{\vec{Y}_1} \\ \vdots \\ P_{\vec{Y}_N} \end{bmatrix} = \begin{bmatrix} P_{\vec{X}_1} \\ \vdots \\ P_{\vec{X}_M} \end{bmatrix} \quad (5.8)$$

In case of given loudspeaker signals  $P_{\vec{Y}_n}$  the calculation of the resulting sound pressures at the listening positions  $P_{\vec{X}_m}$  is straightforward. But since the loudspeaker system is supposed to recreate a desired wave field,  $P_{\vec{X}_m}$  are given and  $P_{\vec{Y}_n}$  need to be calculated. Solving this linear equation system for all 25 considered frequencies leads to the needed loudspeaker signals to create the desired wave field at the listening positions. Replacing the original amplitudes from the source signal by the calculated amplitudes yields the desired wave field at the listening positions. The choice of listening positions has a huge influence on the solution. Choosing as many listening positions as loudspeakers, an explicit, unique solution to the equation system exists. This is applied in both developed audio systems. The applicability of an over-determined equation system with more listening positions than loudspeakers is discussed in the prospects, ch. 7. The listening positions are chosen to sample a listening area. Thereby, the aliasing frequency of the reconstruction is determined by the distance between listening positions rather than between adjacent loudspeakers. This corresponds to the sound field reconstruction approach as discussed in ch. 4.2.2. Of course, a solution to the linear equation system implies 15 manipulations of amplitude and phase. Especially phase manipulations become audible during transients. The precedence fade that will be introduced subsequently in ch. 5.3 does not solve this problem but reduces the audibility. Further compensation methods are discussed in the prospect, ch. 7.

**The ill-conditioning problem:** The linear equation system, eq. 5.8, resembles in principle the Rayleigh I integral, eq. 4.6 given in ch. 4.2.2. In this case, the number of sources is explicitly known as well as their positions in space. Still, this may lead to an ill-conditioning of the propagation matrix, if the sources are assumed to have a monopole-like radiation characteristic. Sources and measurement positions



are relatively close to another considering the wave length of low frequencies. A spacing of 0.3m, respectively 0.65m, between adjacent loudspeakers in the octupole loudspeaker- and the surrounding speaker setup roughly equals halve a wave length of a 570Hz- or, respectively, a 260Hz-frequency. Furthermore, the listening positions are quite close to another. This leads to similar entries between adjacent positions in the propagation matrix if the transfer function is similar for similar directions and distances. The consequence of the ill-conditioning could be a massive change of calculated loudspeaker signals when slightly changing the virtual source position. A relaxation of the equation system can make the calculation more robust. In case of eq. 5.8, the propagation term is already divided into an expansion coefficient  $G(\Delta\vec{r})$  and a directional coefficient  $\Gamma_{\vec{Y}_n}(\varphi)$ . Two solution to solve the ill-conditioning problem are presented in the following, based on an adjustment of  $\Gamma_{\vec{Y}_n}(\varphi)$  which affects the propagation matrix of eq. 5.8.

As a musical instrument can be simplified to a point source with a directional coefficient, it is feasible to consider loudspeakers the same way. Doing so, its directional coefficient can be measured for different angles and interpolated in the same way the musical instruments are treated. This leads to an explicit  $\Gamma_{\vec{Y}_n}(\varphi)$  which is typically different from a monopole being more differentiated in amplitude and especially in phase. Thus, it is likely that the natural radiation characteristic is less similar for adjacent listening positions and that the contribution of adjacent loudspeakers to one listening positions is also dissimilar. This approach has one big advantage: It resembles the actual nature of the loudspeaker. A disadvantage is that it does not guarantee for a robust solution as it remains uncertain if the natural radiation characteristic of the loudspeakers improves the condition of the propagation matrix  $K_{\vec{Y}_n, \vec{X}_m}$ . I refer to this approach as *radiation method*.

Another relaxation method is the minimum energy method (MEM) developed by Bader<sup>16</sup>. It models  $\Gamma_{\vec{Y}_n}(\varphi)$  as a lobe with an opening angle  $\Omega_n(\varphi)$  intermediate between hemisphere at  $n = 0$  and a ray in normal direction at  $n = \infty$ . The linear equation system is solved for different values, starting at  $n = 0$  and going to say  $N = 10$  in steps of 1. The energy of the solution, i.e. the sum of the squared pressures,

<sup>16</sup>The complete theory of the method is out of scope of this thesis. For details, see Bader (2010) and Bader (2014).

is stored. Around the solution with minimum energy, the linear equation system is solved again, this time in steps of 0.1. The  $\Omega_n(\varphi)$  that creates the solution with minimum energy is considered the “correct” driving function. This iterative process prohibits extremely high amplitudes. The solution is valid for half-spaces, i.e. the largest angle difference between a loudspeaker and all listening positions is  $\leq 180^\circ$ . This is not the case for the octupole speaker if the listening positions are distributed evenly around the speaker system. However, if they are located within the same half space, the hemisphere is equivalent to a monopole, since the loudspeakers surround the listening area concavely. In theory, this solution to the ill-conditioning problem has one major advantage: Since it minimizes the reconstruction energy, the calculated solution will make use of constructive interference rather than destructive interference. Therefore, the wave field besides the considered listening positions has lower sound pressures. This reduces the influence of unwanted reflections in an undamped listening room. But practically, in contrast to the NAH case, it is known that the radiation characteristic of the loudspeakers does not conform  $\Gamma_{\vec{Y}_n}(\varphi)$  with a certain  $\Omega(\omega, \varphi)$ , especially if the value of  $\Omega(\omega, \varphi)$  is permanently adjusted.

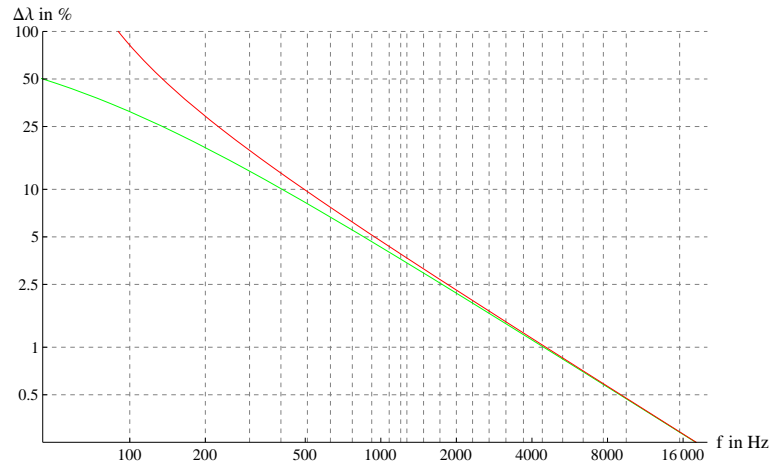
Both methods have opposed strengths: The radiation method keeps up the physical nature of the loudspeaker radiation to a certain degree by considering its RC, though simplified to a complex point source. The RC has to be loaded only and at most an interpolation is necessary for listening positions in between the discretely measured radiation factors. Then, the linear equation system is solved once only. Thus, it is real-time capable. But it does not guarantee for a robust solution. MEM serves a mathematically robust solution. But this solution does not conform the physical behavior of the loudspeakers since it demands an idealized loudspeaker-RC which is permanently changed. Furthermore, the iterative process demands to solve the linear equation system over and over again, as often as iterative steps are implemented. Achieving this is real-time is very difficult. Calculated loudspeaker signals of both methods are analyzed and compared in a simple test in ch. 6.3. The resulting wave field at positions other than the 15 listening positions are analyzed and compared with each other and with the predicted wave fields as well.

Yet, only the radiation method is applied in the developed audio systems. After the equation system is solved, the calculated complex amplitudes replace the original

amplitudes. Thus, 15 spectra are created from the source spectrum, only differing in amplitude and phase at the 25 considered frequencies. An inverse Fourier transform calculates the driving signals for the loudspeakers. Thus, the source signal is divided into overlapping time blocks. These are transferred to frequency domain, processed to find the needed spectra and then transformed back into time domain with overlapping blocks. The calculated overlapping time blocks are linearly cross-faded into another. This is necessary because amplitude- and phase could change between subsequent time blocks resulting in sudden changes in the time series which can be heard as impulsive sound. But during the crossfade the wave field is not reconstructed correctly as the loudspeaker signals do not resemble the calculated solution but a mixture of two solutions. However, the shorter the time block the longer overlap compared to the unfaded sound. For a real time implementation a fast Fourier transform (FFT) needs to be used. The block size determines the frequency accuracy of the FFT and thus the wave length accuracy. So the longer the time block the more accurate the wave field reconstruction due to a higher precision in wave length and due to longer periods of unfaded sound. On the other hand the shorter the time window the more the assumption of stationary sound is valid. Especially in the context of music, which typically involves lots of transients, the shortest possible time window is desirable. An FFT over 512 samples, which corresponds to approximately 11ms yields a frequency accuracy of  $\frac{1}{11\text{ms}} \approx 90\text{Hz}$ , so the worst case derivation between real frequency and FFT-analyzed frequency is  $\pm 45\text{Hz}$ . This still leads to a good wave length approximation, except for the lowest frequencies<sup>17</sup>. Fig. 5.8 illustrates the worst case wave length deviation based on FFT-analyzed and original frequency. The abscissa is the actual source frequency, the ordinate is the deviation between analyzed and actual wave length in percent. Considering wave length accuracy this is the shortest possible time block. Longer time blocks can be chosen if the processed signals contain many quasi stationary parts.

---

<sup>17</sup>Kolundzija et al. chose a duration of  $\approx 21\text{ms}$  for a sound field reconstruction, see Kolundzija et al. (2009a), p. 9, Warusfel et al. chose a block duration of 8ms for a radiation synthesis similar to the octupole speaker setup, see Warusfel et al. (1997), p. 9.



**Figure 5.8.:** Logarithmic plot of the worst case wave length deviation caused by frequency inaccuracy in case of an FFT over 512 Samples. The upper line is a too low assumed frequency, the lower line a too high assumed frequency. The vertical dashed lines visualize the margins of the critical bands.

### 5.3. The Loudspeaker Surrounding

As already demonstrated in ch. 4.2.2, the Rayleigh integrals implicitly model boundaries as rigid surfaces which leads to reflections if the listening area is surrounded by loudspeakers. These reflections could be avoided by a spatial windowing function, i.e. muting those loudspeakers whose normal component does not resemble the local propagation direction of the virtual source sound. Windowing reduces the number of loudspeakers used for the sound field reconstruction. For a WFS system which contains dozens of closely spaced loudspeakers in any direction, such a reduction has little effect, especially for the reconstruction of homogeneous wave fronts. But for the surrounding spectral psychoacoustic WFS system with 15 loudspeakers, any reduction would increase the aliasing frequency or reduce the size of the listening area since an equation system with as many listening positions as active loudspeakers is solved. So a reduction of loudspeakers automatically implies a reduction of listening positions. Therefore, a solution is needed which always involves all 15 speakers, despite their inward normal direction. This is only possible if the sound propagation direction is disregarded for the reconstruction progress. Presuming several assumptions, there is a satisfying solution to this as will be discussed in the following.

Solving the Rayleigh integral in frequency domain or, respectively, the linear equation system eq. 5.8, yields amplitude and phase for all 25 considered frequencies for each loudspeaker. Simply replacing amplitude and phase of the original sound signal in frequency domain does not result in a temporal delay of the signal in time domain after the inverse Fourier transform. Thus, a calculation in frequency domain is only valid for stationary signals. The spectrum as analyzed via Fourier transform is an integral over the considered time window  $\gamma(t - \tau)$ . Music naturally consists of many transients. Yet, analyzing short time windows, a large part of instrumental sound can be regarded as quasi-stationary. Only under this assumption the use of FFT is valid to approximate the spectrum of a sound. Not only for time but also for space the calculation is only valid for stationary signals. Calculating the wave field according to eq. 5.4 in frequency domain neglects the propagation time and thus the different arrival times of sound. If the source signal changes from a time window to the next one, the different traveling times lead to erroneous superpositions of the signals for the duration between the earliest and the latest arriving signal. This is a problem during transients, especially at the onset of notes. The erroneous superposition caused by different arrival times of wave fronts is similar to the erroneous wave fronts that have been described in ch. 4.2.2. It causes very obvious artifacts which make a proper localization of the virtual source almost impossible. If frequencies are considered quasi-stationary within a time-window but change between successive time windows, the calculated loudspeaker signals will only reconstruct the desired wave field approximately correct if the different arrival times at the listening positions are considerably smaller than the time window:

$$\frac{\vec{r}_{\min} - \vec{r}_{\max}}{c} \ll \gamma(t - \tau) \quad (5.9)$$

Here,  $\vec{r}_{\min}$  is the distance between a listening position and the nearest loudspeaker,  $\vec{r}_{\max}$ , respectively, between listening position and farthest loudspeaker. For the chosen setup the largest arrival time difference at any listening position approximately lies between 2.6ms and 6ms, with an average of 4.4ms. This corresponds 126 to 287 samples. Thus, even if a signal changes between successive time windows, the reconstruction will be correct for at least 45.5% to 76.4% of the time. How this

arrival time difference and the reconstruction errors affect the listening experience can be assessed by considering the involved psychoacoustic effects discussed in ch. 2: Erroneous changes in amplitude of consinuous sounds arising between two successive time windows may become audible at a level difference of 0.8dB. For short sounds the JND it is even larger. The worst-case arrival time difference of wave fronts is distinctly below temporal resolution of the auditory system lying around 50ms. Thus, no perception of echo should occur and it is likely that all arriving wave fronts will be integrated into one auditory stream the more features they have in common. Since especially the spatial features synchrony and location are diverse, the loudspeaker signals must resemble in enough other features — like harmonicity, timbre, common fate, etc. — to be perceived as integrated sound. It is this lack in synchrony and location that might lead to artifacts which may result in the perception of a vague, diffuse source location or in the perception of several separate source positions. The reason for that is that the JND in interaural arrival time is much less than the arrival time difference of wave fronts for all listening positions. Therefore, a method is needed to assure for a distinct perceived source position. This could either be done by eliminating erroneous wave fronts especially at the onsets of sounds or by creating a wave field that results in a correct localization due to psychoacoustic panning as in conventional stereophonic sound setups.

**The Precedence Fade:** Since only distances between speakers and listening positions are considered but not transfer time or propagation direction, the precedence effect will appear at the onsets of sound and may last for several seconds, as discussed already in ch. 2.3.4. It shifts the auditory event position — i.e. the perceived source position, as introduced in ch. 2.3.1 — towards the nearest loudspeaker. The result is unwanted but its initiator can be used: If only one loudspeaker actively plays the onsets, the perceived source position will coincide with the loudspeaker position. Already Lopez et al. suggested to make use of that fact for high-order spatio-temporal wave field synthesis systems<sup>18</sup> as extensively discussed in ch. 4, particularly ch. 4.2.2. Technically amplified performances open air or in big halls or stadiums make use of the precedence effect as well. Signals far away from the stage are delayed so much that the wave front of the PA loudspeakers on stage

---

<sup>18</sup>See Lopez et al. (2005).

always arrives first, independent of the listening position. Thus, the signal is amplified but due to the precedence effect the localized sound source stays on the stage.

In this thesis, the speaker whose signal precedes the signals of the others will be called *precedence speaker* and has to be the loudspeaker which lies closest to the connection line of virtual source and listening position. 15 listening positions span the listening area, so the loudspeaker closest to the connection line between virtual source and central listening position should be the precedence speaker. Then, all other loudspeakers can contribute to the sound field reconstruction. Since all loudspeaker signals are derived from the source signal, they are very similar and most likely integrated into one auditory stream, as discussed in ch. 2.4. That means if any signal arrives within less than 50ms after the precedence speaker's signal it should not be localized individually but all signals should have one group location. This group location coincides with the precedence speaker if its signal arrives sufficiently earlier. This assumption is only valid if the speakers show no distinct onset which distinguishes them from the precedence speaker signal. Therefore, it is meaningful to fade them in after the precedence signal. This *precedence fade* must not be too short, otherwise it influences the auditory event direction. An interaural time difference of 0.5 to 5ms can be sufficient to create an auditory event position at the precedence speaker<sup>19</sup>. At an ITD of about 50ms, the impression of an echo emerges. Therefore, the signals should completely fade in between  $5\text{ms} + 6\text{ms} = 11\text{ms}$  and  $50\text{ms} - 6\text{ms} = 44\text{ms}$  to guarantee an ITD between 5ms and 50ms at all listening positions for all virtual source positions. As mentioned earlier, 6ms is the largest arrival time difference possible in the listening area. Unfortunately, this method only allows virtual sources at the 15 loudspeaker positions. Furthermore, this premise reduces maximum size of the listening area to a region where the farthest listening points have a distance of  $\leq 15.1\text{m}$ . Overcoming these restrictions is a topic of future research.

---

<sup>19</sup>See e.g. Dickreiter (1978), p. 82 and Verheijen (1997), p. 17.

## 6. Performance Tests of the Developed Audio Systems

### 6.1. Physical Measurement of the Octupole Speaker

To validate the theory for the recreation of the radiation characteristic, as introduced in ch. 5.1.2, a physical test is conducted in a free field room with controlled parameters.

The bark sound which was initially used to measure the radiation characteristic of the loudspeakers — as described and illustrated in ch. 5.2.2 — is chosen as test signal  $p_{\vec{Q}}(t)$ . It is a 2s-long quasi-stationary signal which contains the central frequency of each Bark band with equal amplitudes. A sound beam is used as radiation characteristic of the virtual source  $\Gamma_{\vec{Q}}(\omega, \varphi)$ . It has a value of 1 at the reference angle  $\varphi_0$  and 0 at all other positions for all frequency bands. The chosen listening positions  $\vec{X}_n, n = 1, 2, \dots, 8$  are arranged concentrically around the virtual source position at a distance of 1m. One listening position every  $45^\circ$  exists as can be seen in fig. 5.6 in ch. 5.2.4. Eq. 6.1 lists all given magnitudes.



$$\Gamma_{\vec{Q}}(\varphi) = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \begin{aligned} \vec{Q} &= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\ P_{\vec{Q}}(\omega) &= 0 + \imath \\ \vec{X}_n &= \begin{bmatrix} 1 \\ \frac{2\pi n}{4} \end{bmatrix}, n = 1, 2, \dots, 8 \end{aligned} \quad (6.1)$$

The wave field at the listening positions in frequency domain  $P_{\vec{X}_n}(\omega)$  is calculated according to eq. 5.1 in ch. 5.1.1:

$$P_{\vec{X}_i}(\omega) = 0 + \imath \Gamma_{\vec{Q}}(\varphi) e^{-\imath k} \quad (6.2)$$

Since the distance between source and listening position  $\Delta r_{\vec{Q}-\vec{X}}$  is 1, the distance term  $G(\omega, \Delta \vec{r}) = \frac{e^{-\imath k r}}{r}$  reduces to  $e^{-\imath k}$ . Thus, at the listening positions the wave field corresponds to a phase-shifted  $\Gamma_{\vec{Q}}(\varphi)$ . These calculated values are to be created by the loudspeaker system. Therefore, the linear equation system from ch. 5.8 is solved:

$$\begin{bmatrix} K_{\vec{Y}_1-\vec{X}_1} & \dots & K_{\vec{Y}_8-\vec{X}_1} \\ \vdots & \ddots & \vdots \\ K_{\vec{Y}_1-\vec{X}_8} & \dots & K_{\vec{Y}_8-\vec{X}_8} \end{bmatrix} \cdot \begin{bmatrix} P_{\vec{Y}_1} \\ \vdots \\ P_{\vec{Y}_8} \end{bmatrix} = \begin{bmatrix} P_{\vec{X}_1} \\ \vdots \\ P_{\vec{X}_8} \end{bmatrix} \quad (6.3)$$

The 25 calculated complex amplitudes simply replace the original ones. Then, an inverse Fourier transform is carried out to receive the 8 loudspeaker signals for the sound field synthesis. The radiated wave field is measured at 72 angles and 1m distance around the center of the loudspeaker system.

Considering the 8 calculated listening positions only, a sound beam should emerge, i.e. an outstanding amplitude at listening position 1 in comparison to positions 2 to 8. The complex amplitudes between the listening positions are not considered for the wave field reconstruction. But for wave lengths larger than the dimensions of the loudspeaker system and the recording positions, few fluctuations are expected and at least the positions adjacent to the 8 listening positions should not differ too much from the calculated amplitudes.

Fig. 6.1 shows the amplitudes as measured in a free field room. The produced sound beam can clearly be seen, having a maximum amplitude on or sometimes close to, listening position 1. At listening positions 2 to 8, the amplitude lies between  $-0.89\text{dB}$  and  $52.42\text{dB}$  lower, with an average of  $-18.41\text{dB}$ . Positions 3 to 7 have a lower amplitude than 2 and 8, the positions adjacent to the lobe of the sound beam. At higher frequencies, the sound beam does not visibly protrude but “fades” to the random values at the other measurement position. Considering the listening positions only, one can see that the amplitude at listening position 1 is still much higher than at positions 2 to 8. The measured amplitudes are summarized in tab. 6.1.

Fig. 6.2 illustrates the measured phases at all 72 measurement positions for all 25 frequencies. Corresponding to eq. 6.2, the phase per frequency should be equal at any listening position. As for the amplitudes, the values at all angles between the 8 listening positions are not considered for the wave field reconstruction. Still, they are determined by the calculated solution. At low frequencies they should not feature many fluctuations due to the smooth, homogeneous, monopole-like radiation characteristic of the speakers in that frequency region and the big wave lengths. At higher frequencies more fluctuations may occur since the radiation patterns of the loudspeakers are more complicated. Furthermore, effects like wave shadows and deflections behind and around the cabinet become more effective and the smaller wave lengths lead to more complicated interference patterns. As can be seen in the figure, no distinct spikes occur below  $2900\text{Hz}$ , except at  $1000\text{Hz}$ . Table 6.2 summarizes the relative phase of each frequency for each position  $\phi_{\text{rel},\vec{X}}(\omega)$  with respect to the measured phase at the first listening position  $\phi_{\vec{X}_1}(\omega)$ , defined as follows:

$f$ in Hz	$P_{\vec{X}_1}$	$P_{\vec{X}_2}$	$P_{\vec{X}_3}$	$P_{\vec{X}_4}$	$P_{\vec{X}_5}$	$P_{\vec{X}_6}$	$P_{\vec{X}_7}$	$P_{\vec{X}_8}$	$\frac{OP_{\vec{X}_{2to8}}}{P_{\vec{X}_1}}$
50	-6.6	-33.88	-34.39	-32.66	-37.38	-34.83	-43.23	-21.92	<b>-25.13</b>
150	0	-34.86	-34.93	-43.75	-43.79	-42.93	-35.38	-34.48	<b>-37.63</b>
250	-3.24	-26.78	-32.46	-29.35	-33.97	-33.37	-30.5	-23.59	<b>-26.01</b>
350	-3.43	-17.1	-24.49	-31.38	-31.3	-27.01	-24.37	-20.58	<b>-20.37</b>
450	-6.92	-37.4	-34.04	-33.8	-34.44	-35.06	-43.44	-35.09	<b>-28.79</b>
570	-2.88	-19.69	-31.36	-28.84	-30.63	-30.78	-29.4	-24.57	<b>-23.97</b>
700	-2.63	-16.49	-21.88	-26.17	-25.96	-29.9	-22.21	-19.62	<b>-19.56</b>
840	0	-33.29	-25.7	-29.54	-29.67	-34.75	-26.47	-24.4	<b>-28.41</b>
1000	0	-43.12	-45	-35.8	-37.73	-38.06	-52.42	-44.23	<b>-40.94</b>
1170	0	-24.04	-30.59	-26.12	-29.37	-47.49	-31.83	-22.87	<b>-28.04</b>
1370	0	-23.03	-26.75	-26.45	-43.4	-29.86	-20.1	-13.04	<b>-22.5</b>
1600	-2.53	-15.26	-29.86	-22.91	-44.84	-22.28	-23.27	-22.88	<b>-20.4</b>
1850	-2.34	-16.55	-38.17	-33.11	-35.4	-27.15	-33.53	-25.1	<b>-24.44</b>
2150	-3.61	-16.06	-35.46	-30.78	-21.88	-22.65	-22.5	-16.61	<b>-18.02</b>
2500	-1.5	-23.3	-28.31	-23.38	-33.99	-28.68	-30.01	-11.25	<b>-20.81</b>
2900	0	-24.54	-31.7	-27.34	-26.51	-48.9	-31.36	-19.88	<b>-27.14</b>
3400	-0.65	-14.59	-17.06	-23.19	-20.92	-19.21	-17.57	-17.77	<b>-17.59</b>
4000	-3.87	-12.81	-14.75	-21.19	-23.41	-30.42	-15.88	-11.2	<b>-12.75</b>
4800	-1.36	-10.1	-19.69	-18.06	-20.56	-25.4	-11.93	-2.25	<b>-10.86</b>
5800	-2.69	-10.41	-23.6	-13.98	-14.84	-31.35	-12.43	-6.49	<b>-10.76</b>
7000	0	-19.66	-13.87	-17.87	-25.36	-17.26	-13.89	-6.64	<b>-14.67</b>
8500	-2.06	-4.07	-20.44	-19.82	-28.31	-18.58	-24.83	-7.26	<b>-11.5</b>
10500	-1.16	-4.12	-20.3	-15.73	-31.86	-16.15	-8.77	-7.07	<b>-10.3</b>
13500	-2.34	-23.39	-12.32	-14.14	-24.17	-12.92	-13.71	-3.61	<b>-10.16</b>
17750	-1.27	-4.05	-17.15	-15.75	-17.34	-7.56	-8.17	-12.53	<b>-10.42</b>
$\emptyset$	<b>-1.3</b>	<b>-13.41</b>	<b>-21.11</b>	<b>-20.95</b>	<b>-24.59</b>	<b>-21.53</b>	<b>-17.93</b>	<b>-11.56</b>	<b>-18.41</b>

**Table 6.1.:** Measured amplitudes of the octupole sound beam test. The amplitudes at positions 2 to 8 are considerably lower than at position 1, with an average of  $-18.41\text{dB}$ . The best performance is achieved below 3400Hz.

$$\phi_{\text{rel},\vec{X}}(\omega) = \begin{cases} \frac{\phi_{\vec{X}_1}(\omega) - \phi_{\vec{X}}(\omega)}{\pi}, & \text{if } |\phi_{\vec{X}_1}(\omega) - \phi_{\vec{X}}(\omega)| \leq \pi \\ \frac{\phi_{\vec{X}_1}(\omega) + \phi_{\vec{X}}(\omega)}{\pi}, & \text{otherwise} \end{cases} \quad (6.4)$$

With this formulation  $\phi_{\text{rel},\vec{X}_1}(\omega)$  is always 0 and  $\phi_{\text{rel},\vec{X}_2 \text{ to } 8}(\omega)$  lie between  $-0.5$  and  $0.5$ . Both extremes represent a phase inversion. The values are given in percent, from  $-50\%$  to  $50\%$ .  $\phi_{\text{rel},\vec{X}_2 \text{ to } 8}(\omega)$  are supposed to lie around 0, which would correspond to equal phases. Additionally, the mean value  $\bar{\phi}$  of the absolute phase differences are given as well as the standard deviation  $\sigma$  of the seven absolute values per frequency.

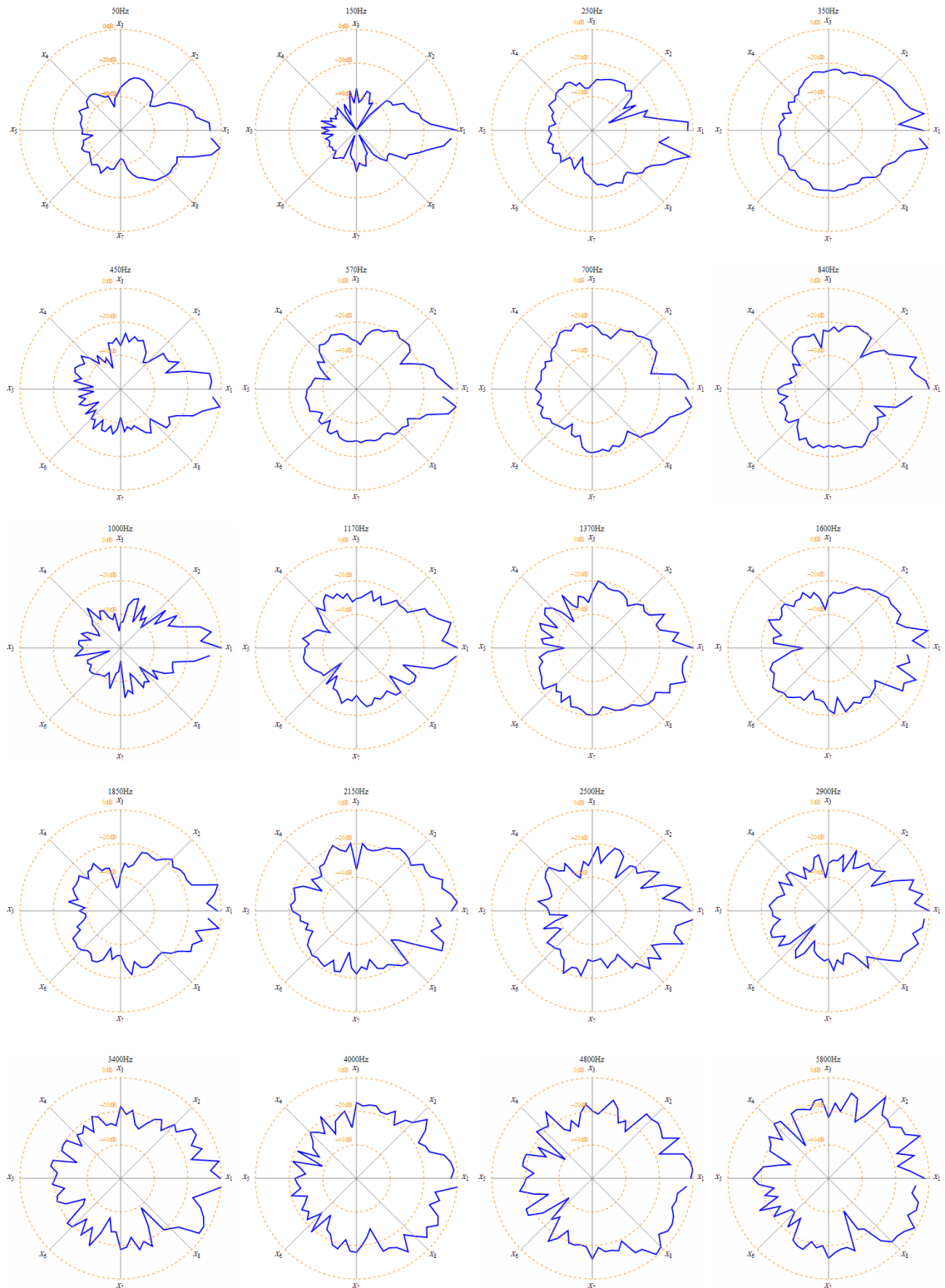
The mean relative phases per frequency at positions 2 to 8 lie between  $6.3\%$  and  $17.2\%$  and have a standard deviation of  $4.86\%$  to  $15.11\%$  of the original phase. The average of all mean values is  $11.84\%$  with an average standard deviation of  $8.93\%$ . Lower frequencies show a quite similar phase per direction and a narrow deviation. At frequencies of  $4\text{kHz}$  and more, the relative phase as well as the standard deviation are notably higher. This corresponds to the measured amplitudes. However, these relative phases are referred to the measured phase at listening position one. Thus, small misplacements of the microphones or noise from the loudspeakers and in the microphone recordings can corrupt the data massively. Although some deviate extremely, they cannot be completely wrong since, due to interference, erroneously created phases would result in incorrect amplitudes which is definitely not the case. Evaluating these “percentages”, one must bear in mind that  $\pm 50\%$  already corresponds to a phase inversion and thus the worst case. For random phase relations the mean value of absolute phase differences should lie around  $25\%$ .

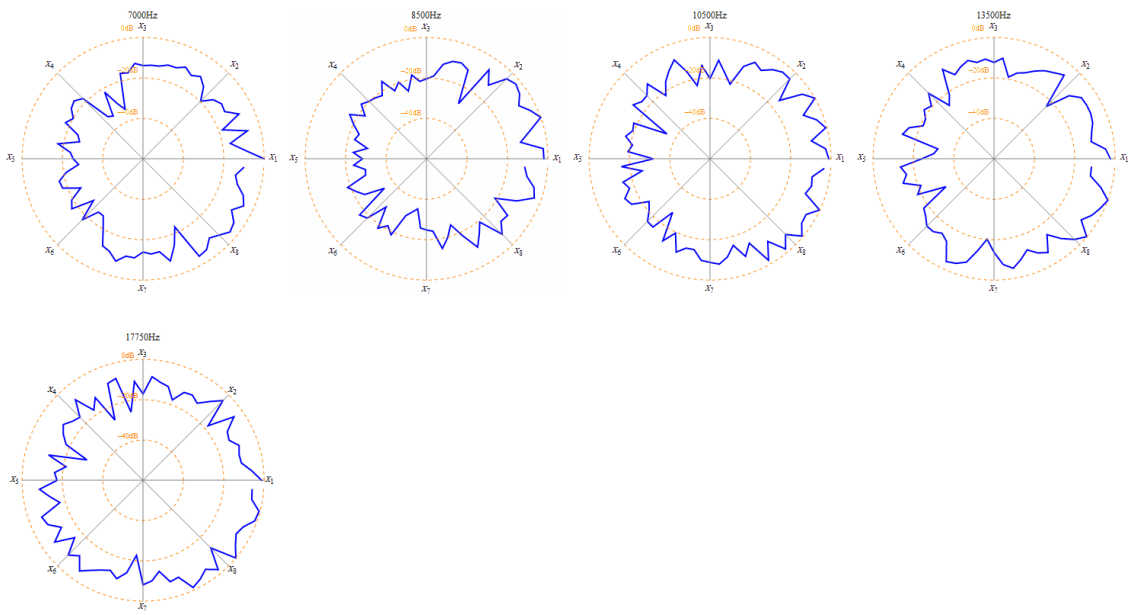
Theoretically, an amplitude of  $0\text{dB}$  at listening positions 1 should lead to a value of  $-\infty$  at positions 2 to 8. Considering the signal to noise-ratio of the microphones, the distinctly audible noise of the loudspeakers and the neglect of diffraction in the calculation, the achieved mean amplitudes of  $-22.85\text{dB}$  at listening positions 2 to 8 are excellent. The radiation patterns of the loudspeakers become much more diverse for every angle at higher frequencies. As the wave length decreases, small misplacements of the loudspeakers or microphones lead to increasing errors.

$f$ in Hz	$\phi_{\bar{X}_1}$	$\phi_{\text{rel}, \bar{X}_2}$ in %	$\phi_{\text{rel}, \bar{X}_3}$ in %	$\phi_{\text{rel}, \bar{X}_4}$ in %	$\phi_{\text{rel}, \bar{X}_5}$ in %	$\phi_{\text{rel}, \bar{X}_6}$ in %	$\phi_{\text{rel}, \bar{X}_7}$ in %	$\phi_{\text{rel}, \bar{X}_8}$ in %	$\emptyset$ $ \phi_{\text{rel}} $	$\sigma$ $ \phi_{\text{rel}} $
50	1.39	16.02	20.00	11.38	3.43	-7.35	9.36	-8.55	<b>10.87</b>	<b>5.56</b>
150	1.94	9.68	4.37	16.14	23.74	19.82	-13.08	7.68	<b>13.5</b>	<b>6.87</b>
250	-0.87	14.44	25.25	8.61	3.56	-2.36	4.27	-5.63	<b>9.16</b>	<b>8.18</b>
350	2.26	19.60	6.47	0.87	-4.52	-2.70	-8.70	-1.24	<b>6.3</b>	<b>6.51</b>
450	-0.83	28.32	14.94	-10.45	-4.55	-1.12	15.38	5.00	<b>11.39</b>	<b>9.21</b>
570	0.90	-30.02	-8.43	-16.99	-3.44	-2.62	13.84	12.40	<b>12.53</b>	<b>9.36</b>
700	1.04	-24.87	-2.21	-5.34	16.00	1.57	11.33	1.42	<b>8.96</b>	<b>8.93</b>
840	-1.09	17.94	14.45	2.76	-5.85	-3.92	-11.59	-1.33	<b>8.26</b>	<b>6.4</b>
1000	1.24	2.99	15.91	-13.05	13.06	0.94	-8.23	13.63	<b>9.69</b>	<b>5.78</b>
1170	1.78	6.54	16.06	7.04	-9.60	-11.29	-11.49	-20.04	<b>11.72</b>	<b>4.86</b>
1370	-1.80	-12.51	-1.69	-15.68	-3.34	2.61	18.74	18.05	<b>10.37</b>	<b>7.6</b>
1600	-0.92	-8.18	1.78	-4.96	20.94	-2.94	16.10	18.87	<b>10.54</b>	<b>7.95</b>
1850	2.40	21.96	-9.71	6.87	28.23	5.97	-9.43	-2.77	<b>12.13</b>	<b>9.33</b>
2150	-2.15	-9.20	-4.57	5.50	12.64	10.72	6.57	-18.75	<b>9.71</b>	<b>4.93</b>
2500	-0.27	13.29	2.38	19.60	-2.43	5.69	7.55	32.36	<b>11.9</b>	<b>10.94</b>
2900	1.00	2.19	-12.76	-3.67	-17.06	-9.53	6.37	1.71	<b>7.61</b>	<b>5.78</b>
3400	2.50	13.38	18.58	10.80	23.06	-6.43	6.63	2.72	<b>11.66</b>	<b>7.24</b>
4000	-2.10	-32.78	-1.19	7.21	8.42	-22.43	1.22	-23.24	<b>13.78</b>	<b>12.34</b>
4800	-0.71	29.84	11.58	-2.97	9.40	25.09	-10.28	26.54	<b>16.53</b>	<b>10.4</b>
5800	-1.02	30.62	9.51	-0.18	33.09	3.91	-12.13	-7.51	<b>13.85</b>	<b>12.9</b>
7000	-0.91	10.11	22.12	-7.07	33.18	-1.74	6.34	7.73	<b>12.61</b>	<b>11.04</b>
8500	0.64	-17.90	-15.78	-9.91	5.09	-32.47	-20.73	-0.05	<b>14.56</b>	<b>10.75</b>
10500	-2.35	-6.34	-32.16	-14.32	-24.42	2.58	-0.35	-24.92	<b>15.01</b>	<b>12.42</b>
13500	2.20	10.27	24.76	32.14	0.87	33.19	-4.90	14.26	<b>17.2</b>	<b>12.98</b>
17750	-0.46	25.35	8.76	6.22	30.20	2.71	0.42	38.67	<b>16.05</b>	<b>21.29</b>
$\emptyset$	—	<b>16.57</b>	<b>12.22</b>	<b>9.59</b>	<b>13.6</b>	<b>8.87</b>	<b>9.4</b>	<b>12.6</b>	<b>11.84</b>	<b>8.93</b>

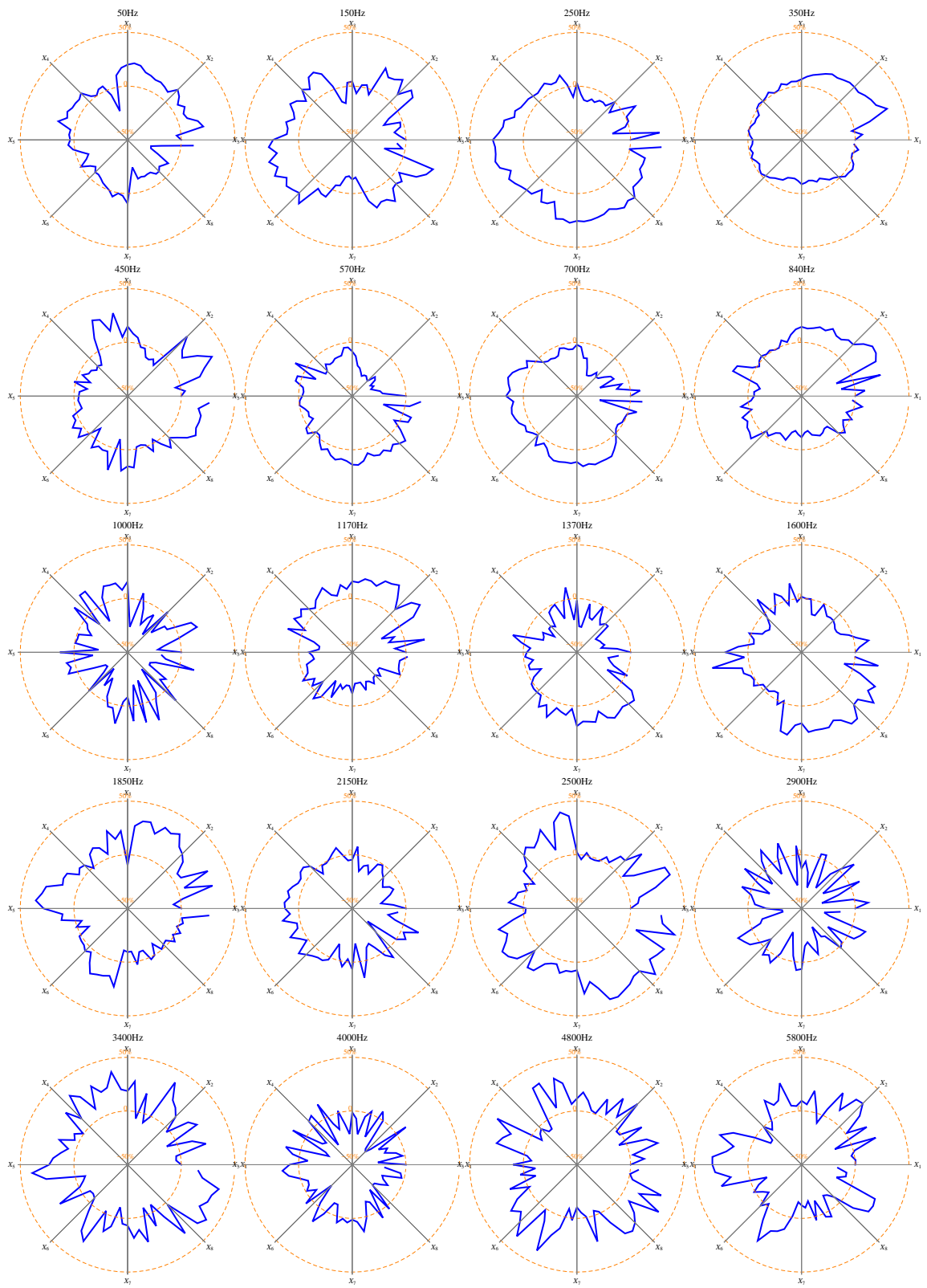
**Table 6.2.:** Measured phases of the octupole sound beam test. Mean relative phase and deviation increase with frequency.

Furthermore, the higher the frequency, the less the assumption of a point source holds true. This explains why the programmed sound beam is much more distinct at low frequencies considering both, amplitude and phase.

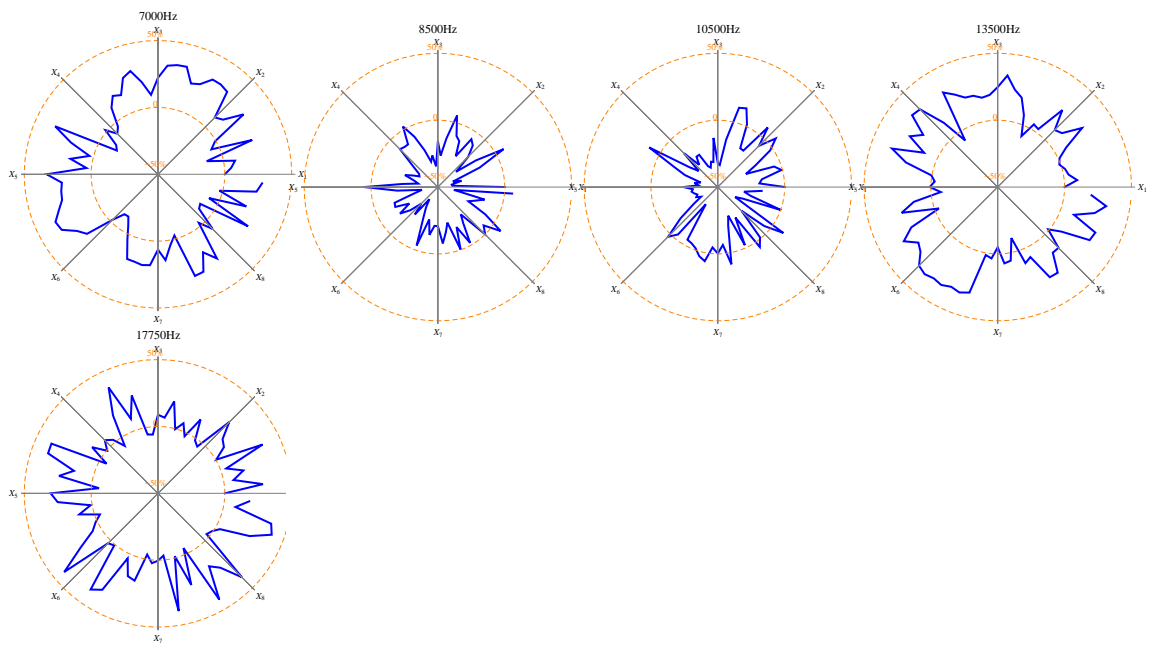




**Figure 6.1.:** Measured amplitudes of the octupole speaker sound beam test. Especially at low frequencies, a clear sound beam is achieved. Even at higher frequencies, the amplitude at listening position 1 is higher than at 2 to 8 in most cases.







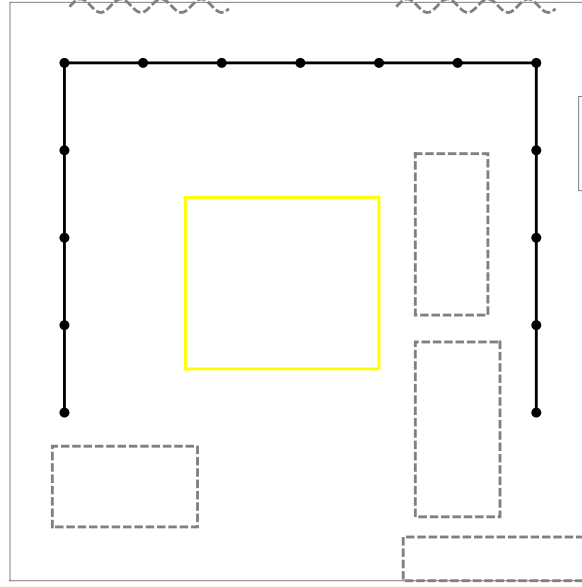
**Figure 6.2.:** Measured relative phases of the octupole speaker sound beam test. At low frequencies, the phase per angle is relatively smooth and homogeneous. At higher frequencies, much more strong fluctuations are visible.

## 6.2. Physical Measurement of the Surrounding Speaker System

The physical performance test on the octupole speaker system proves that the theoretic approach works very well in practice. Even quite extreme radiation characteristics  $\Gamma_{\vec{Q}}(\varphi)$  can be reconstructed accurately at the chosen listening positions. Since both loudspeaker systems are based on the same principles, the test is not repeated for the surrounding speaker system. Instead, it has to undergo several other tests.

The signal processing for the loudspeaker systems assume a free field. The surrounding loudspeaker system is not installed in a free field room but in an ordinary room with parallel, reflecting walls and some furniture. The basic setup has been illustrated already in fig. 5.7 in ch. 5.2.4. The setup, including room, curtains and furniture, is illustrated in fig. 6.3. A complete description of the room as well as a photo are given in appendix A. Such an undamped room with parallel walls is anything but ideal for a sound field reconstruction. Reflections, standing waves and to a minor degree scattering and diffraction disturb the wave field synthesis which only accounts for direct sound in a free field. On the other hand it is quite a natural listening environment similar to a living room. How much the room acoustics degrade the acoustical performance of the speaker system is tested with a simple setup.

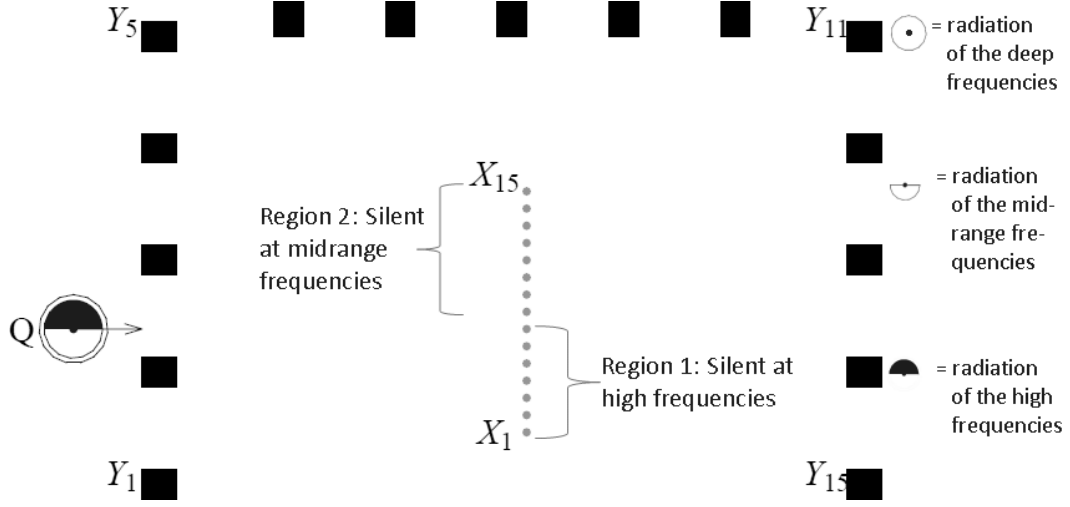
15 listening positions are arranged in a row parallel to and right in between two of the loudspeaker lines. The row of listening positions and the virtual source position build an isosceles, the source faces listening position 8. The source signal is a quasi-stationary signal containing three frequencies  $f_1 = 150\text{Hz}$ ,  $f_2 = 1\text{kHz}$  and  $f_3 = 3.4\text{kHz}$ , representing bass-, midrange-, and treble-frequencies. Each frequency has an individual RC. The complete setup is shown in fig. 6.4, the given magnitudes are listed in eq. 6.5.



**Figure 6.3.:** Listening room including the surrounding loudspeaker setup (connected dots), listening area (solid line), curtains (dashed contour) and tables (dashed rectangles). The dashed rectangle in the lower-right corner is a metal rack.

$$\begin{aligned}
 \vec{Q}(x, y) &= \begin{bmatrix} -0.5\text{m} \\ 1\text{m} \end{bmatrix} \\
 \vec{X}_n(x, y) &= \begin{bmatrix} 1.95\text{m} \\ 0.324\text{m} + n0.05\text{m} \end{bmatrix}, n = 1, 2, \dots, 8 \\
 \Gamma(f_1, \varphi) &= 1 \\
 \Gamma(f_2, \varphi) &= \begin{cases} 1, & \text{between } \varphi_{\vec{X}_1} \text{ and } \varphi_{\vec{X}_7} \\ 0, & \text{otherwise} \end{cases} \\
 \Gamma(f_3, \varphi) &= \begin{cases} 1, & \text{between } \varphi_{\vec{X}_8} \text{ and } \varphi_{\vec{X}_{15}} \\ 0, & \text{otherwise} \end{cases}
 \end{aligned} \tag{6.5}$$

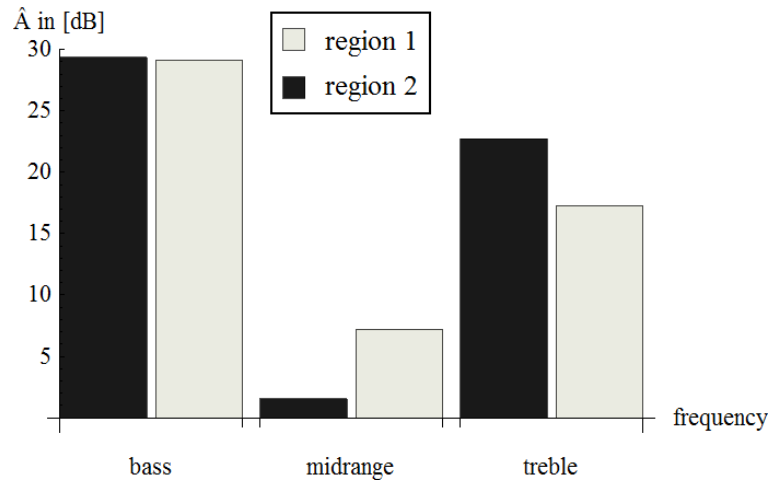
$\Gamma(f_1, \varphi)$  is a monopole and should lead to almost equal amplitudes at all listening positions. The RC for  $f_2$  is a semi-circle, i.e. sound should only reach the listening positions  $\vec{X}_1$  to  $\vec{X}_7$  — summarized as *region 1* — whereas no sound should reach



**Figure 6.4.:** Setup of the physical performance test of the surrounding loudspeaker.

positions  $\vec{X}_8$  to  $\vec{X}_{15}$  — *region 2* — at all. For  $f_3$  it is the opposite way: sound should reach region 2 while region 1 remains silent. The loudspeaker signals are calculated to recreate the correct complex amplitudes at the 15 listening positions under free field conditions. The actual result is recorded at these exact positions. From these recordings, a discrete Fourier transform is performed over the first 80ms. The rest of the signal is ignored to reduce the influence of the reverberation. Reverberation is almost chaotic, no matter if one source is present or more. Hence, the homogeneous distribution of sound pressure strongly reduces the amplitude differences between the regions. Furthermore, the first 80ms are considered to contribute to the impression of the direct sound and thus one might roughly draw conclusions about the perception of sound from the physical results. Due to the small dimensions of the room, the central listening positions, the 15 secondary sources and the small distance between loudspeakers and walls, the first 80ms already contain all first and reflections and many manifold reflections and the time is sufficient for standing waves to occur. The averaged amplitudes per listening region and frequency are plotted in fig. 6.5.

$f_1$  creates an average sound pressure of 29.4dB in listening region 1 and 29.2dB in region 2.  $f_2$  has an average amplitude of 1.5dB in region 1 and 7.2dB in region 2.  $f_3$  has an average sound pressure of 22.7dB in region 1 and 17.2dB in region 2. The result for the bass frequency seems ideal, though the overall sound pressure is



**Figure 6.5.:** Results of the physical measurement of the surrounding speaker system.

quite high. The difference in sound pressure level is only 0.2dB. Due to the large wave lengths, a homogeneous distribution of sound pressure is easy to achieve and not very prone to mistakes due to microphone misplacements. Still, standing waves are distinctly audible and may considerably contribute to the pressure constancy over the listening regions and the overall sound pressure. The low sound pressure in both listening regions for the midrange frequency may be caused by standing waves as well. In contrast to the low frequency, the listening positions seem to lie near a pressure node. The sound pressure in region 2 is 5.7dB higher than in region 1, despite the reflections. This is quite a small difference in comparison to the sound beam with the octupole speaker, which achieved a difference in sound pressure level between 20dB and 40dB in the midrange frequency region. But having in mind that the measured signal contains a high number of early reflections (ER), the difference of almost 6dB is very good. The treble frequency shows a similar performance. It has a difference of 5.5dB between the two regions. The sound pressure level lies between the amplitudes of the deep and the midrange frequency which confirms the idea that standing waves affect deep and midrange frequencies.

Overall, although distinctly worse than the performance of the octupole speaker system in the free field room, the performance of the surrounding speaker system in the undamped room is quite good. The tendencies of the amplitudes per region agree with the predicted calculations for the free field case. Still, the reflections

distinctly influence the sound, leading to a homogenization of the sound pressure of one frequency over the whole room. Standing waves may boost amplitudes of frequencies in some regions and lower them in other regions. On the other hand, the artificial radiation characteristic is very unnatural and extreme. Musical instruments typically have more evenly distributed radiation characteristics and create more similar sound pressures in adjacent areas. Therefore, the actually created amplitudes may deviate less from the desired wave field when synthesizing the sound field of an actual instruments. A psychoacoustic measurement needs to be conducted to ascertain whether the loudspeaker system leads to satisfactory results concerning localization and naturalness of sound. However, the conditions in the room as reflected in the results are not adequate for comparison tests between various radiation characteristics. These would have to take place in a damped room.

### 6.3. Comparison of Radiation Method and Minimum Energy Method

In this section radiation method and minimum energy method are simulated and compared. First, the loudspeaker signals for a virtual monopole are calculated with both methods for the same virtual source position, source signal and listening positions. These should be as small as possible to maximize energy efficiency and to create the lowest possible amount of artifacts by room reflections. Then, the source position is slightly altered. This way, the performance of both methods can be evaluated and compared. Due to the ill posed problem, small changes in source position could lead to high amplitude changes of loudspeaker signals, especially of low frequencies, i.e. of big wave length compared to the distance of the two source positions. Radiation method and MEM are supposed to keep these changes small.

After the evaluation of the loudspeaker signals, the wave fields as created from both methods are analyzed and compared in a simulation. Calculated loudspeaker signals are propagated towards the listening area and beyond to quantify wave field

reconstruction errors between the discrete listening positions and outside of the listening area.

### 6.3.1. Comparison of signal changes for slightly altered source positions

As discussed in ch. 5.2.4, the transfer matrix which describes the propagation of sound from the loudspeakers to the listening positions can be ill-conditioned under certain circumstances. The radiation method is proposed to improve the condition of the matrix. The minimum energy method — as described in ch. 5.2.4 and proposed to improve the conditioning of the transfer matrix of the surrounding speaker system in ch. 5.2.4 — is a promising alternative to the radiation method. Hence, a simulation of both methods is conducted and their results are discussed and compared.

From the simulation several qualitative statements can be derived concerning the amplitudes of the loudspeakers:

1. Magnitude of amplitudes	The lower the reconstruction amplitudes the better
2. Distribution of amplitudes	The closest loudspeaker to the virtual source position should have the highest amplitude
3. Change of amplitudes	Small changes in virtual source position should result in small changes in the loudspeaker signals

**Table 6.3.:** Demands on amplitudes of loudspeakers for a wave field reconstruction.

The lower the amplitudes of the loudspeakers, the more the solution is based on additive interference on the listening positions. This reduces errors of the reconstructed wave field due to reflections. Intuitively, one would expect the maximum sound pressure to come from the loudspeaker which lies closest to the virtual source position. In time-frequency domain, when synthesizing the wave front of a virtual

point source with loudspeakers which are considered as monopoles, this is definitely the case, due to the inverse distance law. But since the surrounding speaker system calculates in frequency domain and inevitably includes signals whose wave fronts reach the listening area from opposing directions, the loudspeaker signals won't resemble the pressure of the source signal propagated to the loudspeaker position. The fact that the loudspeaker signals do not arrive at the listening position simultaneously heavily disturbs the localization. A loud amplitude arriving from a position close to the virtual source position therefore improves the localization. Small changes in source position or radiation characteristic of a source should lead to small changes in the loudspeaker system. If this is not the case, this indicates a badly conditioned transfer matrix.

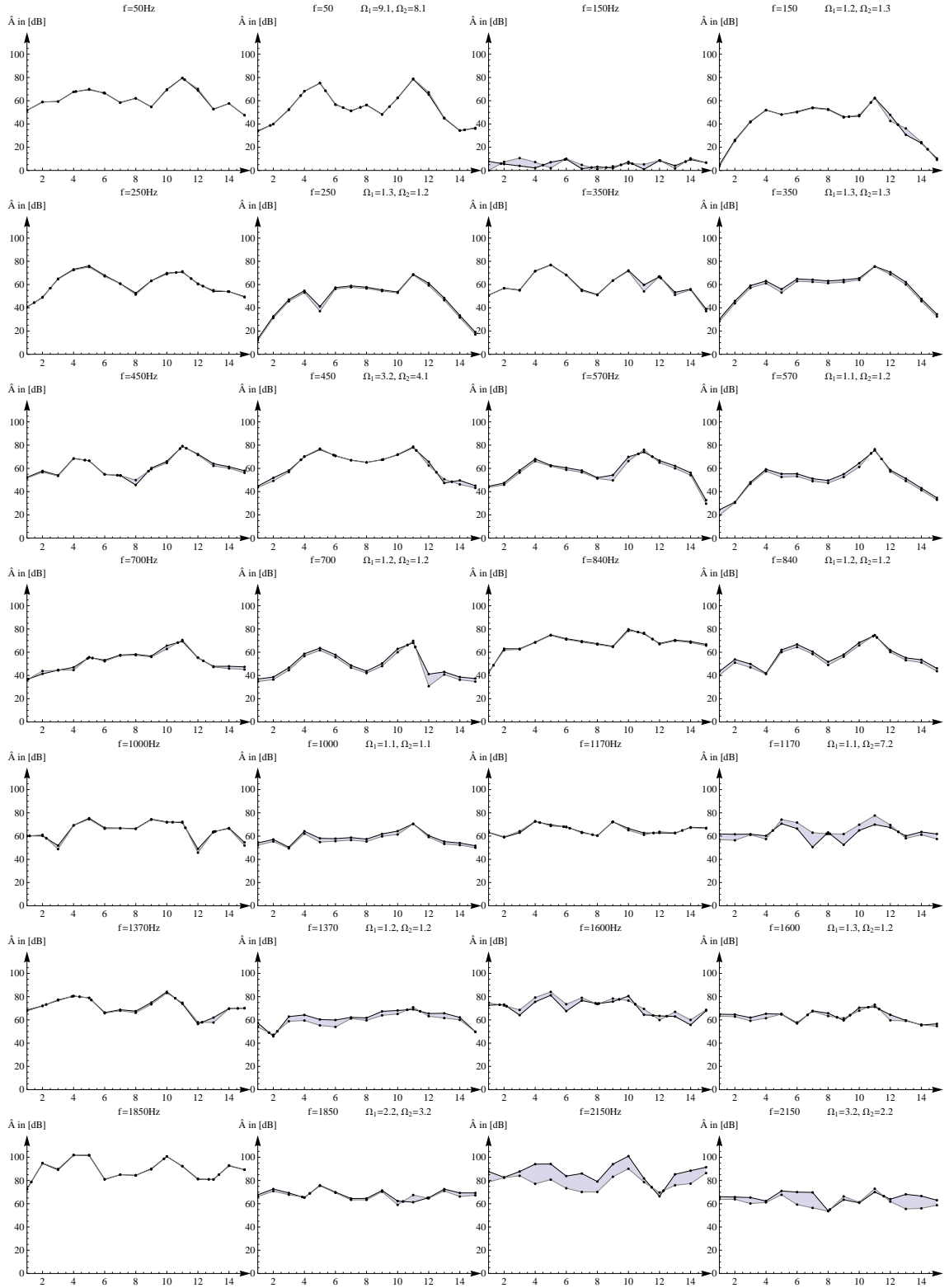
The loudspeaker signals for the wave field reconstruction at 15 listening positions are calculated via radiation method and MEM for the following two virtual source positions of monopole sources:

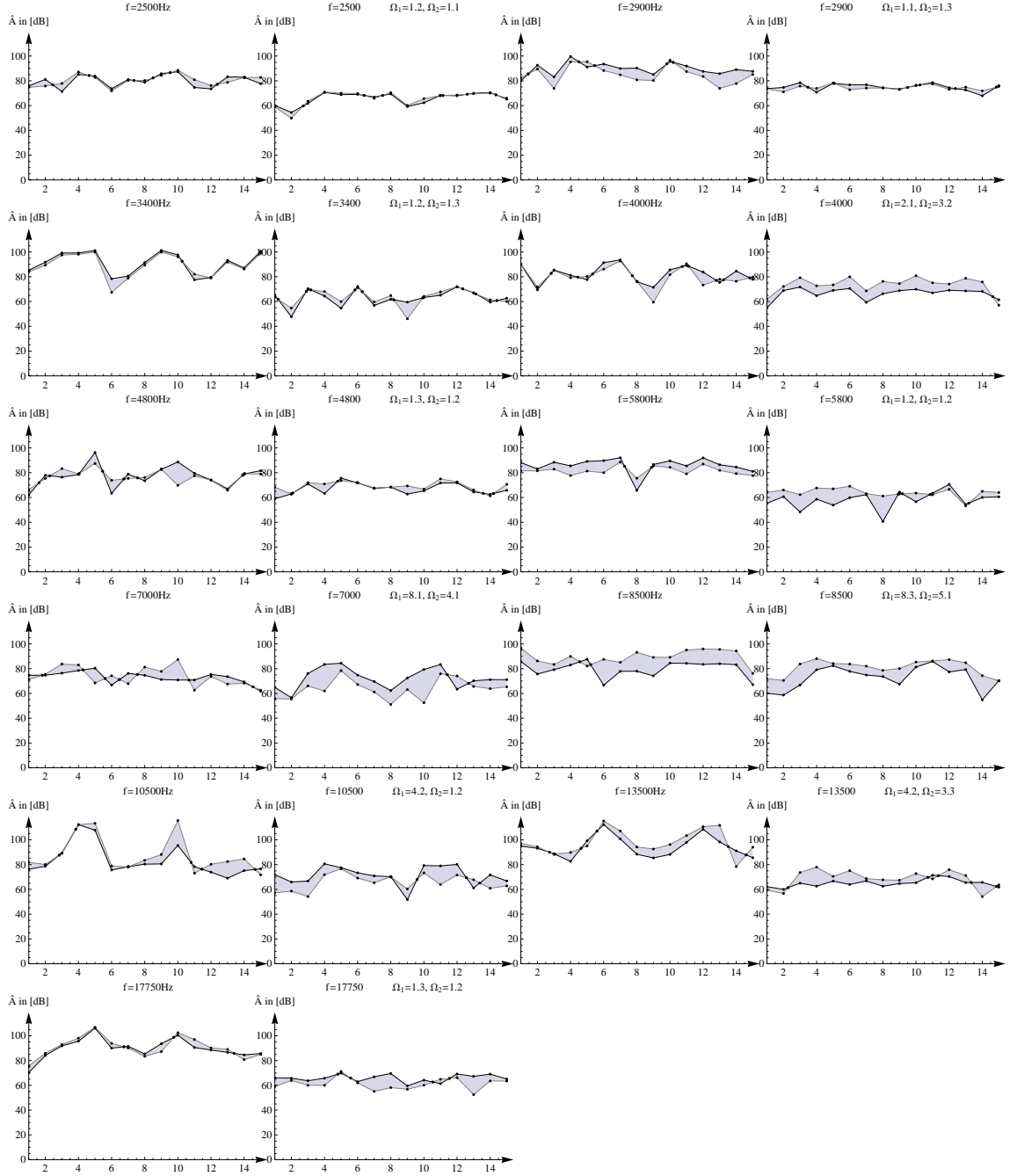
$$\begin{aligned}\vec{Q}_1(x, y) &= \begin{bmatrix} 9\text{m} \\ 3\text{m} \end{bmatrix} \\ \vec{Q}_2(x, y) &= \begin{bmatrix} 9\text{m} \\ 2.9\text{m} \end{bmatrix} \\ \Gamma(\vec{Q}, \varphi) &= 1\end{aligned}\tag{6.6}$$

The solutions of both methods are illustrated for both source positions in fig. 6.6. Radiation method and MEM are illustrated for the same frequencies. The black lines connect the amplitudes for the first virtual source position, the gray lines connect the calculated amplitudes for the second source position. The amplitude difference is highlighted in light gray. The plot labels indicate the frequency and the calculated  $\Omega$  for the MEM.

The relative distribution of loudspeaker amplitudes per frequency are similar for both methods, especially at frequencies 450Hz, 570Hz, 1600Hz, 5800Hz and 8500Hz. The contour of the MEM amplitudes per loudspeaker is more smooth for most







**Figure 6.6.:** Calculated loudspeaker signals via radiation- (left) and minimum energy method (right) for each frequency band for two similar source positions (black and gray lines). Their differences are indicated by the gray filling and should be minimal.

frequencies, compared to the amplitudes of the radiation method. Both methods show only slightly changed loudspeaker amplitudes for the slightly altered source

position. It is conspicuous that via MEM loudspeaker 11 has the highest amplitude of all 15 loudspeaker signals in 11 of 25 frequencies. Via radiation method this is only the case for four frequencies. Loudspeaker 11 lies closest to the virtual source position. As mentioned earlier, it is likely that this improves localization. The calculated amplitudes are summarized in tab. 6.4. The average pressure for source position 1 is listed for the radiation method  $\emptyset P_{1,R}$  and the MEM  $\emptyset P_{1,M}$ . The average absolute pressure for source position 2 is not given but its difference to the pressure of source position 1,  $\emptyset \Delta P_R$  or  $\emptyset \Delta P_M$ , respectively. Also, the standard deviation  $\sigma$  of the amplitudes is given.

Table 6.4 summarizes the calculated pressures of both methods for both source positions. The mean sound pressure level of the loudspeaker signals calculated with MEM are about 10dB lower than calculated via radiation method. Signals of both methods show a similar deviation of sound pressure which corresponds to the similar contours of the calculated amplitudes as already illustrated in fig. 6.6. In both methods the mean amplitude only slightly changes in a magnitude of less than 1dB between the two source positions. With the radiation method, the biggest amplitude change of one frequency from one loudspeaker between the two virtual source positions is 14.33dB. It has a mean change of 7.88dB. With MEM, the biggest amplitude change is 26.86dB. Still, the mean value of all maximum amplitude differences per frequency is less, namely 6.85dB. In general the change in amplitude increases with an increasing difference of  $\Omega$ , as expected. In general, both methods show similar results.

This simulation only informs about two specific sound reconstruction situations. From the results no general conclusion can be drawn about the performance of the two methods. Wave fields in the listening area — i.e. at the calculated listening positions and in between — as calculated from MEM and radiation method are simulated and compared in the following section.

$f$ in Hz	$\emptyset P_{1,R}$	$\sigma P_{1,R}$	$\emptyset \Delta P_R$	$\sigma P_{2,R}$	$ \Delta P_R $ max	$\emptyset P_{1,M}$	$\sigma P_{1,M}$	$\emptyset \Delta P_M$	$\sigma P_{2,M}$	$ \Delta P_M $ max	$\Delta \Omega$
50	61.64	8.58	0.05	8.60	0.58	53.58	14.34	0.18	14.52	0.64	-1.0
150	5.31	2.98	0.41	3.48	7.88	39.87	16.90	-0.31	16.87	5.24	0.1
250	60.49	10.23	-0.33	10.09	1.24	46.85	15.92	-1.58	16.22	4.06	-0.1
350	59.84	10.04	-0.95	10.48	5.58	57.75	12.83	-1.90	13.03	2.91	0.0
450	60.97	8.62	-0.39	8.19	1.86	61.93	11.45	-0.57	12.15	3.45	0.9
570	57.84	10.72	-1.76	11.23	4.60	50.43	13.10	-1.96	13.67	4.72	0.1
700	52.25	8.83	-0.60	8.81	2.90	49.09	10.58	-2.50	11.75	10.40	0.0
840	67.64	8.44	-0.58	8.24	1.29	56.48	9.33	-2.09	9.74	3.25	0.0
1000	64.62	8.06	-0.76	8.96	3.29	58.38	5.28	-1.91	5.43	3.24	0.0
1170	65.30	4.13	-0.09	4.01	1.59	62.38	5.51	1.47	6.92	5.15	6.1
1370	71.37	7.19	-0.51	7.48	4.25	61.48	6.44	-2.58	6.12	5.96	0.0
1600	70.35	7.32	1.99	6.89	3.97	63.31	4.83	-1.21	4.88	4.80	-0.1
1850	89.12	8.49	0.18	8.59	0.56	68.08	4.16	-0.91	3.95	3.31	1.0
2150	86.91	8.09	-8.40	6.12	16.94	65.35	4.44	-4.11	5.14	13.35	-1.0
2500	79.71	5.06	0.60	4.61	5.34	65.62	4.93	0.19	5.76	4.64	-0.1
2900	89.53	4.97	-4.53	7.24	11.90	74.84	2.96	-0.47	2.00	3.97	0.2
3400	90.91	8.95	-1.66	9.63	10.98	62.82	6.62	0.50	6.80	13.41	0.1
4000	82.18	7.35	-2.13	8.48	12.01	66.56	4.62	6.82	6.55	4.40	1.1
4800	77.33	8.89	-1.01	5.99	18.89	66.89	4.66	2.27	3.84	1.96	-0.1
5800	85.84	6.37	-4.27	3.61	9.72	57.96	7.10	5.90	3.70	4.01	0.0
7000	73.01	4.72	0.69	7.71	11.91	72.23	8.28	-8.38	8.09	26.86	-4.0
8500	79.62	6.45	9.67	5.97	5.51	72.65	9.32	8.10	6.18	0.13	-3.2
10500	83.18	12.58	4.36	14.33	5.22	71.16	7.96	-5.57	6.57	15.06	-3.0
13500	94.34	8.56	3.55	9.95	12.71	64.99	3.08	3.23	7.06	11.51	-0.9
17750	89.66	8.13	0.92	8.26	6.34	65.77	3.04	-4.63	4.68	14.76	-0.1
<b><math>\emptyset</math></b>	<b>71.96</b>	<b>7.75</b>	<b>0.22</b>	<b>7.88</b>	<b>6.68</b>	<b>61.46</b>	<b>7.91</b>	<b>0.48</b>	<b>8.07</b>	<b>6.85</b>	<b>-0.16</b>

**Table 6.4.:** Calculated loudspeaker signals via radiation method  $P(x, R)$  and MEM  $P(x, M)$  for two different source positions indicated by subscript  $x$  ( $x = 1 \hat{=} Q_1, x = 2 \hat{=} Q_2$ ).

### 6.3.2. Comparison of Created Wave Fields

Both MEM and radiation method consider the loudspeakers as complex point sources with a certain radiation characteristic  $\Gamma(\varphi)$ . In case of MEM, the radiation characteristic is calculated independently for each frequency band at each time window. The radiation method uses the measured radiation patterns of each frequency band. The different sound radiations lead to different solutions of the linear equation system and thus yield different source signals and different sound field inside and outside the listening area except for the 15 discrete listening positions for which the equation system is solved.

Ideally, the methods create correct wave fields at these positions and in between, i.e. in the complete listening area, for frequencies below the aliasing frequency. In a simulation, wave fields created from the loudspeaker signals resulting from both methods are calculated and compared with each other and the expected original wave field inside and besides the listening area assuming free field conditions. Source position, signal and radiation characteristic are the same as the previous test in ch. 6.3.1:

$$\begin{aligned}\vec{Q}(x, y) &= \begin{bmatrix} 9\text{m} \\ 3\text{m} \end{bmatrix} \\ \Gamma(\omega, \varphi) &= 1\end{aligned}\tag{6.7}$$

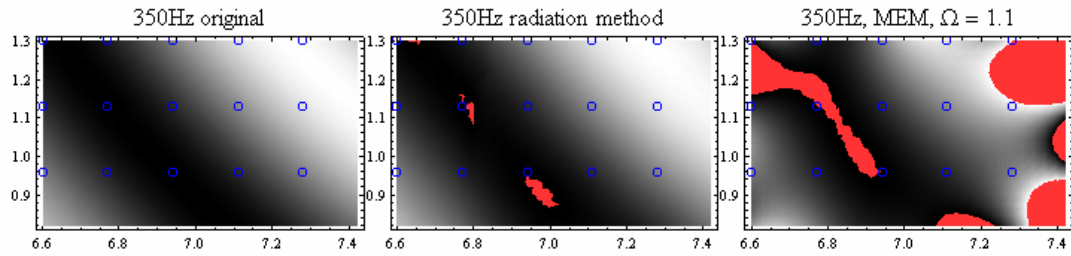
Only the listening positions are altered which explains the different  $\Omega$ -values in solutions of the MEM. In fig. 6.7, the results of the simulation can be seen exemplary for one frequency. The results for all frequencies are illustrated and discussed in more detail in appendix C. Three wave fields are calculated for the 15 listening positions which have a distance of 15cm between adjacent positions to sample the listening area (LA). The wave field is calculated, furthermore, for 5 additional positions in between and beside these listening positions, so in sum the wave field is calculated at 540 positions inside and outside the listening area. Together, all 540 positions constitute the extended LA. A fourth wave field is calculated applying

MEM but solving the linear equation system with  $\Omega$  by 0.1. This narrows the radiation of the speakers which is supposed to improve the condition of the propagation matrix. This procedure has led to better reconstruction results in several other applications.

Table 6.5 summarizes the results of the sound field reconstructions in the LA and in an extended LA. Here, not the real parts are compared, as plotted in fig. 6.7, but the amplitudes of the spectral components.

The critical frequency for the given setup can be calculated from eq. 4.24 given in ch. 4.2.2. It is dependent on the distance of the listening positions and the angle between listening positions and source. The absolute worst case aliasing frequency for this constellation of listening positions lies at about 1.1kHz. The aliasing frequency for the constellation with these listening positions and the given source positions lies between 1.4kHz and 1.9kHz. It can be seen that all methods work best below the critical frequency. That is why the mean synthesis errors of frequencies below the aliasing frequency are given in the table. As expected, the mean synthesis error is smaller in the listening area than in the extended listening area. Remembering from ch. 2.1 that the JND in amplitude change is about 0.8dB the errors are expected to be audible except for the 7 to 8 lowest frequency bands. MEM with both  $\Omega$ -values result in similar loudspeaker signals and also create errors of similar magnitude which lie about 15dB higher than the synthesis errors produced by the radiation method. This is the case for frequencies below and above the aliasing frequency inside the main as well as the extended listening area.

In conclusion the radiation method seems to be superior to MEM for the purpose of sound field synthesis because it takes the actually measured radiation characteristics into account and is thus closer to reality than MEM. Both methods achieve a similar robustness but the synthesis error in the listening area and beyond is much smaller for the radiation method. Therefore, the radiation method is applied for the following listening test in ch. 6.4.1.



**Figure 6.7.:** Exemplary comparison of simulated real part of the wave field from the original source (left) and the wave field recreations by radiation method (center) and minimum energy method (right). The circles indicate the 15 discrete listening positions. The contrasted areas denote clipping, i.e. real parts that lie above the highest or below the lowest real value of the original sound field.

$f$ in Hz	$\emptyset P - P_R$ extended LA in dB	$\emptyset P - P_R$ LA in dB	$\emptyset P - P_{M_1}$ extended LA in dB	$\emptyset P - P_{M_1}$ LA in dB	$\emptyset P - P_{M_2}$ extended LA in dB	$\emptyset P - P_{M_2}$ LA in dB
50	0.04	0.02	0.02	0	0.01	0
150	1.9	0.61	0.08	0.02	0.1	0.02
250	0.09	0.06	0.18	0.04	0.2	0.04
350	0.13	0.09	4.78	1.18	4.35	0.89
450	0.16	0.09	9.73	2.72	9.23	2.34
570	0.87	0.41	13.92	4.6	13.27	4.26
700	0.31	0.15	20.15	10.11	19.49	9.99
840	2.24	0.84	22.53	15.02	22.43	14.69
1000	5.36	4.91	16.15	15.34	18.4	17.83
1170	5.39	4.74	32.17	31.63	32.12	31.55
1370	2.34	1.85	21.9	21.99	22.11	22.18
$\emptyset < f_{\text{alias}}$	<b>1.94</b>	<b>1.45</b>	<b>18.8</b>	<b>16.4</b>	<b>18.84</b>	<b>16.57</b>
1600	5.16	4.78	22.27	22.07	22.35	22.12
1850	2.8	2.97	24.46	22.07	24.87	22.41
2150	9.13	8.33	32.86	32.48	33.24	32.84
2500	8	7.56	20.71	21.01	20.75	21.07
2900	5.95	5.98	30.17	30.4	30.56	30.75
3400	5.05	5.04	22.45	22.37	22.36	22.23
4000	16.37	16.17	40.07	40.08	39.85	39.87
4800	21.24	21.13	29.47	29.56	27.5	27.57
5800	7	6.66	25.91	25.77	26.54	26.38
7000	5.23	5.19	22.84	22.27	22.49	21.9
8500	11.58	11.35	29.99	29.46	29.72	29.16
10500	18.6	16.56	13.62	13.72	12.83	12.95
13500	18	17.66	18.25	18.24	17.42	17.45
17750	11.42	11.51	11.11	11.09	11.06	11.01
$\emptyset > f_{\text{alias}}$	<b>12.42</b>	<b>11.97</b>	<b>27.77</b>	<b>27.57</b>	<b>27.63</b>	<b>27.4</b>
$\emptyset f_{\text{all}}$	<b>9.22</b>	<b>8.76</b>	<b>24.87</b>	<b>24.24</b>	<b>24.77</b>	<b>24.14</b>

**Table 6.5.:** Wave field errors in simulation for the radiation method and the MEM with original and slightly increased  $\Omega$ -values in the listening area and the extended listening area



## 6.4. Psychoacoustic Investigations of the Surrounding Speaker System

In this section, the performance of the surrounding loudspeaker system is evaluated by means of listening tests. First, localization, perceived distinctness, naturalness and spaciousness are investigated. They are compared between mono, stereo and the introduced sound field synthesis approach with and without the precedence fade as described in ch. 5.3.

### 6.4.1. A comparison of Spectral Psychoacoustic Wave Field Synthesis, Stereo and Mono

To test the performance of the developed wave field synthesis application for the recreation of radiation characteristics of musical instruments in a natural listening environment, a listening test is conducted. The test compares mono, stereo and the psychoacoustic wave field synthesis approach with and without the precedence fade as discussed in ch. 5.3. The test takes place in the same room as the second physical measurement described previously in this chapter.

Two groups of five musicology students are advised to stand in the listening area, locate the sound and mark the perceived source position(s) or area(s). The probands mark their own position in the room and their viewing direction. They stand still but are allowed to move trunk and head. The sound signal is a 30sec sample of a Bach sonata for violin (BWV 1001 adagio) in a dry recording, presented successively in mono, stereo, the developed WFS approach without the precedence fading and WFS with precedence fading, each with a different virtual source position. The subjects were not informed about the applied method or the number of active loudspeakers to create the virtual source or phantom source. Music is the most natural and meaningful stimulus for a comparison of audio systems with the aim to recreate the RC of musical instruments. Furthermore, a violin piece is a challenge for the WFS approaches since onsets are not very distinct and due to vibrato

permanent frequency changes occur. Both WFS examples simulate the measured violin radiation characteristic. After each presentation the probands subjectively answer three questions concerning three features of the sound<sup>1</sup>:

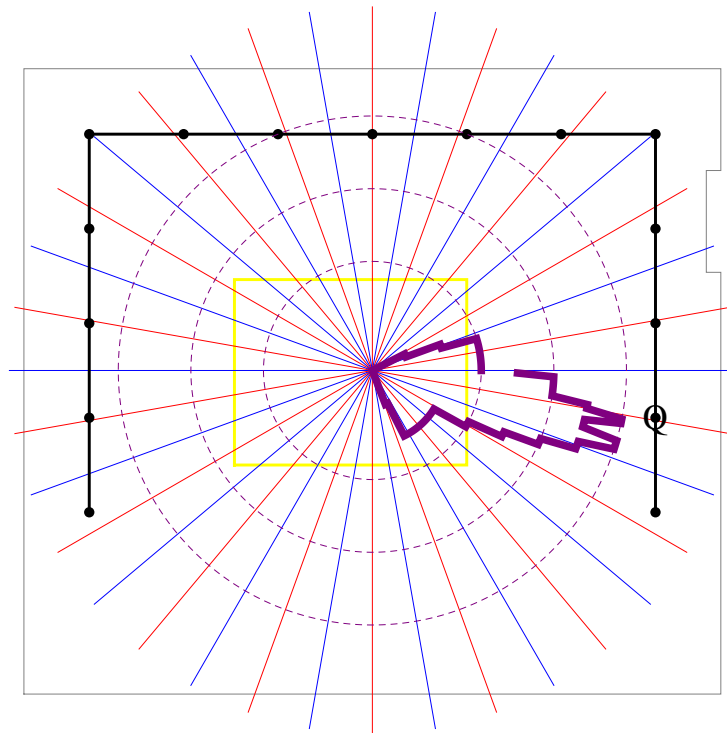
1. How distinctly is the sound locatable and are there several sources?  
(from 1 “not at all” to 5 “very distinctly” and “yes”/“no”)
2. How natural is the sound?  
(from 1 “entirely unnatural” to 5 “entirely natural”)
3. How spatial is the sound? (from 1 “not a bit” to 4 “very spatial”)

These features were not described any further to prevent the subjects from explicitly paying attention to specific sound features. Hence, they evaluate the sound according to their own interpretations. The last question has only four distinctions to force to decide a tendency. The questionnaire and further information about the subjects are presented in appendix D.

Figures 6.8 to 6.11 are polar plots which illustrate the angles of the auditory events from the center of the speaker surrounding as marked by the subjects. The diagram shows the number of subjects who perceived the sound as coming from the respective angle. Dashed circles are at 3, 5 and 7 of 10 subjects. The  $Q$  represents the virtual source position, the solid lines denote the angles each 10°. The stereo basis and the two used loudspeakers for creating the phantom source are indicated by a double arrow. The plots are projected on the listening room for reasons of clarity.

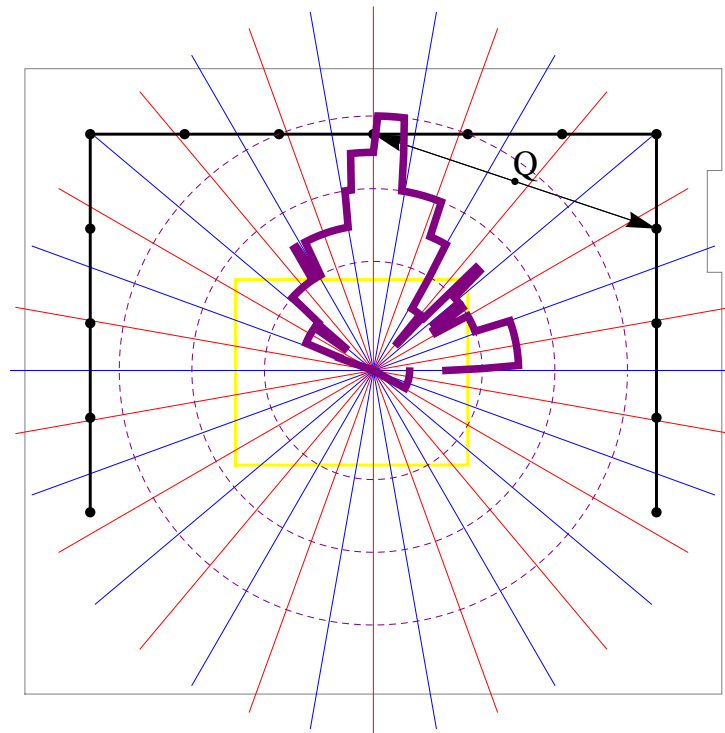
---

<sup>1</sup>Similar parameters were asked in a subjective WFS listening test in Chung et al. (2007).



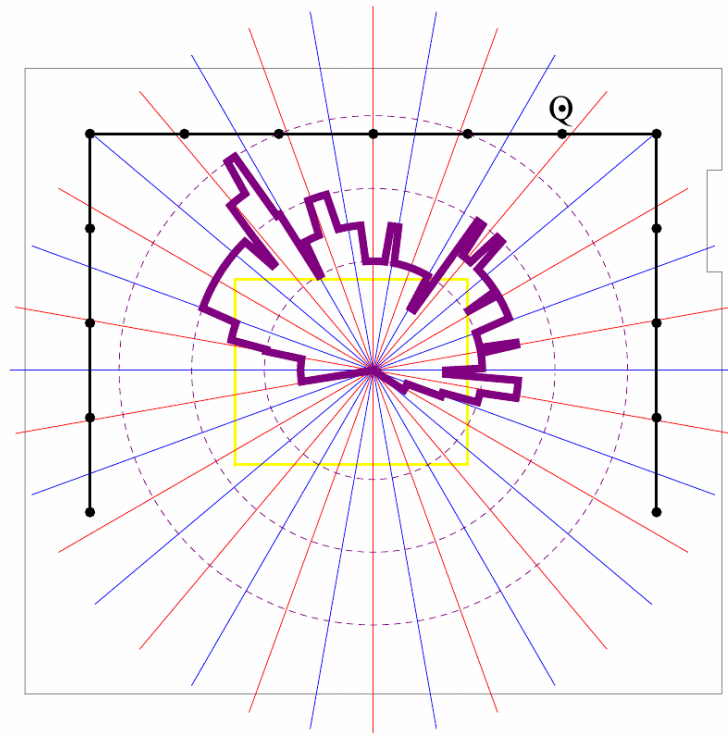
**Figure 6.8.:** Results of the listening test of a mono signal.

Several sources? [0] “Yes” [10] “No”  
 Distinctness: [0] 1 [0] 2 [0] 3 [5] 4 [5] 5  
 Modus: 4 and 5, Median: 4 to 5  
 ar. Mean: 4.5  
 Naturalness: [0] 1 [3] 2 [7] 3 [0] 4 [0] 5  
 Modus: 3, Median: 3  
 ar. Mean: 2.7  
 Spaciousness: [2] 1 [2] 2 [5] 3 [1] 4  
 Modus: 3, Median: 3  
 ar. Mean: 2.5



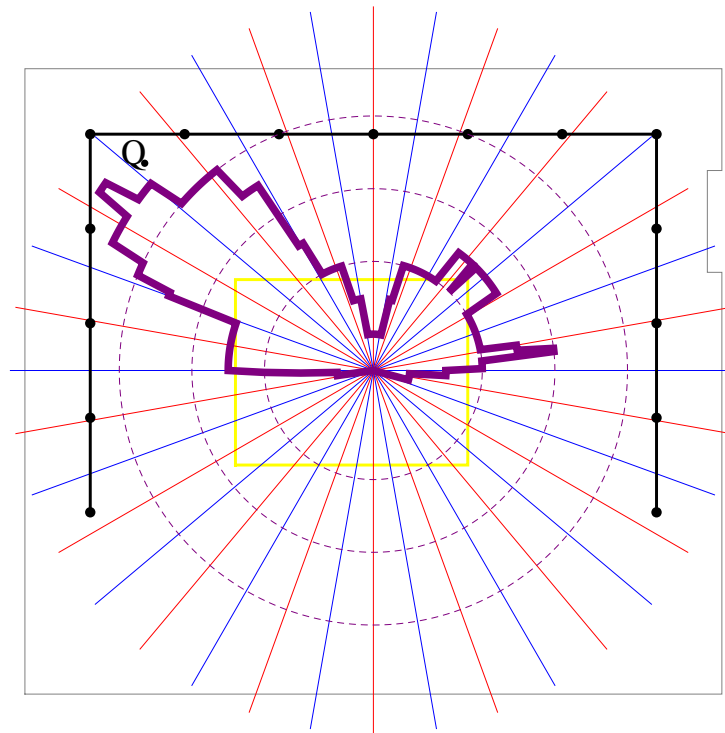
**Figure 6.9.:** Results of the listening test of a stereo signal. The arrows represent the loudspeaker basis and point at the active speakers.

Several sources? [5] "Yes" [5] "No"  
 Distinctness: [0] 1 [0] 2 [6] 3 [2] 4 [2] 5  
 Modus: 3, Median: 3  
 ar. Mean: 3.6  
 Naturalness: [0] 1 [1] 2 [7] 3 [2] 4 [0] 5  
 Modus: 3, Median: 3  
 ar. Mean: 3.1  
 Spaciousness: [0] 1 [4] 2 [5] 3 [1] 4  
 Modus: 3, Median: 3  
 ar. Mean: 2.7



**Figure 6.10.:** Results of the listening test of the wave field synthesis signal without the precedence fade.

Several sources? [9] "Yes" [1] "No"  
 Distinctness: [0] 1 [2] 2 [6] 3 [2] 4 [0] 5  
 Modus: 3, Median: 3  
 ar. Mean: 3  
 Naturalness: [1] 1 [6] 2 [3] 3 [0] 4 [0] 5  
 Modus: 2, Median: 2  
 ar. Mean: 2.2  
 Spaciousness: [0] 1 [1] 2 [6] 3 [3] 4  
 Modus: 3, Median: 3  
 ar. Mean: 3.2



**Figure 6.11.:** Results of the listening test of the wave field synthesis signal with the precedence fade.

Several sources? [8] “Yes” [2] “No”  
 Distinctness: [0] 1 [7] 2 [1] 3 [2] 4 [0] 5  
 Modus: 2, Median: 2  
 ar. Mean: 2.5  
 Naturalness: [2] 1 [3] 2 [3] 3 [2] 4 [0] 5  
 Modus: 2 and 3, Median: 2 to 3  
 ar. Mean: 2.5  
 Spaciousness: [1] 1 [1] 2 [4] 3 [4] 4  
 Modus: 3 and 4, Median: 3  
 ar. Mean: 3.1

The arithmetic mean is only valid if the level of measurement is assumed to be an interval scale, which is probably not the case in this situation with vague terms and

only 4 to 5 possible answers. Thus, modus and median are more suitable and valid to describe the mean estimation.

As can be seen in fig. 6.8, the mono source is located quite correctly with a small range of perceived auditory event angles. All subjects perceived one distinct source position with a medium degree of naturalness and small to medium degree of spaciousness.

The range of located auditory event angles in case of the stereo example is larger than for mono. This coincidences with the inferior perceived distinctness. Furthermore, half of the probands perceived several sources. Most salient of the stereo example is that the commonest perceived angle is not in the region of the phantom source but of the active speaker which is closer to the listening area. The sound is rated more naturally and spaciouly sounding than the mono sound.

The WFS approach without the precedence fade leads to the impression of several sources for most subjects. The distinctness is medium but the perceived source angle lies between  $-100^\circ$  and  $122^\circ$ . Only in a small range more than half of the students presumed a source position which was not even close to the virtual source position. Many subjects marked single speakers. It is almost a random distribution. The sound is perceived quite unnatural but very spatial.

Figure 6.11 shows that WFS with the precedence fade leads to a rather correct localization of the virtual source for most subjects. But still eight of ten perceived more than one source. This is slightly less than without the precedence fade. A second perceived source region protrudes  $90^\circ$  shifted. The overall distribution is wider than in case of stereo or mono but less than without the precedence fade. It has distinct heaps and valleys. The degree of spaciousness is similar to the WFS approach without the precedence fade. This is an exceptional finding since distinctness in locatability is usually considered as the reason for a small perceived source width<sup>2</sup>. The spaciousness is higher than for mono and stereo but at the cost

---

<sup>2</sup>See e.g. Ahrens (2012), pp. 198ff.

of distinctness. The naturalness improved compared to the WFS approach without the precedence fade. But it stays smaller than for the conventional audio systems.

**Locatability:** The fact that the single speaker is located best and has the most distinct direction can be explained by many reasons. Since the virtual source position is actually the real source position, wave front curvature, radiation direction, amplitude, phase and arrival time at any point in the room portend to the source position. Stereo and WFS create partly wrong propagation directions plus unnatural wave fields beyond the discrete listening positions. These may be one reason for the deviant auditory event direction. Furthermore, only the room reflections of the mono source conform the reflections of a source at that position. But the big localization blur even of the mono source — although head-movements were allowed during listening — in comparison to the presented listening tests in free field rooms, as presented in ch. 2.3.2, indicates a distinct negative influence of the locatability due to very early reflections and standing waves. The earliest reflection is so early and undamped that a sort of time- and amplitude based panning between the loudspeaker position and the mirror source position arises which is dependent on listening position. This might explain the high number of slightly false localizations. Two slightly different positions are perceived as source positions for 7 subjects although every subject stated that only one source seems to exist.

Stereo and WFS create many more ER which entirely differ from the natural reflections of a real source at the virtual source position. The more speakers involved the more unnatural reflections occur which corrupt the localization. The phantom source of the stereo example seems additionally to be a “victim” of the precedence effect. It is located at the speaker position which is closer to the listening area. The WFS approach without consideration of the precedence effect leads to a slightly worse distinctness than stereo but in both cases the commonest perceived source position is the closest loudspeaker position.

This is not the case for WFS with precedence fade but this approach apparently suffers under the missing head related filtering-cues. Many subjects locate two sources: The correct virtual source position and a position mirrored at the axis



of their ears. In the tested loudspeaker setup, misplacements of the loudspeakers up to some centimeters are possible due to imprecisions of the used pedestals and brackets. Research of Ahrens et al. showed that for a WFS setup even loudspeaker misplacements of 1cm may lead to audible effects such as perceived changes in timbre, distance or apparent size of the virtual source<sup>3</sup>. So there is a chance that a more accurate setup may improve the naturalness as well as the localization of the virtual source.

**Naturalness:** The stereo sample is perceived as most natural. The reason for that may be familiarization with stereo. Furthermore, in the mono and stereo example phases in the loudspeaker signals are the same as in the recording. Only the overall amplitude of the stereo example is not the same between both speakers though they are very similar in this case, where the phantom source position lies quite central between the two active loudspeakers.

The perceived naturalness of the WFS examples is worse. This may be caused by unnatural amplitude and phase relations beyond the calculated listening spots. Furthermore, changing phases sometimes leads to an unnatural sound impression since phase relations are audible especially during the transients. E.g. it may happen that sharp attacks blur or seem to fade. This also explains why this effect is worse for the unfaded WFS. It radiates up to 15 unnatural sounding transients whereas in the other approach it is only one.

**Spaciousness:** Although no reflections are modeled, a spaciousness occurred, being fewest for mono, slightly more for stereo and most for WFS. This effect is supposed to occur by synthesizing the radiation characteristic of an instrument. But it is not tested yet whether the perceived spaciousness becomes smaller for “unspacious” sounding radiations like a monopole. Furthermore, WFS is not yet compared to other examples using all 15 speakers, e.g. 15 equal signals or 15 randomly phased signals. Therefore, no statement of the controllability of spaciousness can be made at this point.

---

<sup>3</sup>See Ahrens et al. (2010).

## 7. Discussion and Prospects

Music and space are associated in many ways. Frequency discrimination, the perception of beatings and roughness and phenomena like masking result, among other things, from the spatial distribution of the traveling wave inside the cochlea. The ability of sound source localization and identification are complex processes. On the one hand, these processes are affected by auditory scene analysis, on the other hand, they affect it, as well. Without grouping in terms of auditory stream segregation, attributes like timbre, rhythm or melody could neither be composed nor perceived in a musical manner.

Already the pure direct sound of musical instrument has both spatial physical properties and psychoacoustic qualities. Wave field synthesis has the potential to implement a number of these parameters by synthesizing a physically correct reproduction of original wave fields. Many applications exist already and implementing room acoustics and the radiation characteristics of musical instruments is still a matter of research.

In this thesis, a spectral psychoacoustic wave field synthesis approach has been derived and implemented in two loudspeaker systems. It can be traced back to the discrete Rayleigh I integral to create a desired sound field within a listening area by superposition of loudspeaker signals. The necessary calculations are done in frequency domain. Several psychoacoustic considerations are applied: The spectrum of the input signal is divided into 25 critical frequency bands according to the Bark scale. Only the loudest frequency of each frequency band is processed. The others are considered as being partly or completely masked. The dimensions of the listening area and the distance between adjacent loudspeakers is chosen in accor-

dance with integration times of the auditory system to provide for a perceptually satisfying sound field despite different arrival times and directions of wave fronts.

Fading in all but one loudspeaker signal during onsets improves localization. This precedence fade only works because the loudspeaker signals stem from the same input signal and are therefore so similar that they are likely to be integrated into one stream by the auditory system and thus have a common group location. The fading also improves naturalness of the sound because artifacts at onsets are softened due to the fading. Compared to other wave field synthesis system the developed systems have a low number of speakers. This reduces the necessary computational power and allows for an easy installation in common rooms like living rooms.

The main aim of the two loudspeaker systems is to reconstruct the sound radiation characteristics of musical instruments as measured from anechoic far field recordings from 128 angles in horizontal plane. By considering the musical instruments as complex point sources with an angular weighting factor  $\Gamma(\omega, \varphi)$ , the propagated wave field of such a source with any arbitrary input signal can be calculated and reconstructed at as many listening points as loudspeakers present. The complex point source model is also the basis of the radiation method which is applied to improve the condition of the propagation matrix and to relax the linear equation system to be solved. By sampling the listening volume instead of the separation plane between source volume and listening volume, a trade-off between maximum frequency and listening area size is possible: The aliasing frequency of the loudspeaker system can be increased at the cost of extent of the listening area.

That the presented wave field reconstruction principle works could be demonstrated by physical measurements under laboratory conditions, i.e. in a free field room, as well as under natural listening conditions, i.e. in a furnished room with the dimensions of a living room. Simulations have revealed the precision of the reconstruction inside and slightly beside the sampled listening area. In listening tests with musical material it has been demonstrated that this wave field synthesis approach increases the perceived spaciousness of direct sound while remaining a fairly natural sound impression and a good source localization when applying the precedence fade. The fact that the perceived naturalness and spaciousness almost stay the same with

and without the precedence fade, although the localization clearly improves when applying it, indicates that the perception of spaciousness is not bounded to distinctness and correctness of localization. Also the radiation method excelled in simulations and listening tests.

In principle, the presented wave field synthesis method is compatible to other audio systems with discrete channels, such as the stereo triangle, the 5.1 setup of Dolby Digital or Ambisonics. It does not have to be applied as standalone audio system but can also be implemented in other systems as well. Above the aliasing frequency, conventional wave field synthesis systems either apply OPSI or the sub band approach — which lead to coloration effects and/or restrict the listening area to one sweet spot — or the aliasing wave fronts are simply resigned to since they do not harm the perceived sound coloration and localization much. However, the presented approach could be implemented alternatively, being able to recreate the spectrum of high frequencies correctly at least in a “sweet area” rather than a sweet-spot only. Misleading interaural time difference cues due to the wave fronts that arrive from positions other than the intended source position will occur when the method is applied. But since interaural time differences are utilized for frequencies below a few kHz only, this should not harm the localization precision. The main benefit is the implementation of the radiation characteristics of musical instruments in the sound field synthesis. Conventional wave field synthesis systems typically create virtual monopole sources. In the lower frequency range of many musical instruments, the sound radiation can in fact be considered roughly as monopole-like. At higher frequencies the radiation patterns become more complicated. This is where the spectral psychoacoustic wave field synthesis comes into play, creating complex point sources which approximate this radiation characteristic.

The spectral psychoacoustic wave field synthesis has several general strengths and weaknesses which are summarized in table 7.1.

**Future steps:** Until now, note onsets have been chosen manually or they have been defined as first signal after some samples of silence in the source signal. Applying an onset detection algorithm may automatize this process reliably. An al-

Strengths	Weaknesses
Reconstruction of instruments' RCs possible	Largest listening area dimension $\leq 15.1\text{m}$
Listening area rather than a sweet spot	Audible artifacts during transients
Adjustability of listening area extent and aliasing frequency	Onset detection necessary
Compatibility to other loudspeaker setups, such as stereo, 5.1 or ambisonics	Room acoustics may heavily corrupt the sound fields
Combinability with conventional WFS	
Real time capable	

**Table 7.1.:** Summary of general strengths and weaknesses of spectral psychoacoustic wave field synthesis.

gorithm to treat not explicitly one frequency per fixed critical frequency band but to detect maskers and masking thresholds and thus identify unmasked frequencies to be treated might be closer to the actual perception of listeners and lead to better perceived synthesis results especially concerning naturalness and timbre. Of course, it is the masking threshold of the sound field at the listener rather than of the source signal itself that is crucial. This has to be taken into account like, in principle, already implemented in AC-3.

Developing a solver that solves the linear equation system, eq. 5.8, with minimal manipulation of phase has the potential to create a desired sound field without audible artifacts even for transient source signals. Furthermore, the extent of the listening area may be increased within the given limits by calculating the desired wave field for even more listening points than loudspeakers present. Then, an over-determined equation system needs to be solved to calculate the loudspeaker signals. This is possible because the propagation matrix is fully populated.

Implementing the presented spectral psychoacoustic WFS in conventional WFS for frequencies above the aliasing frequency may reveal an improvement of naturalness and an increase of perceived spatial extent of virtual sources. This would have to be investigated in listening tests. Also, comparisons with OPSI and the sub band approach would be interesting. If naturalness can be improved, listening tests with constant source signals but different RCs might finally reveal the direct relationship

between radiation characteristics and apparent source width or other impressions of source extent.

# Bibliography

- Ahrens, Jens, *Analytic Methods of Sound Field Synthesis*, Berlin, Heidelberg: Springer 2012.
- Ahrens, Jens and Spors, Sascha, “Analytical driving functions for higher order Ambisonics”, in: *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 373–376.
- Ahrens, Jens and Spors, Sascha, “Sound Field Reproduction Employing Non-Omnidirectional Loudspeakers”, in: *Audio Engineering Society Convention 126*, 2009a.
- Ahrens, Jens and Spors, Sascha, “Spatial Encoding and Decoding of Focused Virtual Sound Sources”, in: *Ambisonics Symposium*, Graz 2009b.
- Ahrens, Jens, Geier, Matthias and Spors, Sascha, “Perceptual Assessment of Delay Accuracy and Loudspeaker Misplacement in Wave Field Synthesis”, in: *Audio Engineering Society Convention 128*, 2010.
- Ajdler, Thibaut et al., “Sound Field Analysis along a Circle and its Application to HRTF Interpolation”, in: *J. Acoust. Soc. Am.* 56(3), 2008: pp. 156–175.
- Albrecht, Bernhard et al., “An Approach for Multichannel Recording and Reproduction of Sound Source Directivity”, in: *Audio Engineering Society Convention 119*, 2005.
- Allen, Jont B., “Nonlinear Cochlear Signal Processing and Masking in Speech Perception”, in: Jacob Benesty, M. Mohan Sondhi and Yiteng Huang (Eds.), *Handbook of Speech Processing*, chap. 03, Berlin, Heidelberg: Springer 2008, pp. 27–60.

- Avizienis, Rimas et al., “A Compact 120 Independent Element Spherical Loudspeaker Array with Programable Radiation Patterns”, in: *Audio Engineering Society Convention 120*, 2006.
- Baalman, Marije, *On Wave Field Synthesis and electro-acoustic music, with a particular focus on the reproduction of arbitrarily shaped sound sources*, Saarbrücken: VDM 2008.
- Backus, John, *The Acoustical Foundations of Music*, New York: W. W. Norton & Co. 1969.
- Bader, Rolf, “Reconstruction of radiating sound fields using minimum energy method”, in: *J. Acoust. Soc. Am.* 127(1), 2010: pp. 300–308.
- Bader, Rolf, *Nonlinearities and Synchronization in Musical Acoustics and Music Psychology*, Berlin Heidelberg: Springer 2013.
- Bader, Rolf, “Microphone Array”, in: Thomas D. Rossing (Ed.), *Handbook of Acoustics*, Berlin Heidelberg: Springer 2014, pp. 1179–1207.
- Berkhout, A. J., “A Holographic Approach to Acoustic Control”, in: *J. Audio Eng. Soc* 36(12), 1988: pp. 977–995.
- Berkhout, A.J., Diemer de Vries and Vogel, P., “Acoustic control by wave field synthesis”, in: *J. Acoust. Soc. Am.* 93(5), 1993: pp. 2764–2778.
- Blauert, Jens, *Räumliches Hören*, Stuttgart: Hirzel 1974.
- Blauert, Jens, *Räumliches Hören. Nachschrift - Neue Ergebnisse und Trends seit 1972*, Stuttgart: Hirzel 1985.
- Blauert, Jens, *Spatial Hearing. The Psychophysics of Human Sound Source Localization*, revisited edn., Massachusetts: MIT University Press 1997.
- Blauert, Jens, “3-D-Lautsprecher-Wiedergabemethoden”, in: *DAGA 2008 – Dresden*, 2008, pp. 25–26.
- Blauert, Jens and Braasch, Jonas, “Räumliches Hören”, in: Stefan Weinzierl (Ed.), *Handbuch der Audiotechnik*, chap. 3, Berlin, Heidelberg: Springer 2008, pp. 87–122.



- Blauert, Jens and Xiang, Ning, *Acoustics for Engineers. Troy Lectures*, 2nd edn., Berlin, Heidelberg: Springer 2009.
- Bleda, Sergio et al., “An Approach to Discrete-Time Modelling Auralization for Wave Field Synthesis Applications”, in: *Audio Engineering Society Convention 118*, 2005.
- Bregman, Albert S., *Auditory Scene Analysis*, Massachusetts: MIT Press 1990.
- Bruhn, Herbert, “Tonpsychologie—Gehörpsychologie—Musikpsychologie”, in: Herbert Bruhn, Rolf Oerter and Helmut Rösing (Eds.), *Musikpsychologie. Ein Handbuch*, 4th edn., Reinbek bei Hamburg: Rowohlt 2002a, pp. 439–451.
- Bruhn, Herbert, “Verarbeitung einzelner Schallereignisse”, in: Herbert Bruhn, Rolf Oerter and Helmut Rösing (Eds.), *Musikpsychologie. Ein Handbuch*, 4th edn., Reinbek bei Hamburg: Rowohlt 2002b, pp. 666–670.
- Bruhn, Herbert, “Wahrnehmung und Repräsentation musikalischer Strukturen”, in: Herbert Bruhn, Rolf Oerter and Helmut Rösing (Eds.), *Musikpsychologie. Ein Handbuch*, 4th edn., Reinbek bei Hamburg: Rowohlt 2002c, pp. 452–459.
- Bruhn, Herbert and Michel, Dieter, “Hören im Raum”, in: Herbert Bruhn, Rolf Oerter and Helmut Rösing (Eds.), *Musikpsychologie. Ein Handbuch*, 4th edn., Reinbek bei Hamburg: Rowohlt 2002, pp. 650–655.
- Burns, Thomas H., “Sound Radiation Analysis of Loudspeaker Systems Using the Nearfield Acoustic Holography (NAH) and the Application Visualization System (AVS)”, in: *Audio Engineering Society Convention 93*, 1992.
- Cho, Wan-Ho, Ih, Jeong-Guon and Boone, Marinus M., “Holographic Design of a Source Array Achieving a Desired Sound Field”, in: *J. Audio Eng. Soc* 58(4), 2010: pp. 282–298.
- Chung, Hyunjoo et al., “Hybrid Sound Field Processing for Wave Field Synthesis System”, in: *Audio Engineering Society Convention 122*, 2007.
- Corteel, Etienne, “Synthesis of Directional Sources Using Wave Field Synthesis, Possibilities, and Limitations”, in: *EURASIP Journal on Advances in Signal Processing* 2007, 2007.

- Corteel, Etienne and Nicol, Rozenn, “Listening Room Compensation for Wave Field Synthesis. What Can Be Done?”, in: *Audio Engineering Society Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction*, 2003.
- Daniel, Jérôme, “Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format”, in: *Audio Engineering Society Conference: 23rd International Conference: Signal Processing in Audio Recording and Reproduction*, 2003.
- Daniel, Jérôme, Nicol, Rozenn and Moreau, Sébastien, “Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging”, in: *Audio Engineering Society Convention 114*, 2003.
- David jr., Edward E., “Aufzeichnung und Wiedergabe von Klängen”, in: Klaus Winkler (Ed.), *Die Physik der Musikinstrumente*, Heidelberg: Spektrum 1988, pp. 150–160.
- Davis, Mark F., “Audio and Electroacoustics”, in: Thomas D. Rossing (Ed.), *Handbook of Acoustics*, chap. 18, New York: Springer 2007, pp. 743–781.
- Dickreiter, Michael, *Handbuch der Tonstudioteknik*, vol. 1, 2 edn., Munich et al.: De Gruyter 1978.
- Dickreiter, Michael, *Handbuch der Tonstudioteknik*, vol. 1, 5. völlig neu bearbeitete und ergänzte edn., Munich et al.: De Gruyter 1987.
- von Ehrenfels, Christian, “Über Gestaltqualitäten”, in: *Vierteljahrsschrift für wissenschaftliche Philosophie* 14, 1890: pp. 249–292.
- Elen, Richard, “Ambisonics. The Surround Alternative”, 2001, last visited: 22.11.2010, URL <http://www.ambisonic.net/pdf/ambidvd2001.pdf>.
- Fastl, Hugo, “Temporal masking effects: II. Critical band noise masker”, in: *Acoustica* 36, 1977: pp. 317–331.
- Fastl, Hugo, “Praktische Anwendungen der Psychoakustik”, in: *DAGA 2010 – Berlin*, 2010, pp. 5–10.

- Fletcher, Neville H. and Rossing, Thomas D., *The Physics of Musical Instruments*, 2nd edn., New York: Springer 2008.
- Fraunhofer-Institut für Digitale Medientechnologie IDMT, “Jubiläum in Bregenz”, 2014, last visited: 26.04.2016, URL [www.idmt.fraunhofer.de/de/Press\\_and\\_Media/press\\_releases/2014/Zehn\\_Jahre\\_Thueringer\\_3D\\_Klang\\_fuer\\_die\\_groesste\\_Seebuehne\\_der\\_Welt.html](http://www.idmt.fraunhofer.de/de/Press_and_Media/press_releases/2014/Zehn_Jahre_Thueringer_3D_Klang_fuer_die_groesste_Seebuehne_der_Welt.html).
- Friedrich, Hans Jörg, *Tontechnik für Mediengestalter. Töne hören — Technik verstehen — Medien gestalten*, Berlin, Heidelberg: Springer 2008.
- Garner, Wendell R., *The Processing of Information and Structure*, New York, Toronto, London, Sydney: Lawrence Erlbaum 1974.
- Gauthier, Philippe-Aubert and Berry, Alain, “Adaptive Wave Field Synthesis for Sound Field Reproduction: Theory, Experiments, and Future Perspectives”, in: *Audio Engineering Society Convention 123*, 2007.
- Geier, Matthias et al., “Perceptual Evaluation of Focused Sources in Wave Field Synthesis”, in: *Audio Engineering Society Convention 128*, 2010.
- Gelfand, Stanley A., *Hearing. An introduction to psychological and physiological acoustics*, second edn., New York and Basel: Informa 1990.
- Goertz, Anselm, “Lautsprecher”, in: Stefan Weinzierl (Ed.), *Handbuch der Audiotechnik*, chap. 8, Berlin, Heidelberg: Springer 2008, pp. 421–490.
- Hall, Donald E., *Musikalische Akustik. Ein Handbuch*, Mainz: Schott 2008.
- Hammond, Joe and White, Paul, “Signals and Systems”, in: David Havelock, Sonoko Kuwano and Michael Vorländer (Eds.), *Handbook of Signal Processing in Acoustics*, chap. 1, New York: Springer 2008, pp. 3–16.
- Haustein, B. G., “Hypothesen über die einohrige Entfernungswahrnehmung des menschlichen Gehörs”, in: *Hochfrequenztechnik und Elektroakustik* 78, 1969: pp. 46–57.
- Haustein, B. G. and Schirmer, W., “Messeinrichtung zur Untersuchung des Richtungslokalisationsvermögens”, in: *Hochfrequenztechnik und Elektroakustik* 79, 1970: pp. 96–101.

- Horbach, Ulrich et al., “Numerical Simulation of Wave Fields Created by Loudspeaker Arrays”, in: *Audio Engineering Society Convention 107*, 1999.
- IOSONO GmbH, “IOSONO — The future of spatial audio”, 2008, last visited: 23.01.2011, URL <http://www.iosono-sound.com/>.
- Kassakian, Peter and Wessel, David, “Characterization of Spherical Loudspeaker Arrays”, in: *Audio Engineering Society Convention 117*, San Francisco 2004.
- Kim, Youngtae et al., “Optimal Filtering for Focused Sound Field Reproductions Using a Loudspeaker Array”, in: *Audio Engineering Society Convention 126*, 2009.
- Kling, J. W. and Riggs, Lorrin A. (Eds.), *Woodworth & Schlossberg’s Experimental Psychology*, 3rd edn., New York: Holt, Rinehart and Winston 1971.
- Kolundzija, Mihailo, Faller, Christof and Vetterli, Martin, “Designing Practical Filters for Sound Field Reconstruction”, in: *Audio Engineering Society Convention 127*, 2009a.
- Kolundzija, Mihailo, Faller, Christof and Vetterli, Martin, “Sound Field Reconstruction: An Improved Approach For Wave Field Synthesis”, in: *Audio Engineering Society Convention 126*, 2009b.
- Kostek, Bozena, *Perception-Based Data Processing in Acoustics*, Berlin, Heidelberg, New York: Springer 2005.
- Larcher, Véronique and Jot, Jean-Marc, “Techniques d’interpolation de filtres audio-numériques. Application à la reproduction spatiale des sons sur écouteurs”, in: *Congrès Français d’Acoustique, Marseille, France*, Marseille 1999.
- Lerch, Alexander, “Bitdatenreduktion”, in: Stefan Weinzierl (Ed.), *Handbuch der Audiotechnik*, chap. 16, Berlin, Heidelberg: Springer 2008, pp. 849–884.
- Lopez, José Javier et al., “A Sub-band Approach to Wave-Field Synthesis Rendering”, in: *Audio Engineering Society Convention 118*, Barcelona 2005.
- Magalhães, Marcelo Bruno S. and Tenenbaum, Roberto A., “Sound Sources Reconstruction Techniques: A Review of Their Evolution and New Trends”, in: *Acta Acustica United With Acustica* 90, 2004: pp. 199–220.

- Martín, R. San et al., “Influence of the Source Orientation on the Measurement of Acoustic Parameters”, in: *Acta Acustica United With Acustica* 93, 2007: pp. 387–397.
- Mechel, Fridolin, *Room Acoustical Fields*, Berlin, Heidelberg: Springer 2013.
- Mechel, Fridolin P., “General Linear Fluid Acoustics”, in: Fridolin P. Mechel (Ed.), *Formulas of Acoustics*, 2nd edn., chap. B, Berlin, Heidelberg, New York: Springer 2008, pp. 5–58.
- Menzel, D. et al., “Binaurale Raumsyntese mittels Wellenfeldsynthese — Realisierung und Evaluierung”, in: *Fortschritte der Akustik — DAGA '06*, Braunschweig 2006, pp. 255–256.
- Menzies, Dylan and Al-Akaidi, Marwan, “Nearfield binaural synthesis and ambisonics”, in: *J. Acoust. Soc. Am.* 121(3), 2007: pp. 1559–1563.
- Merziger, Gerhard and Wirth, Thomas, *Repetitorium der höheren Mathematik*, 5th edn., Springer: Binomi 2006.
- Meyer, E., Burgtorf, W. and Damaske, P., “Eine Apparatur zur elektroakustischen Nachbildung von Schallfeldern. Subjektive Hörwirkungen beim Übergang Kohärenz–Inkohärenz”, in: *Acustica* 15, 1965: pp. 339–344.
- Meyer, John, Meyer, Perrin and Baird, Justin, “Far-Field Loudspeaker Interaction: Accuracy in Theory and Practice”, in: *Audio Engineering Society Convention 110*, 2001.
- Meyer, Jürgen, “Musikalische Akustik”, in: Stefan Weinzierl (Ed.), *Handbuch der Audiotechnik*, chap. 4, Berlin, Heidelberg: Springer 2008, pp. 123–180.
- Meyer, Jürgen, *Acoustics and the Performance of Music. Manual for Acousticians, Audio Engineers, Musicians, Architects and Musical Instrument Makers*, 5th edn., Bergkirchen: Springer 2009.
- Middlebrooks, John C. and Green, David M., “Sound Localization by Human Listener”, in: *Annu. Rev. Psychol.* 42, 1991: pp. 135–159.

- Müller, Swen, “Measuring Transfer-Functions and Impulse Responses”, in: David Havelock, Sonoko Kuwano and Michael Vorländer (Eds.), *Handbook of Signal Processing in Acoustics*, chap. 5, New York: Springer 2008, pp. 65–85.
- Morikawa, Daisuke and Hirashara, Tatsuya, “Signal frequency necessary for horizontal sound localization”, in: *Acoust. Sci. & Tech.* 31(6), 2010: pp. 417–419.
- Morse, Philip M. and Ingard, K. Uno, *Theoretical Acoustics*, Princeton: Princeton University Press 1986.
- Myers, David G., *Psychologie*, 2. erweiterte und aktualisierte edn., Berlin, Heidelberg: Springer 2008.
- Oellers, Helmut, “Die virtuelle Kopie des räumlichen Schallfeldes”, 2010, last visited: 27.09.2010, URL <http://www.syntheticwave.de/>.
- Ono, Kazuho, Pulkki, Ville and Karjalainen, Matti, “Binaural Modeling of Multiple Sound Source Perception. Coloration of Wideband Sound”, in: *112th Audio Engineering Society Convention*, Munich 2002.
- Otondo, Felipe and Rindel, Jens Holger, “The Influence of the Directivity of Musical Instrument in a Room”, in: *Acta Acustica United With Acustica* 90, 2004: pp. 1178–1184.
- Otondo, Felipe and Rindel, Jens Holger, “A New Method for the Radiation Representation of Musical Instrument in Auralization”, in: *Acta Acustica United With Acustica* 91, 2005: pp. 902–906.
- Pelzer, S., Pollow, M. and Vorländer, M., “Auralization of a virtual orchestra using directivities of measured symphonic instrument”, in: *Proceedings of the Acoustics 2012 Nantes Conference*, 2012, pp. 2379–2384.
- Pierce, Alan D., “Basic Linear Acoustics”, in: Thomas D. Rossing (Ed.), *Handbook of Acoustics*, chap. 3, New York: Springer 2007, pp. 25–111.
- Pollow, Martin and Behler, Gottfried K., “Variable Directivity for Platonic Sound Sources Based in Spherical Harmonics Optimization”, in: *Acta Acustica United With Acustica* 95, 2009: pp. 1082–1092.

- Preibisch-Effenberger, R., *Die SchallokalisationsfähSchall des Menschen und ihre Audioetaudiom Verwendung zur klinischen Diagnostik*, Ph.D. thesis, Technical University of Dresden, Dresden, 1966.
- Pätynen, Jukka and Lokki, Tapio, “Directivities of Symphony Orchestra Instruments”, in: *Acta Acustica United With Acustica* 96, 2010: pp. 138–167.
- Pulkki, Ville, “Multichannel Sound Reproduction”, in: David Havelock, Sonoko Kuwano and Michael Vorländer (Eds.), *Handbook of Signal Processing in Acoustics*, chap. 38, New York: Springer 2008, pp. 747–760.
- Rabenstein, Rudolf and Spors, Sascha, “Sound Field Reproduction”, in: Jacob Benesty, M. Mohan Sondhi and Yiteng Huang (Eds.), *Handbook of Speech Processing*, chap. 53, Berlin, Heidelberg: Springer 2008, pp. 1095–1114.
- Rabenstein, Rudolf, Spors, Sascha and Steffen, Peter, “Wave Field Synthesis Techniques for Spatial Sound Reproduction”, in: Eberhard Hänsler and Gerhard Schmidt (Eds.), *Topics in Acoustic Echo and Noise Control. Selected Methods for the cancellation of Acoustical Echoes, the Reduction of Background Noise, and Speech Processing*, chap. 13, Signals and Communication Technology, Berlin, Heidelberg: Springer 2006, pp. 517–545.
- Reisinger, Gunther, *Einsatz von stereophonen Aufnahmetechniken für die räumliche Übertragung ausgedehnter Schallquellen mit Hilfe der Wellenfeldsynthese*, Diploma thesis, University of Applied Sciences Düsseldorf, Düsseldorf, 2003.
- Reisinger, Markus, *Neue Konzepte der Tondarstellung bei Wiedergabe mittels Wellenfeldsynthese*, Diploma thesis, University of Applied Sciences Düsseldorf, Düsseldorf, 2002.
- Roederer, Juan G., *The Physics and Psychophysics of Music. An Introduction*, fourth edn., New York: Springer 2008.
- Rosenthal, David F. and Okuno, Hiroshi G., *Computational auditory scene analysis*, Mahwah: Lawrence Erlbaum 1998.
- Rossing, Thomas D., *The Science of Sound*, 2nd edn., Reading (Massachusetts): Addison-Wesley 1990.

- Ruckert, Martin, *Understanding MP3. Syntax, Semantics, Mathematics and Algorithms*, Wiesbaden: GWV 2005.
- Schanz, Günther W., *Stereo-Taschenbuch. Stereo-Technik für den Praktiker*, Eindhoven: Philips 1966.
- Schmidhuber, Maximilian, Völn, Florian and Fastl, Hugo, “Psychoakustische Experimente zum Einfluss des Ventriloquismuseffekts auf Richtungsunterschiedsschwellen (Minimum Audible Angles) in der Horizontalebene”, in: *Fortschritte der Akustik – DAGA*, Düsseldorf 2011, pp. 577–578.
- Spors, S., Kuntz, A. and Rabenstein, R., “An Approach to Listening Room Compensation with Wave Field Synthesis”, in: *Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality*, 2003.
- Spors, Sascha, “Extension of an Analytic Secondary Source Selection Criterion for Wave Field Synthesis”, in: *Audio Engineering Society Convention 123*, 2007.
- Spors, Sascha, “Investigation of Spatial Aliasing Artifacts of Wave Field Synthesis in the Temporal Domain”, in: *Fortschritte der Akustik - DAGA*, 2008.
- Spors, Sascha and Ahrens, Jens, “A Comparison of Wave Field Synthesis and Higher-Order Ambisonics with Respect to Physical Properties and Spatial Sampling”, in: *Audio Engineering Society Convention 125*, 2008.
- Spors, Sascha et al., “Sound Field Synthesis”, in: Yiteng Huang and Jacob Benesty (Eds.), *Audio Signal Processing. For Next-Generation Multimedia Communication Systems*, chap. 12, New York, Boston, Dordrecht, London, Moskau: Springer 2004, pp. 323–344.
- Spors, Sascha et al., “Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering”, in: *J. Acoust. Soc. Am.* 122(1), 2007: pp. 354–369.
- Spors, Sascha, Rabenstein, Rudolph and Ahrens, Jens, “The Theory of Wave Field Synthesis Revisited”, in: *Audio Engineering Society Convention 124*, 2008.



- Strube, Gerhard, “Lokalisation von Schallereignissen”, in: Herbert Bruhn, Rolf Oerter and Helmut Rösing (Eds.), *Musikpsychologie. Ein Handbuch in Schlüsselbegriffen*, Munich: Urban & Schwarzenberg 1985, pp. 65–69.
- Technical University Berlin, “Wave Field Synthesis at the TU Berlin”, 2015, URL [www.ak.tu-berlin.de/menue/forschung/wellenfeldsynthese/parameter/en/](http://www.ak.tu-berlin.de/menue/forschung/wellenfeldsynthese/parameter/en/).
- Teutsch, Heinz, *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition*, Berlin, Heidelberg: Springer 2007.
- Theile, Günther, *Über die Lokalisation im überlagerten Schallfeld*, Ph.D. thesis, Technischen Universität Berlin, Berlin, 1980.
- Verheijen, Edwin, *Sound Reproduction by Wave Field Synthesis*, Ph.D. thesis, Delft University of Technology, Delft, 1997.
- Vorländer, Michael, *Auralization. Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, Berlin, Heidelberg: Springer 2008.
- de Vries, Diemer, “Sound Reinforcement by Wave Field Synthesis: Adaption of the Synthesis Operator to the Loudspeaker Directivity Characteristics”, in: *J. Audio Eng. Soc.* 44(12), 1996: pp. 1120–1131.
- de Vries, Diemer, Start, Evert W. and Valster, Vincent G., “The wave field synthesis concept applied to sound reinforcement. Restrictions and solutions”, in: *Audio Engineering Society Convention 96*, Amsterdam 1994.
- Wang, DeLiang and Brown, Guy J., *Computational Auditory Scene Analysis*, Hoboken: IEEE Press 2006.
- Warusfel, Olivier and Misdariis, Nicolas, “Sound Source Radiation Syntheses: From Performance to Domestic Rendering”, in: *Audio Engineering Society Convention 116*, 2004.
- Warusfel, Olivier, Derogis, Philippe and Caussé, Rene, “Radiation Synthesis with Digitally Controlled Loudspeakers”, in: *Audio Engineering Society Convention 103*, 1997.

- Webers, Johannes, *Handbuch der Tonstudioteknik. Analoges und Digitales Audio Recording bei Fernsehen, Film und Rundfunk*, 8. neu bearbeitete edn., Poing: Franzis 2003.
- Werner, Lynne A., “Overview and Issues in Human Auditory Development”, in: Lynne A. Werner, Richard R. Fay and Arthur N. Popper (Eds.), *Handbook of Auditory Research*, chap. 01, New York: Springer 2012, pp. 1–18.
- Wöhe, Walter, “Grundgleichungen des Schallfeldes und elementare Ausbreitungsvorgänge”, in: Wolfgang Fasold, Wolfgang Kraak and Werner Schirmer (Eds.), *Taschenbuch Akustik. Teil 1*, chap. 1.2, Berlin: Verlag Technik 1984, pp. 23–31.
- Williams, Earl G., *Fourier Acoustics. Sound Radiation and Nearfield Acoustical Holography*, Cambridge: Academic Press 1999.
- Wittek, Helmut, *Perceptual differences between wavefield synthesis and stereophony*, Ph.D. thesis, University of Surrey, Guilford, 2007.
- Ziemer, Tim, “Wave Field Synthesis by an Octupole Speaker System”, in: Luiz Naveda (Ed.), *SysMus09 Proceedings*, 2009, pp. 89–93.
- Ziemer, Tim, *Wave Field Synthesis. Theory and Application*, Magister thesis, University of Hamburg, 2011.
- Ziemer, Tim, “Sound Radiation Characteristics of a Shakuhachi with different Playing Techniques”, in: *Proceedings of the International Symposium on Musical Acoustics*, Le Mans 2014, pp. 549–555.
- Ziemer, Tim et al., “Ein Technical Report zum Thema der ‘Just noticeable differences’ (JNDs) zeitlicher Unterschiede in musikalischen Signalen”, 2007, last visited: 11.02.2013, URL [www.systmuwi.de/Pdf/Technical%20Reports/Technical%20Report-JND,%20Below,%20Ziemer,%20etc.pdf](http://www.systmuwi.de/Pdf/Technical%20Reports/Technical%20Report-JND,%20Below,%20Ziemer,%20etc.pdf).
- Zotter, Franz, *Analysis and Synthesis of Sound-Radiation with Spherical Arrays*, Ph.D. thesis, University of Music and Performing Arts, Graz, 2009.
- Zotter, Franz et al., “Capturing the Radiation Characteristics of the Bonang Barung”, in: *3rd Congress of the Alps Adria Acoustics Association*, Graz 2007.

---

Zurek, Patrick M. and Saberi, Kourosh, “Lateralization of two-transient stimuli”,  
in: *Perception and Psychophysics* 65(1), 2003: pp. 95–106.

Zwicker, Eberhard and Fastl, Hugo, *Psychoacoustics. Facts and Models*, second  
updated edn., Berlin, Heidelberg: Springer 1999.

## A. Technical Data

In this section the data of the used recording and reproducing equipment is named in technical, geometrical, and practical respects.

**Microphone Array** For measuring the radiation characteristic of musical instrument, a circular microphone array is deployed. The array contains 128 equidistant omnidirectional MCE-4000 electret microphones with a frequency range from 20Hz to 20kHz and a signal-to-noise ratio of  $> 58\text{dB}$ . It is hung up parallel to the floor at a height of 1m with a radius of 1m. That is one microphone every  $\frac{2\pi}{128} \approx 2.8125^\circ$  with a distance between adjacent microphones of  $\sqrt{(1 - \cos \frac{2\pi}{128})^2 + \sin^2 \frac{2\pi}{128}} \approx 0.05\text{m}$ . The signals are digitalized and preamplified by a RME-system and synchronized via word clock. Two digital interfaces (“Multichannel Audio Digital Interface” MADI) transfer the signals with a sampling rate of 48kHz and a sampling depth of 24bit to a PC. The recording duration is two seconds, so 128 signals à 96000 samples are recorded with a sample depth of 24bit (MADI transfers the signals with a 32bit coding, so the stored sample depth is higher than the actual sample depth). The microphone array is hung up in a free field room, so the recordings contain only direct sound <sup>1</sup>. A loudspeaker in the center of the array, playing a 500Hz-tone serves as calibrator. Its radiation characteristic in that frequency area is assumed to be practically monopole-shaped. The reciprocal of the recorded amplitudes is taken as calibration factor. The signal of damaged microphones is replaced by the arithmetic mean of its two neighbors.

---

<sup>1</sup>Further information on the equipment, see Bader (2010), p. 304.

**Octupole Speaker** (Partly taken from my earlier paper<sup>2</sup>) The octahedron shaped octupole speaker is self-built and intended to recreate the measured radiation characteristics of musical instruments. This platonic solid has eight wooden equilateral triangular planes, each parallel to the opposing. The side length is  $0.3m$ , so the height of one pyramid is  $\frac{\sqrt{2}}{2}0.3m \approx 0.21m$ . The face centers are also the centers of the broadband loudspeaker membranes which have a  $0.08m$  radius. The volume of the loudspeaker cabinet is  $\frac{0.3^3}{3}\sqrt{2} \approx 0.13m^3$ . A photo of the octupole speaker system is given in fig. A.1.



**Figure A.1.:** Photo of the octupole speaker system in the free field room.

The lower  $0.03m$  of the octahedron is sawed off for the loudspeaker cables to go through and for the speaker array to be mounted on a right circular cylinder with a  $0.03m$  diameter. This pedestal with its quadratic  $0.3m^2$  wide base is  $0.91m$  high, so the center of the speaker array is in height of the microphone array.

---

<sup>2</sup>Ziemer (2009).

Through the parallel arrangement of the speakers 0th order radiation (monopole), 1st order radiation (dipole), 2nd order radiation (quadrupole), 3rd order radiation (octupole), and all its combinations can be approximated, presuming the speakers to radiate as simple point sources. But especially at high frequencies this assumption does not hold. The loudspeakers show a complicated radiation pattern. Considering the speakers as complex point sources and measuring their complex amplitude for different angles, their radiated wave fields can be predicted more accurately. By varying the relative amplitudes and phases of the individual speakers, the radiation characteristic can be controlled.

**Loudspeaker surrounding** The loudspeaker surrounding contains 15 Adam Professional Audio A5 Studio/Multimedia monitor speaker in height of head, i.e. the center of the woofer is about 1.70m high, see fig. A.2. Its dimensions are  $28.5 \times 17.2 \times 20$ cm (h/w/d), the woofer has a 14.7cm diameter, the tweeter an equivalent membrane with  $7.1\text{cm}^2$  expanse. The crossover frequency lies at 2.2kHz, the frequency response  $\pm 3\text{dB}$  comprises 55Hz to 35kHz, two 25W Root Mean Square (RMS) amplifiers are built in. XLR connector to 6.3mm audio jack connects the speakers with a 16-channel external sound card which is connected with a PC via FireWire. A rectangular array surrounds the listening area from three sides (left, front, right). Distance between adjacent speaker centers amounts to 65cm.



**Figure A.2.:** Photo from the loudspeaker surrounding system in the wave field synthesis room in the Institute of Musicology at the University of Hamburg.

## A.1. Calibration

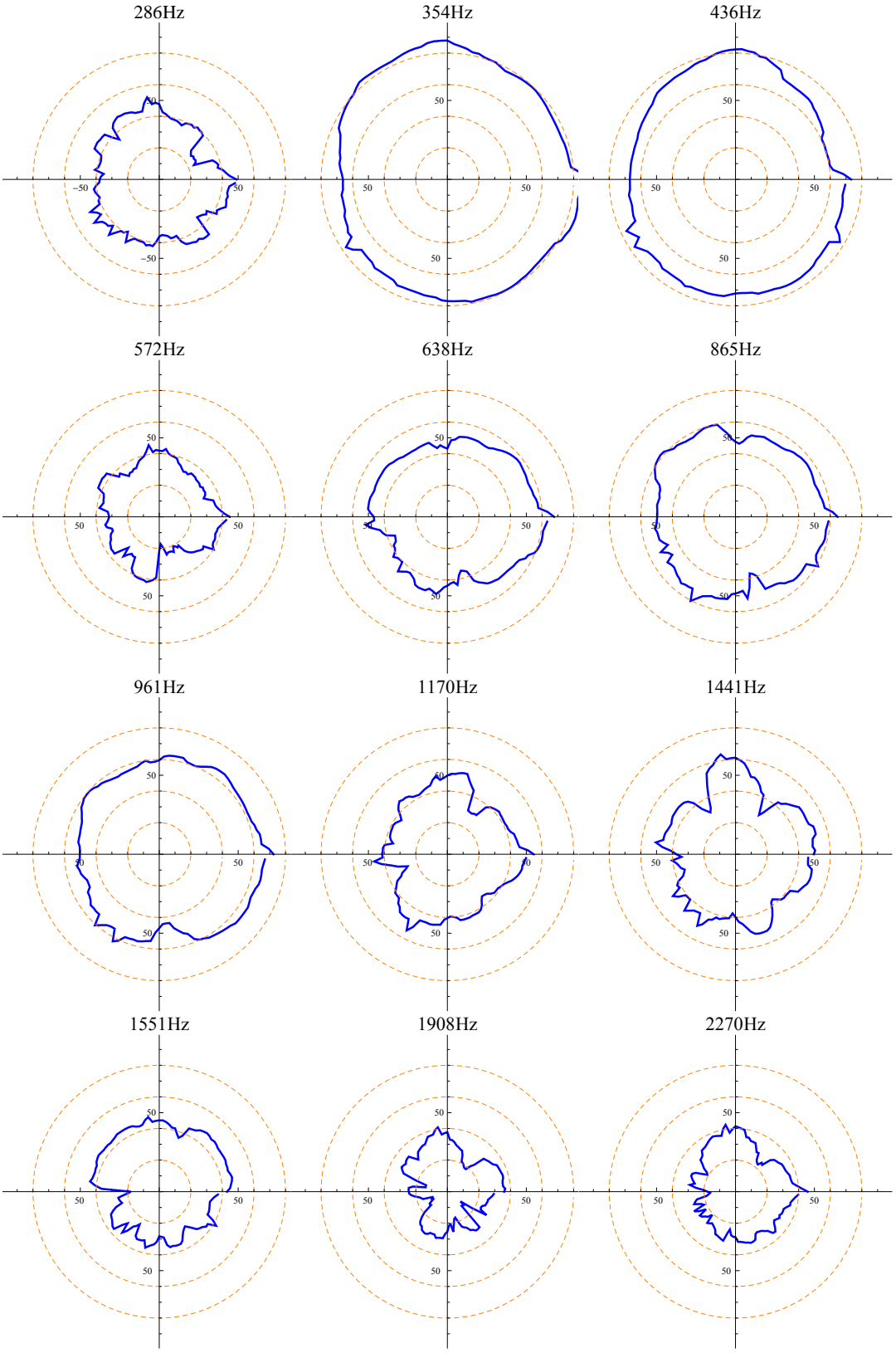
For matters of calibration, an omnidirectional loudspeaker is put in the center of the microphone array in the free-field room, playing a quasi-stationary 100Hz-tone. This sound is recorded simultaneously by all microphones and stored with a sampling frequency of 48000Hz and 24bit sample depth. Assuming that the loudspeaker is a perfect monopole for deep frequencies, all recordings should have equal amplitudes and phases. All recordings have a duration of two seconds. One second from the middle of the signal is extracted, the rest is disregarded. This ensures that transients like the nonlinear onset are eliminated. A discrete Fourier analysis of the extracted signals delivers the complex amplitudes with a precision of 1Hz. The reciprocal value of the 100Hz-amplitudes are the calibration factors by which the measured spectra are multiplied to receive the calibrated signal. If a microphone is corrupt, it is replaced by a new one and the calibration is repeated. If a microphone still delivers a corrupted signal during the measurement, the signal is simply replaced by the signal of the neighboring microphone.

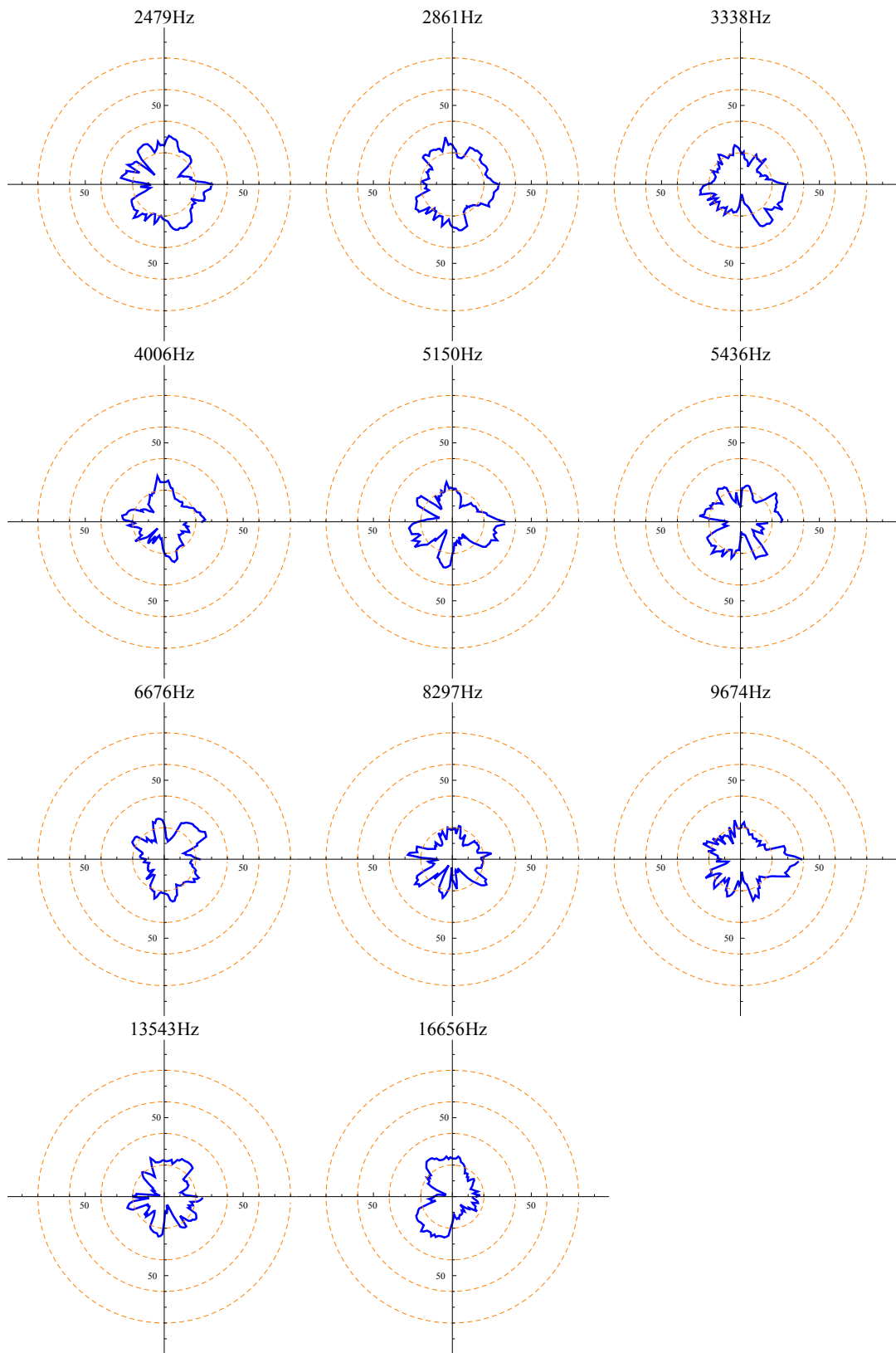


## B. Radiation characteristics

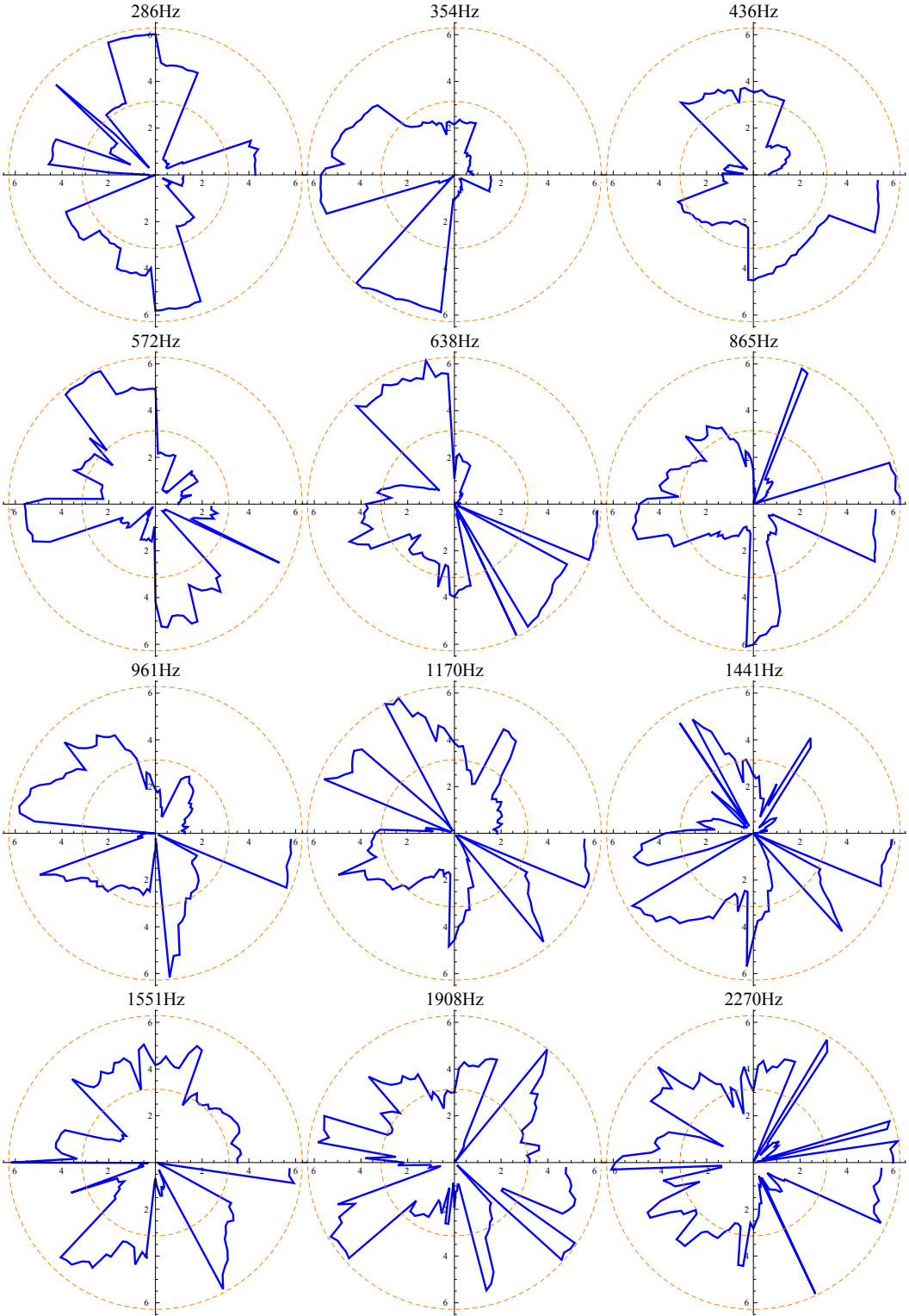
The following pages show exemplary radiation characteristics. First the radiated amplitudes per direction of a shakuhachi are presented, followed by phases of the corresponding frequencies. The shakuhachi sounds do not contain frequencies in the lowest and highest critical band, therefore only 23 radiation patterns are measured and plotted. Then, the radiation characteristics of the loudspeakers used in the loudspeaker surrounding are shown, also divided into polar plots of the amplitudes followed by polar plots of the phase.

The amplitudes are presented in dB without normalization. The instrumentalist, respectively the loudspeaker membrane, faces the right position.





**Figure B.1.:** RC of a shakuhachi (amplitudes).



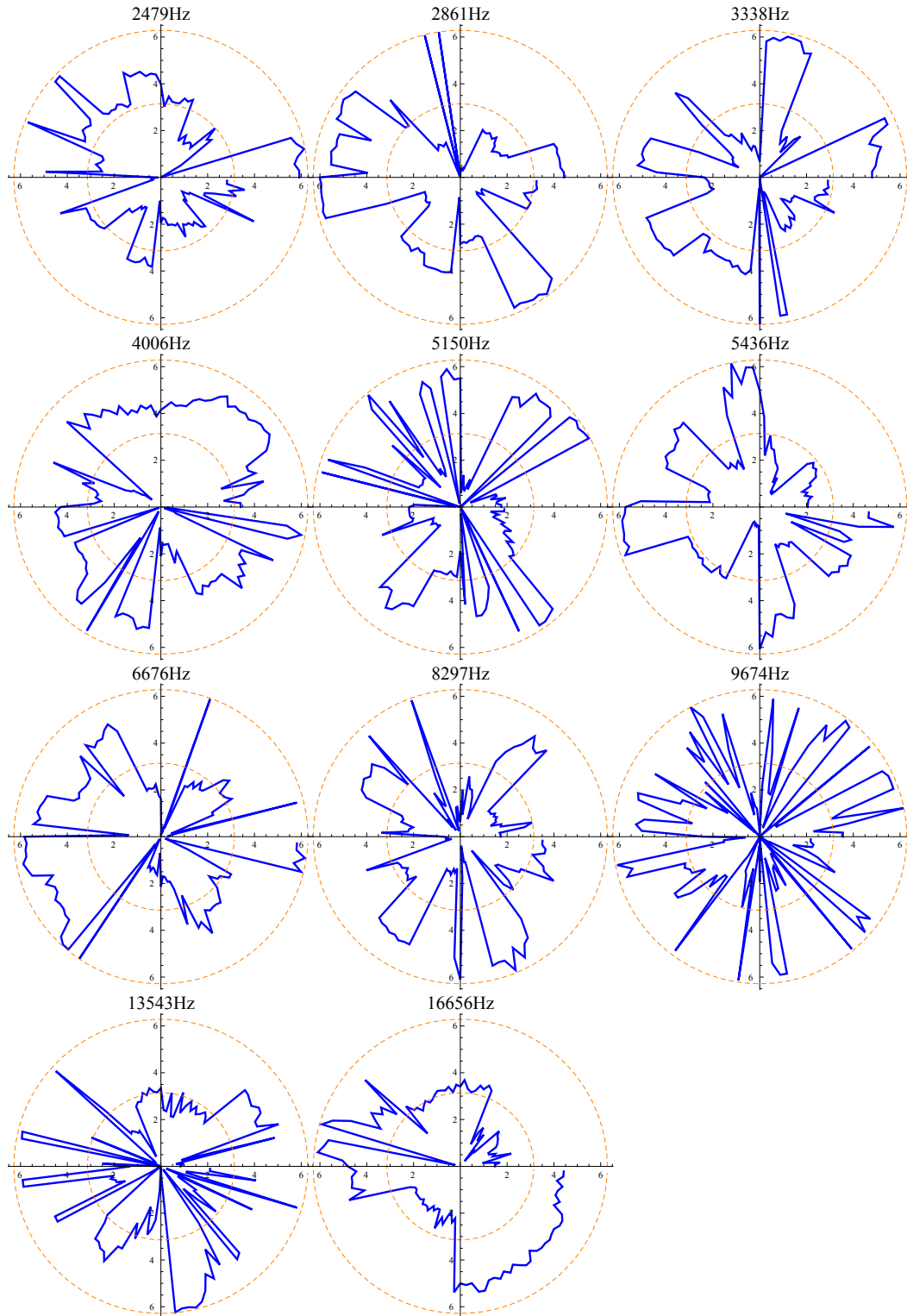
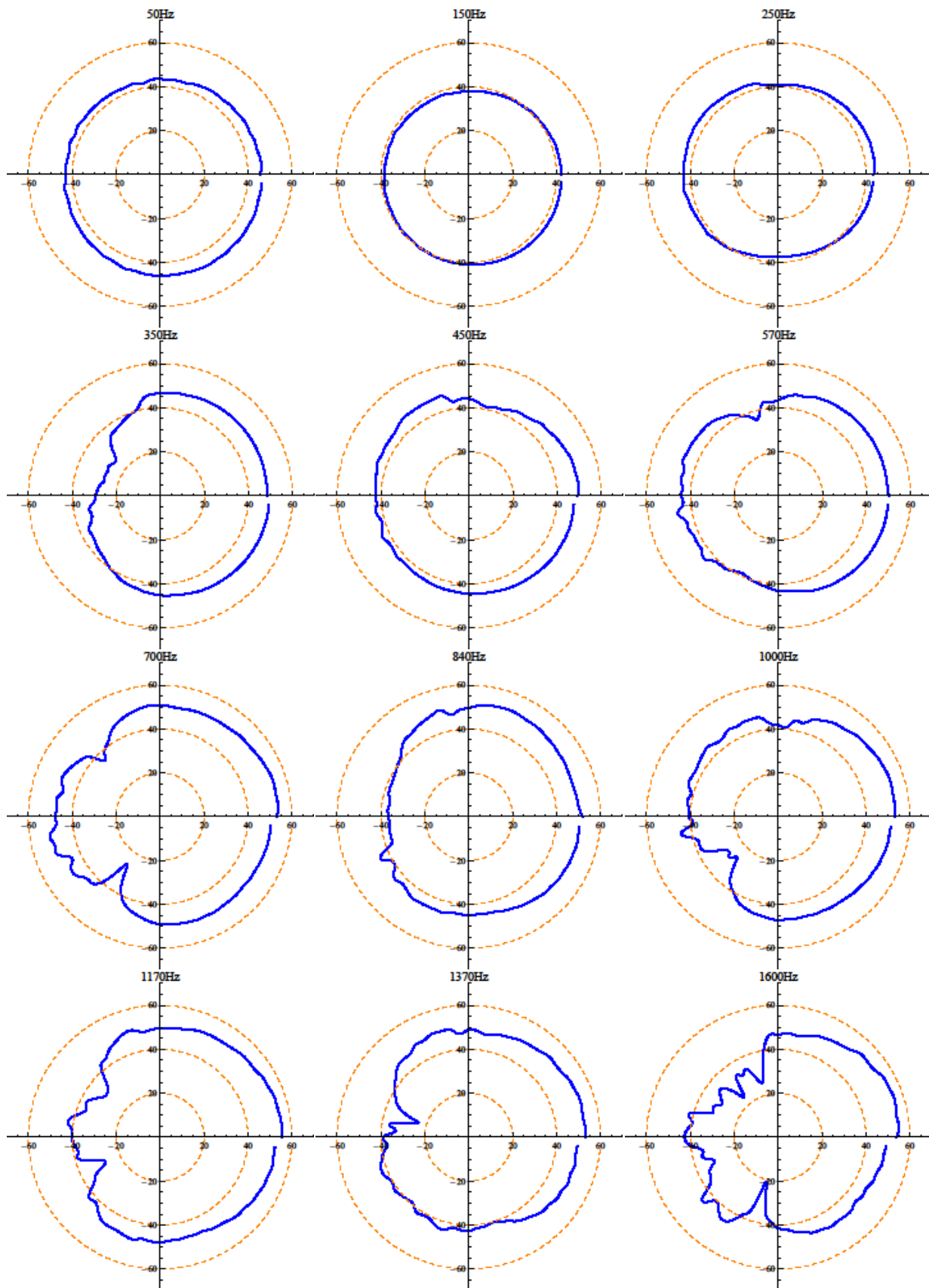
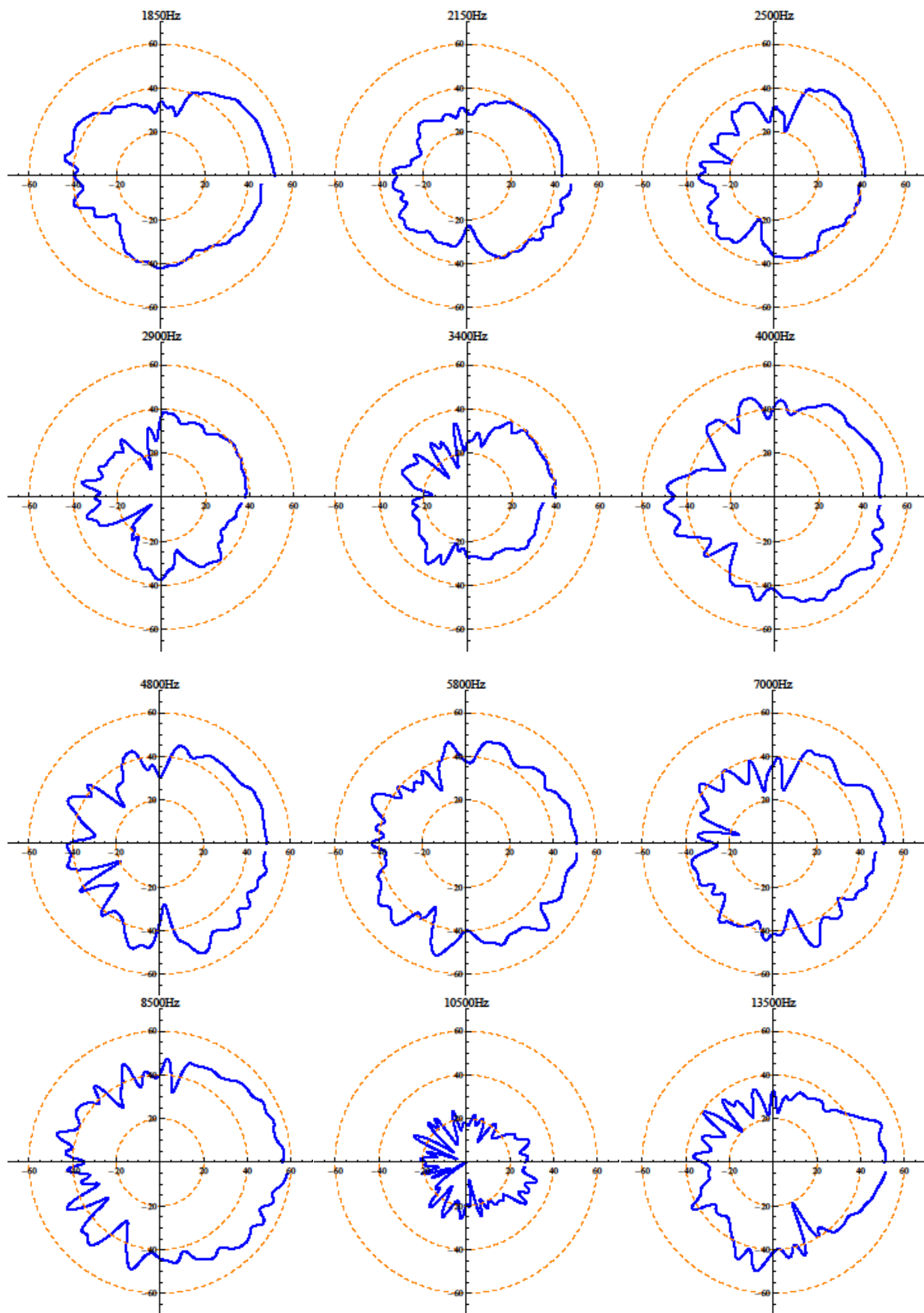
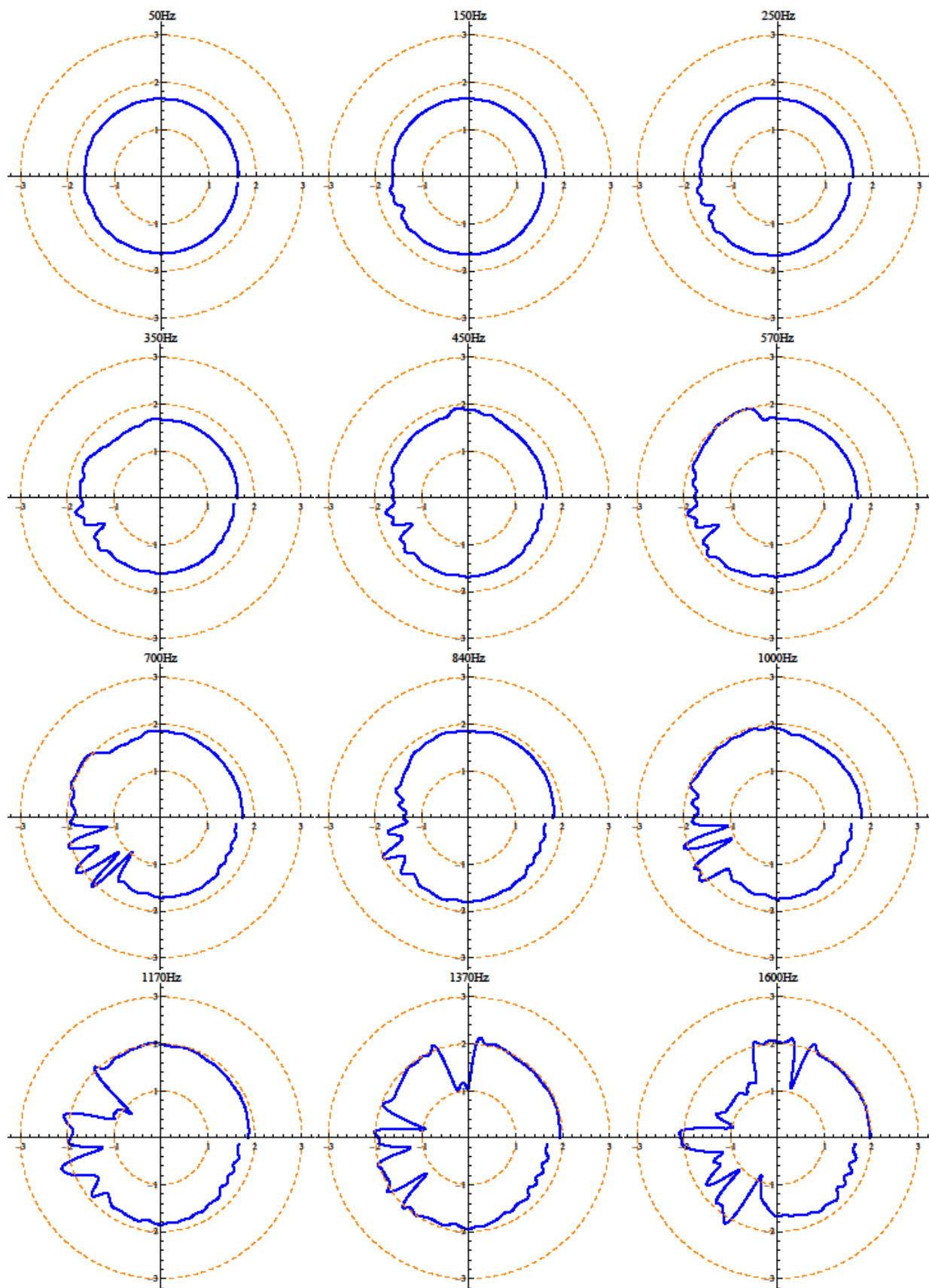


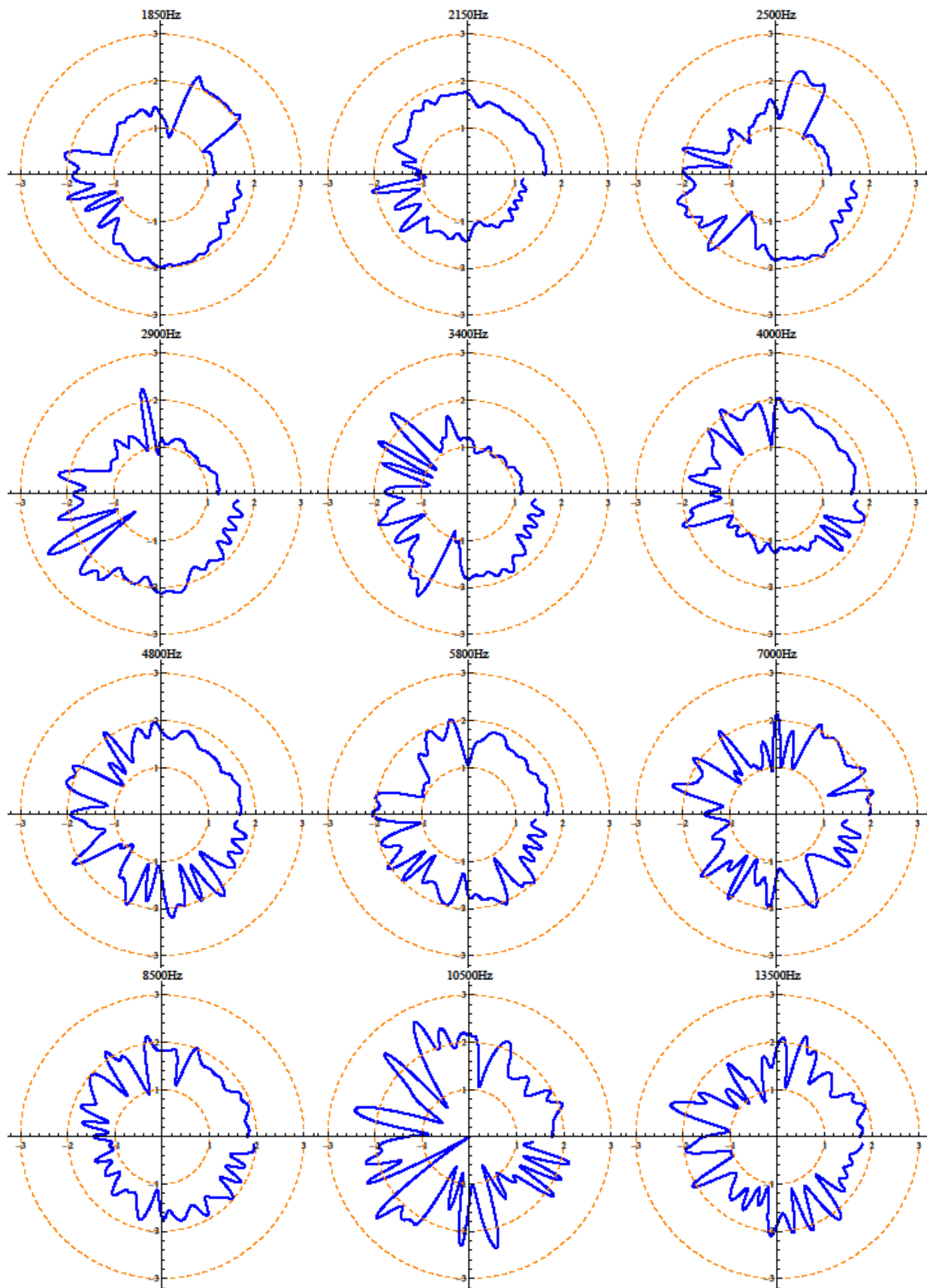
Figure B.2.: RC of a shakuhachi (phases).









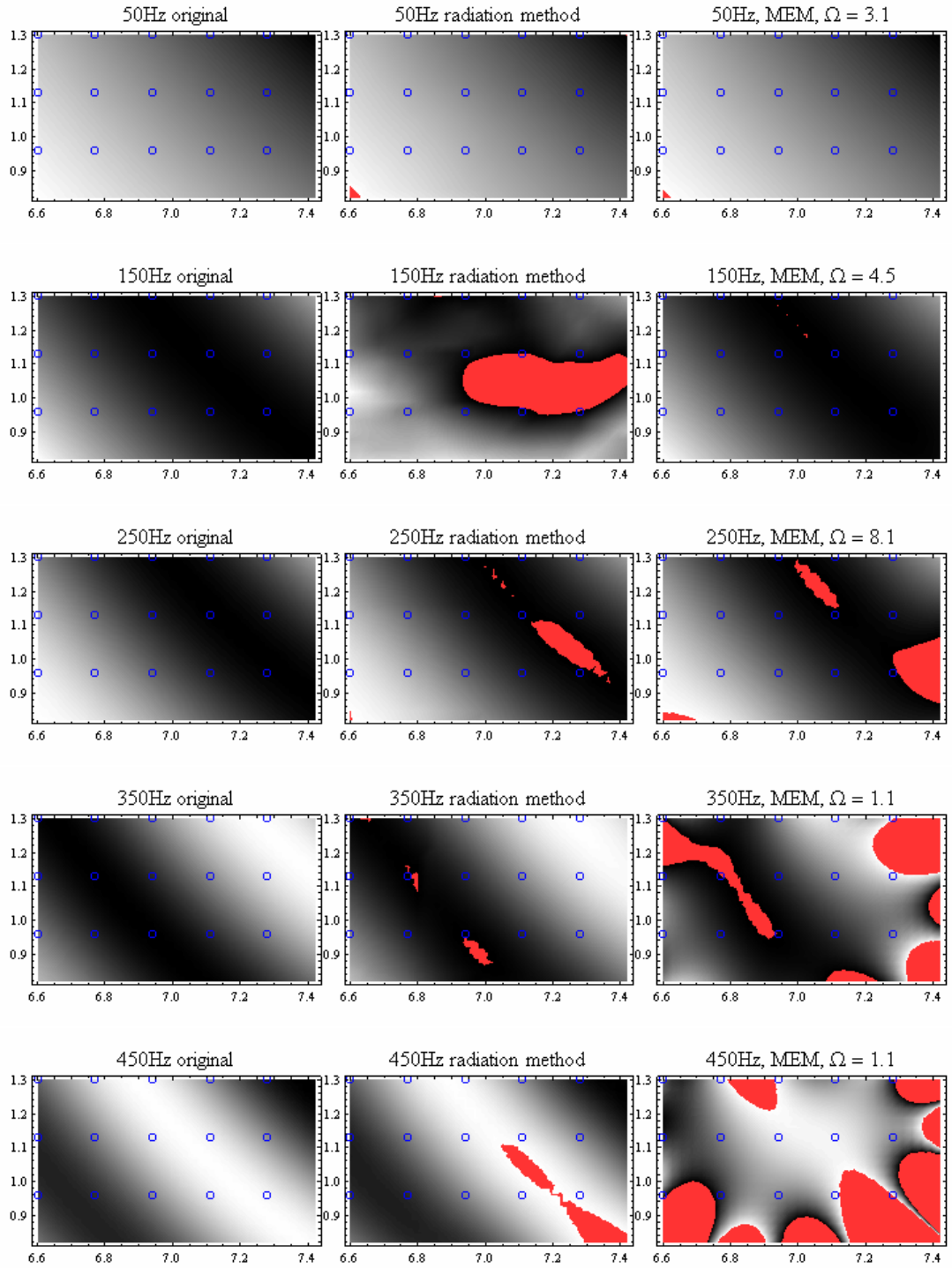


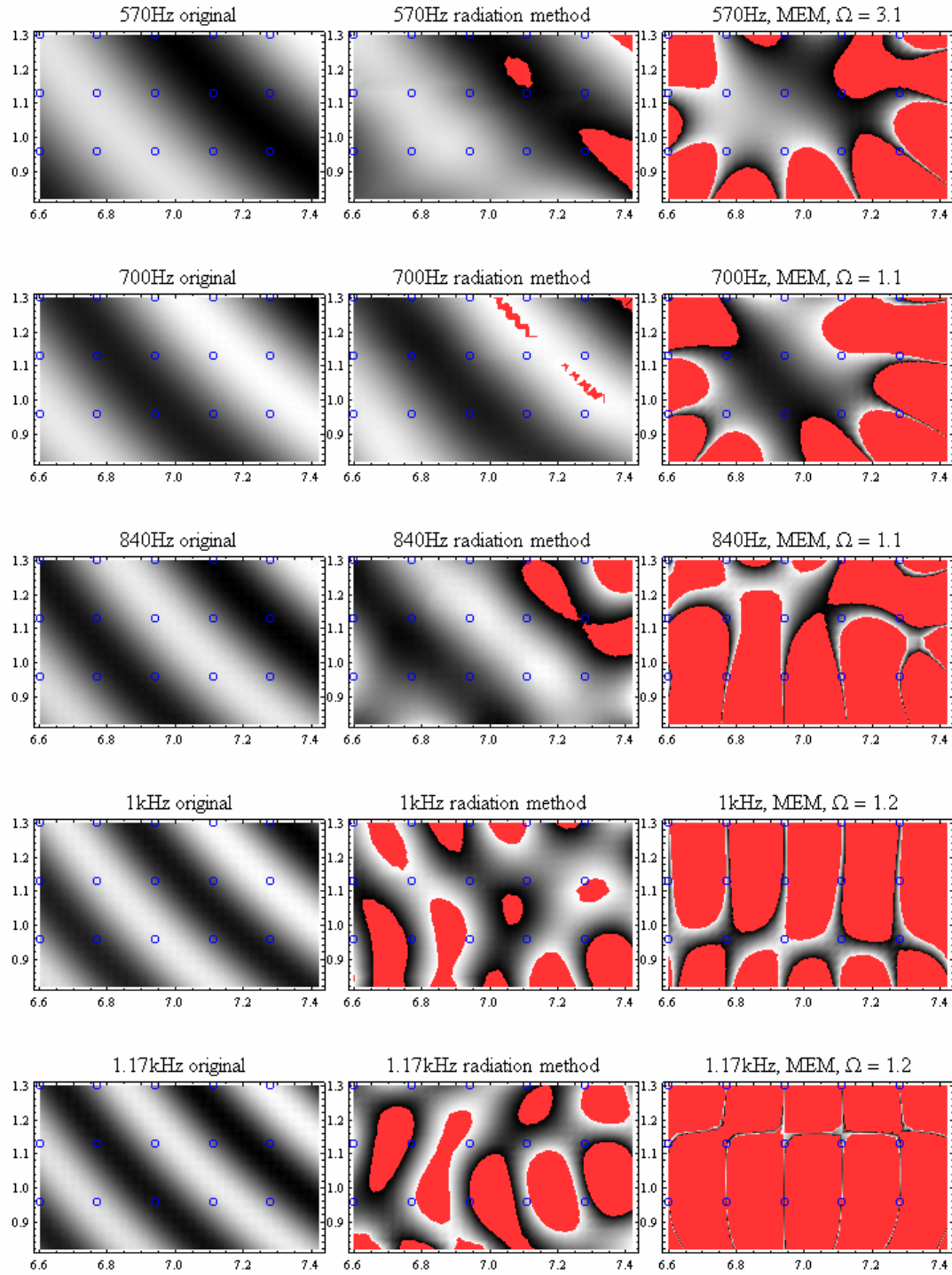
## C. Wave Fields Created from Radiation- and Minimum Energy Method

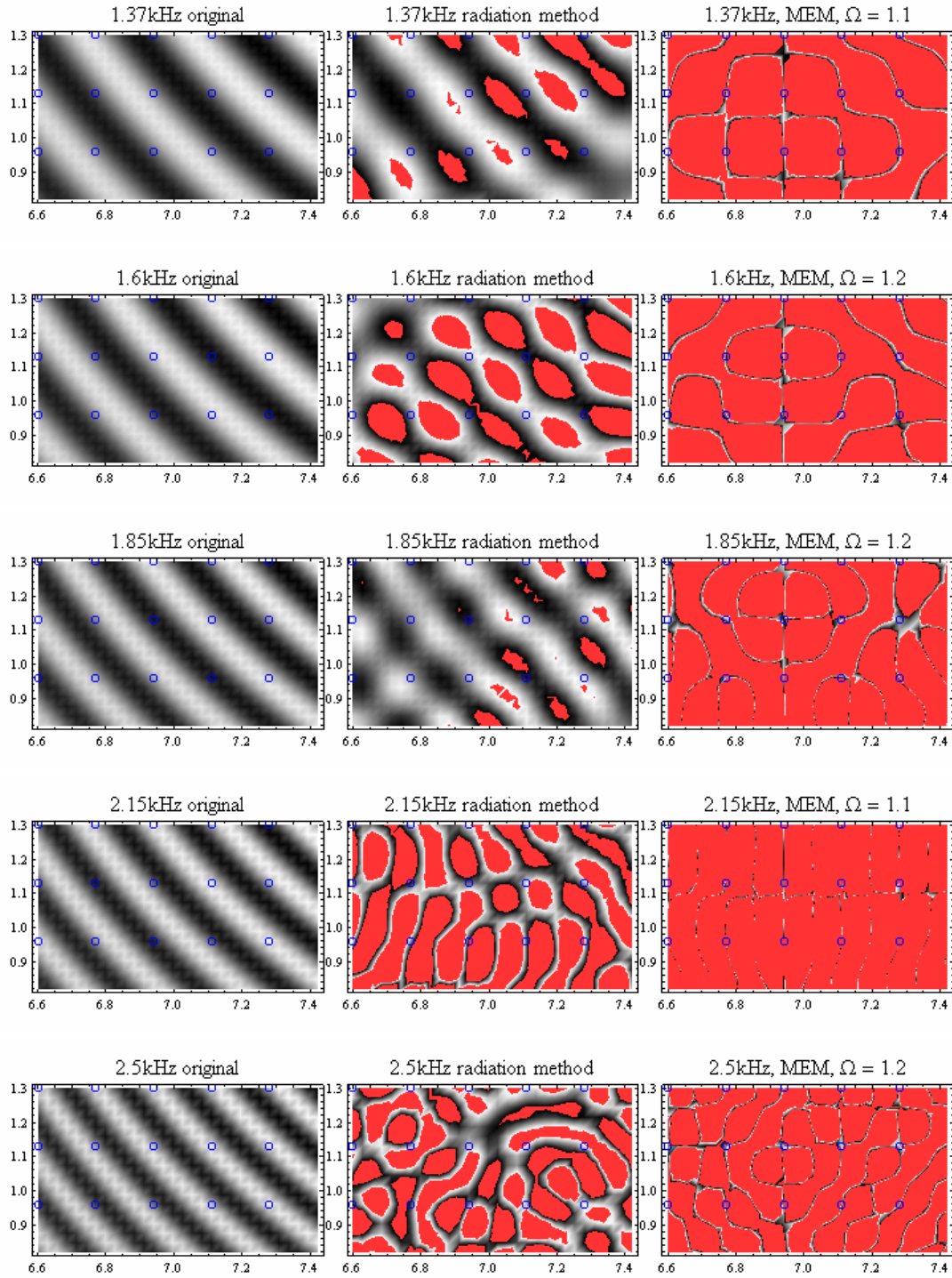
On the next few pages three different calculated wave fields are shown and compared. The first one, fig. C.1 (left), is the wave field as created from a monopole source at position  $\vec{Q} = \begin{bmatrix} 9\text{m} \\ 3\text{m} \end{bmatrix}$  playing the bark sound known from ch. 5.2.2. The second one, fig. C.1 (center), is the wave field as reconstructed by the loudspeaker surrounding applying the radiation method. The third one, fig. C.1 (right), is the wave field as reconstructed applying the MEM on the loudspeaker surrounding. The wave field as reconstructed applying MEM with an increased  $\Omega$ , as discussed in ch. 6.3.2, is not plotted since it does not alter the resulting wave field much.

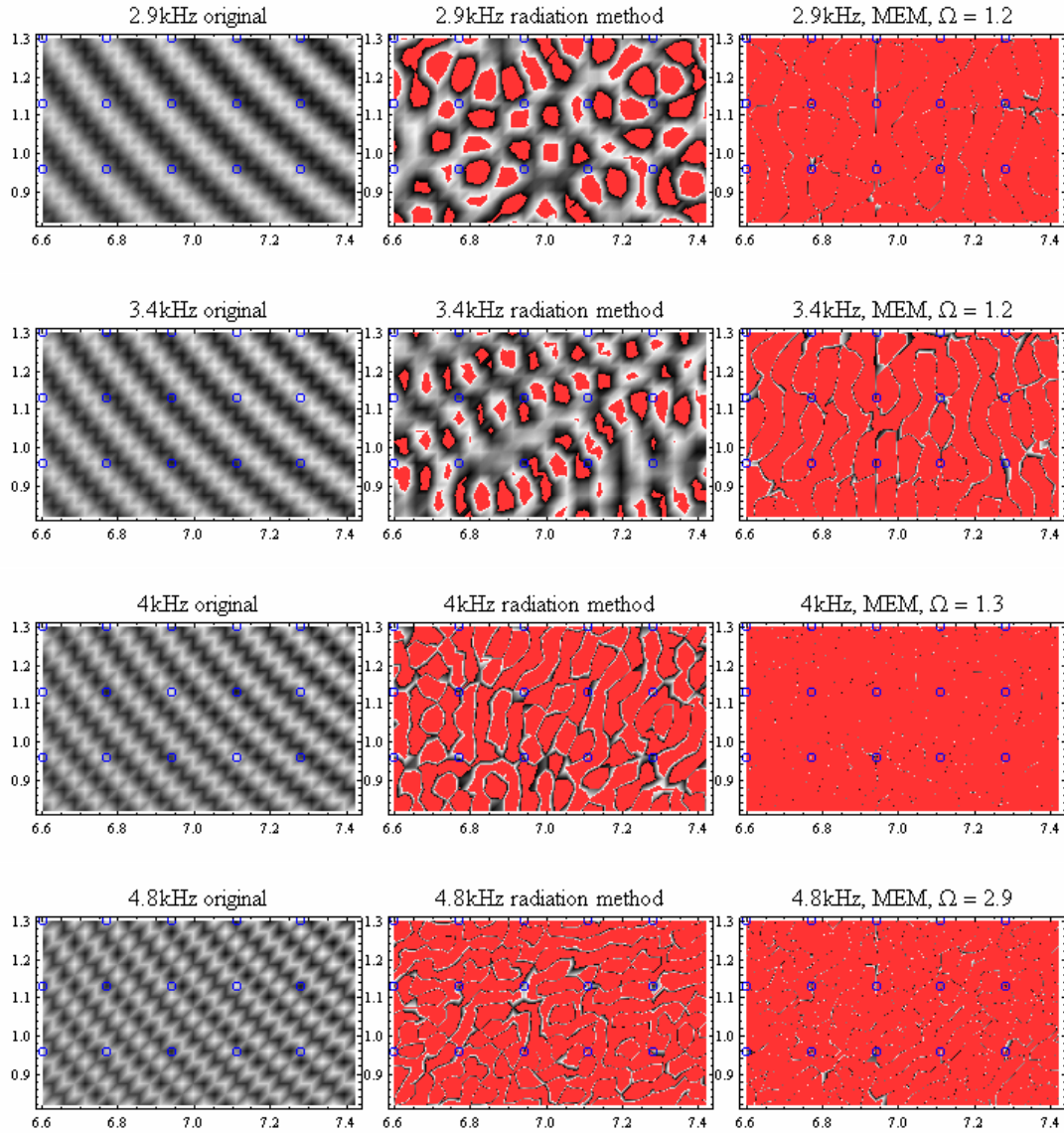
They are all calculated for the 15 listening positions — which are marked by the circles — assuming free field conditions. Since the linear equation system is solved for these exact 15 positions, the wave field is perfectly recreated here with both methods. The real part of the sound field is plotted for the 540 positions, linearly interpolated between the discrete position. They are scaled to the minimum (white) and maximum (black) value of the original wave field for each frequency. The contrasted areas are regions in which the reconstructed sound field exceeds the original wave field, i.e. lies above the maximum or below the minimum value of the original wave field to be reconstructed.

It can be seen that both reconstructions work considerably better at low frequencies. The reconstruction between the calculate listening positions from the radiation method becomes distinctly worse above the critical frequency. The wave field reconstructed by the minimum energy method deviates much more in most cases and shows bad performance even below the critical frequency. Here, reconstruction errors arise mainly at the positions outside the listening area. With increasing frequency, errors spread and above 1.17kHz MEM creates erroneous sound fields at almost every position except the 15 listening positions. Frequencies above 4.8kHz are not plotted because the interpolation becomes delusive at these frequencies due to aliasing. However, as already mentioned in ch. 6.3.2, the reconstruction results become worse at higher frequencies.









**Figure C.1.:** Real part of calculated wave fields from the original source (left) and the sound field reconstruction by radiation method (center) and minimum energy method (right) for different frequencies. The sharp-edged areas denote clipping, i.e. values outside the range of original real values.

## D. Questionnaire

The subjects entered the listening room and the listening test was explained as follows:

“Bitte nehmt Euch einen Testbogen und einen Stift und sucht Euch eine Position in dem markierten Hörbereich aus. An dieser Position bleibt Ihr bitte für die gesamte Dauer des Hörversuchs stehen. Es ist Euch erlaubt Kopf und Oberkörper zu drehen, solange die Beine an derselben Stelle verharren.

Ihr habt nun alle einen aus zwei Zetteln bestehenden Testbogen. Auf diesem sind zunächst 4 gleiche Abbildungen zu sehen. Diese symbolisieren die Lautsprecheranordnung in Schwarz; die Punkte stellen die Lautsprecherpositionen dar. Die roten und blauen Linien dienen zur Orientierung. Sie verlaufen alle  $10^\circ$  vom Mittelpunkt der Lautsprecheranordnung nach außen.

Bitte markiert jetzt in der ersten Abbildung mit einem “X” Eure Position und von dem “X” ausgehend mit einem Pfeil Eure Blickrichtung.

Ich spiele Euch nun viermal das gleiche Musikstück von 30 Sekunden Dauer vor. Bitte markiert in den Abbildungen jeweils die von Euch wahrgenommene Instrumentenposition bzw. den -bereich. Anschließend habt Ihr Zeit die drei Fragen je Durchlauf zu beantworten. Bei der ersten Frage ist eine Doppelantwort möglich, falls der erste Fall zutrifft.

Im Anschluss an die Hörbeispiele füllt Ihr bitte noch die zwei Angaben zu Eurer Person aus.”



The results were already presented in ch. 6.4.1. Further information about the subjects:

All subjects are students of systematic musicology (as major or minor subject) at the University of Hamburg. The age is between 20 and 34 years

Modus: 22 and 23years

Median: 22.5years

Arithmetic mean: 24 years,

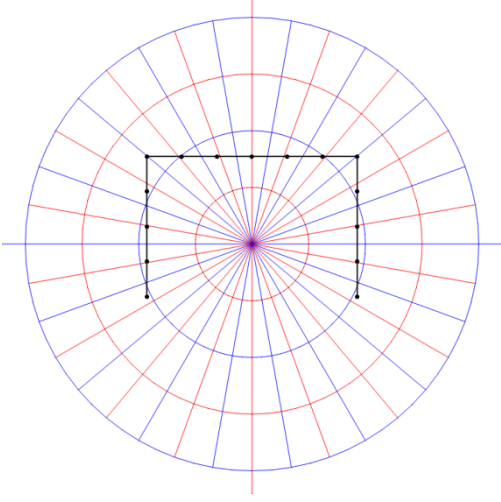
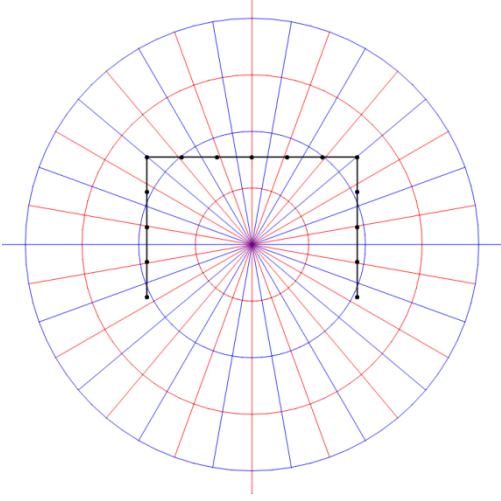
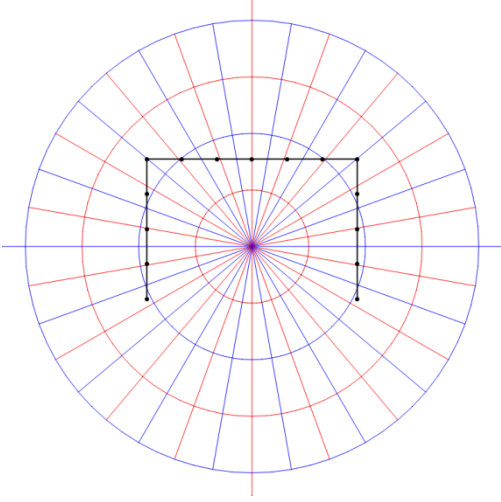
the hight lies in a range between 1.63m and 1.89m

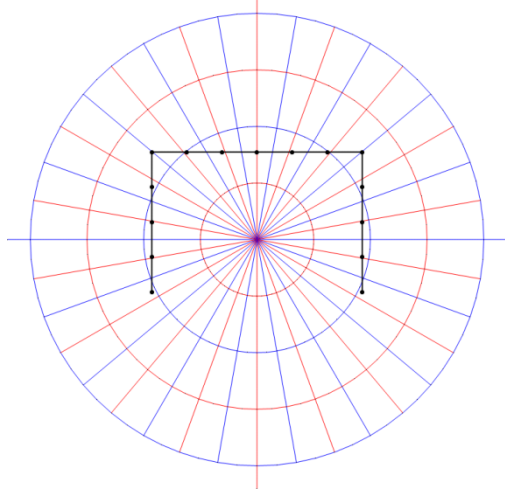
Modus: 1.73m and 1.80m

Median: 1.785m

Arithmetic mean: 1.777m.

The questionnaire will be shown on the following two pages.

<p>↓ 1 ↓</p> 	<p>Wie deutlich ist die Schallquelle lokalisierbar?</p> <p><input type="checkbox"/> mehrere Schallquellen lokalisierbar</p> <p><input type="checkbox"/> überhaupt nicht</p> <p><input type="checkbox"/> sehr vage</p> <p><input type="checkbox"/> ungefähr</p> <p><input type="checkbox"/> deutlich</p> <p><input type="checkbox"/> sehr deutlich</p>	<p>Wie natürlich ist der Instrumentenklang (von völlig unnatürlich 1 bis sehr natürlich 5)?</p> <p><input type="checkbox"/> 1   <input type="checkbox"/> 2   <input type="checkbox"/> 3   <input type="checkbox"/> 4   <input type="checkbox"/> 5</p> <p>Wie räumlich ist der Klang?</p> <p><input type="checkbox"/> gar nicht räumlich</p> <p><input type="checkbox"/> kaum räumlich</p> <p><input type="checkbox"/> ein wenig räumlich</p> <p><input type="checkbox"/> sehr räumlich</p>
<p>↓ 2 ↓</p> 	<p>Wie deutlich ist die Schallquelle lokalisierbar?</p> <p><input type="checkbox"/> mehrere Schallquellen lokalisierbar</p> <p><input type="checkbox"/> überhaupt nicht</p> <p><input type="checkbox"/> sehr vage</p> <p><input type="checkbox"/> ungefähr</p> <p><input type="checkbox"/> deutlich</p> <p><input type="checkbox"/> sehr deutlich</p>	<p>Wie natürlich ist der Instrumentenklang (von völlig unnatürlich 1 bis sehr natürlich 5)?</p> <p><input type="checkbox"/> 1   <input type="checkbox"/> 2   <input type="checkbox"/> 3   <input type="checkbox"/> 4   <input type="checkbox"/> 5</p> <p>Wie räumlich ist der Klang?</p> <p><input type="checkbox"/> gar nicht räumlich</p> <p><input type="checkbox"/> kaum räumlich</p> <p><input type="checkbox"/> ein wenig räumlich</p> <p><input type="checkbox"/> sehr räumlich</p>
<p>↓ 3 ↓</p> 	<p>Wie gut ist die Schallquelle lokalisierbar?</p> <p><input type="checkbox"/> mehrere Schallquellen lokalisierbar</p> <p><input type="checkbox"/> überhaupt nicht</p> <p><input type="checkbox"/> sehr vage</p> <p><input type="checkbox"/> ungefähr</p> <p><input type="checkbox"/> deutlich</p> <p><input type="checkbox"/> sehr deutlich</p>	<p>Wie natürlich ist der Instrumentenklang (von völlig unnatürlich 1 bis sehr natürlich 5)?</p> <p><input type="checkbox"/> 1   <input type="checkbox"/> 2   <input type="checkbox"/> 3   <input type="checkbox"/> 4   <input type="checkbox"/> 5</p> <p>Wie räumlich ist der Klang?</p> <p><input type="checkbox"/> gar nicht räumlich</p> <p><input type="checkbox"/> kaum räumlich</p> <p><input type="checkbox"/> ein wenig räumlich</p> <p><input type="checkbox"/> sehr räumlich</p>

↓ 4 ↓		
	<p>Wie deutlich ist die Schallquelle lokalisierbar?</p> <p><input type="checkbox"/> mehrere Schallquellen lokalisierbar</p> <p><input type="checkbox"/> überhaupt nicht</p> <p><input type="checkbox"/> sehr vage</p> <p><input type="checkbox"/> ungefähr</p> <p><input type="checkbox"/> deutlich</p> <p><input type="checkbox"/> sehr deutlich</p>	<p>Wie natürlich ist der Instrumentenklang (von völlig unnatürlich 1 bis sehr natürlich 5)?</p> <p><input type="checkbox"/> 1   <input type="checkbox"/> 2   <input type="checkbox"/> 3   <input type="checkbox"/> 4   <input type="checkbox"/> 5</p> <p>Wie räumlich ist der Klang?</p> <p><input type="checkbox"/> gar nicht räumlich</p> <p><input type="checkbox"/> kaum räumlich</p> <p><input type="checkbox"/> ein wenig räumlich</p> <p><input type="checkbox"/> sehr räumlich</p>

Alter:

Größe:

## E. Summary of the Results

In this thesis a method to implement the radiation characteristics of musical instruments in wave field synthesis systems is developed. It is applied and tested in two loudspeaker systems. Because the loudspeaker systems have a comparably low number of loudspeakers the wave field is synthesized at discrete listening positions by solving a linear equation system. Thus, for every constellation of listening and source position all loudspeakers can be used for the synthesis. The calculations are done in spectral domain, denying sound propagation velocity at first. This approach causes artefacts in the loudspeaker signals and synthesis errors in the listening area which are compensated by means of psychoacoustic methods. With these methods the aliasing frequency is determined by the extent of the listening area whereas in other wave field synthesis systems it is determined by the distance of adjacent loudspeakers.

Musical instruments are simplified as complex point sources to gain, store and propagate their radiation characteristics. This method is the basis of the newly developed “Radiation Method” which improves the matrix conditioning of the equation system and the precision of the wave field synthesis by implementing the radiation characteristics of the driven loudspeakers. In this work, the “Minimum Energy Method” — originally developed for acoustic holography — is applied for matters of wave field synthesis for the first time. It guarantees a robust solution and creates softer loudspeaker driving signals than the Radiation Method but yields a worse approximation of the wave field beyond the discrete listening positions.

Psychoacoustic considerations allow for a successful wave field synthesis: Integration times of the auditory system determine the spatial dimensions in which the

wave field synthesis approach works despite different arrival times and directions of wave fronts. By separating the spectrum into frequency bands of the critical band width, masking effects are utilized to reduce the amount of calculations with hardly audible consequences. By applying the “Precedence Fade”, the precedence effect is used to manipulate the perceived source position and improve the reproduction of initial transients of notes. Based on Auditory Scene Analysis principles, “Fading Based Panning” creates precise phantom source positions between the actual loudspeaker positions.

Physical measurements, simulations and listening tests prove evidence for the introduced methods and reveal their precision. Furthermore, results of the listening tests show that the perceived spaciousness of instrumental sound not necessarily goes along with distinctness of localization.

The introduced methods are compatible to conventional multi channel audio systems as well as other wave field synthesis applications.

## F. Kurzfassung der Ergebnisse

In dieser Arbeit wird eine Methode entwickelt, um die Abstrahlcharakteristik von Musikinstrumenten in Wellenfeldsynthesesystemen zu implementieren. Diese wird in zwei Lautsprechersystemen umgesetzt und getestet. Aufgrund der vergleichsweise geringen Anzahl an Lautsprechern wird das Schallfeld an diskreten Hörpositionen durch Lösung eines linearen Gleichungssystems resynthetisiert. Dadurch können für jede Konstellation aus Quellen- und Hörposition alle Lautsprecher für die Synthese verwendet werden. Hierzu wird zunächst in Frequenzebene, unter Vernachlässigung der Ausbreitungsgeschwindigkeit des Schalls gerechnet. Dieses Vorgehen sorgt für Artefakte im Schallsignal und Synthesefehler im Hörbereich, die durch psychoakustische Methoden kompensiert werden. Im Vergleich zu anderen Wellenfeldsyntheseverfahren wird bei diesem Vorgehen die Aliasingfrequenz durch die Größe des Hörbereichs und nicht durch den Lautsprecherabstand bestimmt.

Musikinstrumente werden als komplexe Punktquellen vereinfacht, wodurch die Abstrahlung erfasst, gespeichert und in den Raum propagiert werden kann. Dieses Vorgehen ist auch die Basis der neu entwickelten “Radiation Method”, die durch Einbeziehung der Abstrahlcharakteristik der verwendeten Lautsprecher die Genauigkeit der Wellenfeldsynthese erhöht und die Konditionierung der Propagierungsmatrix des zu lösenden Gleichungssystems verbessert. In dieser Arbeit wird erstmals die für die akustische Holografie entwickelte “Minimum Energy Method” auf Wellenfeldsynthese angewandt. Sie garantiert eine robuste Lösung und erzeugt leisere Lautsprechersignale und somit mehr konstruktive Interferenz, approximiert das Schallfeld jenseits der diskreten Hörpositionen jedoch schlechter als die Radiation Method.

Zahlreiche psychoakustische Überlegungen machen die Umsetzung der Wellenfeldsynthese möglich: Integrationszeiten des Gehörs bestimmen die räumlichen Dimensionen in der die Wellenfeldsynthesemethode — trotz der aus verschiedenen Richtungen und zu unterschiedlichen Zeitpunkten ankommenden Wellenfronten — funktioniert. Durch Teilung des Schallsignals in Frequenzbänder der kritischen Bandbreite wird unter Ausnutzung von Maskierungseffekten die Anzahl an nötigen Rechnungen mit kaum hörbaren Konsequenzen reduziert. Mit dem “Precedence Fade” wird der Präzedenzeffekt genutzt, um die wahrgenommene Schallquellenposition zu beeinflussen. Zudem wird dadurch die Reproduktion transienter Einschwingvorgänge verbessert. Auf Grundlage von Auditory Scene Analysis wird “Fading Based Panning” eingeführt, um darüber hinaus eine präzise Schallquellenlokalisation jenseits der Lautsprecherpositionen zu erzielen.

Physikalische Messungen, Simulationen und Hörtests weisen nach, dass die neu eingeführten Methoden funktionieren und zeigen ihre Präzision auf. Auch zeigt sich, dass die wahrgenommene Räumlichkeit eines Instrumentenklangs nicht der Lokalisationssicherheit entspricht.

Die eingeführten Methoden sind kompatibel mit konventionellen Mehrkanal-Audiosystemen sowie mit anderen Wellenfeldsynthesesystemen.

## G. List of Publications Resulting from the Dissertation

Ziemer, Tim, “Wave Field Synthesis by an Octupole Speaker System”, in: Luiz Naveda (Ed.): *SysMus09 Proceedings*, 2009, pp. 89–93.

Ziemer, Tim, *Wave Field Synthesis. Theory and Application*, (magister thesis), University of Hamburg, 2011.

Ziemer, Tim, “A Psychoacoustic Approach to Wave Field Synthesis”, in: *Audio Engineering Society Conference: 42nd International Conference: Semantic Audio*, 2011, pp. 191–197.

Ziemer, Tim, “Psychoacoustic Effects in Wave Field Synthesis Applications”, in: Albrecht Schneider, Arne von Ruschkowski (Eds.): *Systematic Musicology. Empirical and Theoretical Studies*, Frankfurt am Main 2011, pp. 153–162.

Ziemer, Tim, “Sound Radiation Characteristics of a Shakuhachi with different Playing Techniques”, in: *Proceedings of the International Symposium on Musical Acoustics*, 2014, pp. 549–555.